

Identifying Tribes on Twitter Through Shared Context

Gloor, P. A., Fronzetti Colladon, A., de Oliveira, J. M., Rovelli, P., Galbier, M., & Vogel, M.

This is the accepted manuscript after the review process, but prior to final layout and copyediting. **Please cite as:**

Gloor, P. A., Fronzetti Colladon, A., de Oliveira, J. M., Rovelli, P., Galbier, M., & Vogel, M. (2019). Identifying Tribes on Twitter Through Shared Context. In Y. Song, F. Grippa, P. Gloor, & J. Leitão (Eds.), Collaborative Innovation Networks (pp. 91–111). Cham, Switzerland: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-17238-1_5

An extended version of this work is also available here:

<https://doi.org/10.1016/j.ijinfomgt.2019.03.011>

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Identifying Tribes on Twitter through Shared Context

Peter Gloor¹, Andrea Fronzetti Colladon², Joao Marcos de Oliveira³, Paola Rovelli⁴, Manuel Galbier⁵, Manfred Vogel⁶

¹ MIT Center for Collective Intelligence, 245 First Street, 02142 Cambridge, MA, USA, pgloor@mit.edu

² Department of Engineering, University of Perugia, Via G. Duranti 93, 06125, Perugia, Italy, andrea.fronzeticolladon@unipg.it

³ galaxyadvisors AG, Aarau, Switzerland, jmarcos@galaxyadvisors.com

⁴ Department of Economics and Management, Free University of Bozen-Bolzano, Piazza Università 1, 39100, Bolzano, Italy, paola.rovelli@unibz.it

⁵ University of Applied Sciences Northwestern Switzerland, Brugg, Switzerland, manuel.galbier@fhnw.ch

⁶ University of Applied Sciences Northwestern Switzerland, Brugg, Switzerland, Manfred.vogel@fhnw.ch

Abstract This paper introduces *Tribefinder*, a novel system able to reveal Twitter users' tribal affiliations. *Tribefinder* establishes to which tribes individuals belong through the analysis of their tweets and the comparison of their vocabulary. These tribal vocabularies are previously generated based on the vocabulary of tribal influencers and leaders selected using *Tribecreator*. To demonstrate its functionality, in the case presented in this paper, the system was calibrated in three specific tribal macro-categories: *alternative reality*, *lifestyle*, and *recreation*. Apart from describing the methodology we used to create this system, we also provide some practical examples of its use, thus giving a first indication of its potential. Finally, we present the results of the adoption of a T-SNE visualization approach, useful to verify whether tribe members cluster closely together.

1 Introduction

A tribe is “a network of heterogeneous persons linked by a shared passion or emotion” (Cova and Cova 2002). In other words, a tribe is a means whereby individuals experience a sense of community and share strong emotional links, common culture, passions, and vision of life (Cova 1996; Cova and Cova 2002; Richardson 2013). Individuals break up in several different tribes and each of them may belong to many smaller and larger tribes, playing different roles and wearing different

masks (Cova 1996; Cova and Cova 2002). Individuals' expressed behaviors reveal to what tribes they belong and how they perceive their own identity (e.g., Garry et al. 2008). Indeed, each tribe has its own peculiarities, behaviors, rituals, traditions, myths, values, beliefs, hierarchy, and vocabulary (Cova and Pace 2006), which support the identification of individuals' tribal affiliations.

Observing the emergence of "tribalism" (Bauman 1990; Maffesoli 1996), it became clear that understanding its "tribes" is essential for firms' survival (e.g., Holzweber et al. 2015), being especially important for marketing (e.g., Goulding et al. 2013; Kozinets 1999). To extend traditional marketing strategies (Addis and Podesta 2005; Canniford 2011), scholars have started suggesting to firms to rethink their marketing activities (Cova and Cova 2002; Moutinho et al. 2007), taking in account the existence and behavior of their consumer tribes - i.e. "tribal marketing" (Cova and Cova 2002). Tribes' characteristics may indeed affect the success of a marketing campaign, even if few studies exist so far on how they can be used as a strategic resource (Cova and Cova 2002). Marketing actions should be designed depending on the tribes that have to be addressed given the characteristics of the firm, its brand, and the product or service it offers (Moutinho et al. 2007). At the same time, knowing what types of tribes are particularly attracted by a specific product or brand may be a powerful instrument to improve marketing of this product or brand. In doing so, firms have the possibility to design their marketing actions in line with the individual and social needs of tribes' members (e.g., Cova 1996; Holzweber et al. 2015), thus maximizing the probability of success. However, the identification of tribes is difficult and requires different and special efforts (Cova and Cova 2002). Moreover, the advent of the Internet and the growing use of social media as marketing instruments (Burton and Soboleva 2011) challenge even more the identification of the so-called *virtual tribes*, meaning tribes that nowadays form by communication technologies (Cova and Pace 2006). This, in turn, calls for new methodologies to properly identify these virtual tribes. This is particularly true given the limits of the traditional approaches used by existing studies on consumer tribes - e.g. ethnography and netnography (Cova and White 2010; Goulding et al. 2013; Hamilton and Hewer 2010), focus groups, (Dionísio et al. 2008; Moutinho et al. 2007), interviews, (Cova and Cova 2002; Cova and Pace 2006; Holzweber et al. 2015), and surveys, (Taute and Sierra 2014) - which do not allow to automatically and systematically identify virtual tribes and their characteristics.

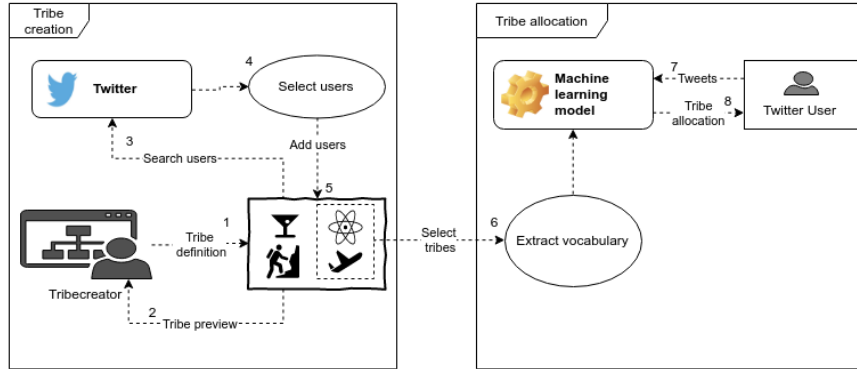
Our paper presents a novel system, called *Tribefinder*, to identify virtual tribes (hereafter: *tribes*). Leveraging Twitter, it analyzes an individual's tweets and categorizes her/him into tribes belonging to three specific tribal macro-categories: *alternative reality*, *lifestyle*, and *recreation* (De Oliveira and Gloor 2018). While these macro-categories have been chosen just as examples to demonstrate how the system works, *Tribefinder* can be easily extended to other macro-categories depending on the user's needs. Through this system it is possible to automatically classify any individual into her/his tribal affiliations by any macro-category that is of interest for the analyst. After presenting our novel system and its validating it, we provide a T-SNV visualization of tribe members. T-SNE offers an intuition on the quality of the

tribe classification generated by *Tribefinder* by verifying whether individual tribe members cluster closely together.

2 Developing a System to Reveal Tribes

The continuous stream of tweets is an important source of information (e.g., Bringay et al. 2011), which offers a powerful setting for studying and identifying tribes of individuals. The goal of the proposed system *Tribefinder* is to categorize Twitter users into alternative orthogonal tribes. This is made possible by extracting information about key people, brands, and topics from their tweets. *Tribefinder* provides as output the tribal affiliations of an individual, consistent with three tribal macro-categories: *alternative reality*, *lifestyle*, and *recreation*. These three specific macro-categories have been chosen to provide an example of how the system functions. This is not intended to be a limitation of the system. *Tribefinder* can be personalized depending on the interests of who uses the system. Applying the same methodology, *Tribefinder* is currently being extended to similarly identify user-defined tribal macro-categories.

Fig. 1. *Tribefinder* system architecture.



The *Tribefinder* system consists of two main components (Fig. 1): the *tribe creation* and the *tribe allocation* modules. To create, and then train *Tribefinder*, a user first has to identify key individuals who represent the different predefined tribes for each tribal macro-category (e.g., the tribes *nerd*, *fatherlander*, *spiritualist*, and *tree-hugger* for the macro-category *alternative reality*). Through this process, a large sample of Twitter users is generated belonging to each of these newly created tribes, defined by the concepts, ideas, and artifacts that may describe them. A tribe can be idealized as a concept, idea, or artifact that its members believe in or like (De Oliveira and Gloor 2018). More specifically, this search is performed using *Tribecreator* (De Oliveira and Gloor 2018), a Web tool that allows users to automatically

find individuals by keywords expressing concepts, ideas and beliefs, using four search functions. New tribe users can be searched based on the match between the tribe’s general characteristics and the individual’s (i) Twitter profile description, (ii) tweets, (iii) followers, and (iv) friends (i.e., those whom s/he follows)¹.

The Twitter timeline of the users that likely belong to each predefined tribe is then gathered by *Tribecreator*. This collected data is subsequently used to create a tribal vocabulary and the machine-learning model to find the tribal affiliations of a given individual. However, our system also utilizes this information to get a preliminary understanding of the tribal affiliations of the individuals previously extracted. The characteristics of such a newly created tribe can be visualized in three ways. First, *Tribecreator* draws a network of the tribe’s members, to have a first idea about the most influential individuals. Second, a hashtag word cloud can be generated, to identify the top hashtags. Third, the most popular posted links can be shown.

As mentioned before, once a tribe has been created, its tribal vocabulary is computed. This final step to make the system learn on how to associate random individuals with specific tribes consists of the analysis of the language these influential tribal leaders use through deep learning. In so doing, classifiers are created using embedding and LSTM (long short-term memory) models. Specifically, these classifiers work by collecting the Twitter feeds of all the users from the tribes that *Tribefinder* is training on. On these, embedding is applied to map words into vectors, which are then used as input for the following LSTM models. LSTM models are deep learning models specially designed to analyze sequential data, which are used in this case to analyze not only what individuals say on social media, but also how they say it. The model thus tries to learn how to predict a tribal affiliation for a single tweet. Once a tribe is predicted for each tweet, *Tribefinder* sums up the result to have a tribe distribution for the user timeline. In other words, analyzing recurring concepts in the tweets of influential leaders, *Tribefinder* identifies the textual patterns that characterize each tribe and generates a specific tribal vocabulary.

The following Table 1 summarizes tribal macro-categories and actual tribes we identified. Specifically, *Tribefinder* uses three macro-categories to define individuals’ tribal affiliations (i.e., *alternative reality*, *lifestyle*, and *recreation*). Looking for instance at the *alternative reality* to which individuals belong, *Tribefinder* separates them into four tribes: *nerds*, *treehuggers*, *spiritualists*, and *fatherlanders*. The so-called *nerds* are technocrats who believe in a global world ruled by capital and technology, the *treehuggers* fight for protecting the environment, while the *spiritualists* are individuals who mainly focus their attention on the spiritual side of things. On the opposite side, the *fatherlanders* are ultra-patriots who want to recreate the national states of the early twentieth century.

Using *Tribefinder* and the tribal vocabulary it learned, it is now possible to establish the tribal affiliations of every Twitter user. In practice, *Tribefinder* analyzes the individual’s word usage in her/his tweets and then assigns the corresponding

¹ For followers and friends, their tweets are analyzed to understand whether the individual is connected with accounts that post tweets on topics related to the tribe.

alternative reality, lifestyle, and recreation tribal affiliation based on the similarities with the specific tribal vocabularies.

Table 1. Tribefinder tribal macro-categories and tribes.

Tribal macro-category	Tribes	Description
Alternative reality	Fatherlander	They believe in God and fatherland, and that their fatherland is the best one. They cling to the good old times, hold the idea of the family in high regard and have little time for foreigners
	Nerd	They believe that progress, science and technology are a blessing. They want to overcome death and colonize Mars. They are fans of globalization and network with each other
	Spiritualist	They believe in a subjective experience of a sacred dimension. They find strength in contemplation, and their behavior is driven by the search for sacred meaning
	Treehugger	They believe in the limits of growth and in the protection of nature. They challenge some elements of technological progress (e.g., gene manipulation) and welcome others (e.g., alternative energies)
Lifestyle	Fitness	They love doing sports and are addicted to training. They show an almost compulsive engagement in any form of physical exercise
	Sedentary	Opposite to the fitness addicted, they are characterized by much sitting and little physical exercise
	Vegan	They follow a plant-based diet avoiding all animal foods, as well as avoiding using animal products
	Yolo	They follow the motto “You only live once” and they think that one should make the most of the present without worrying about the future (“carpe diem”). As a consequence, they often adopt impulsive and reckless behavior
Recreation	Art	They are interested in any form of art (e.g., paintings, sculptures, music, dance, literature, films), of which they appreciate the beauty and emotional power
	Fashion	They are interested in popular or the latest style of clothing, hair, decoration, or behavior
	Sport	They love watching any kind of sport on TV, and attending sports events. Some also actually like to practice these sports
	Travel	They love travelling around in the world, for both pleasure and business, experiencing different cultures and environments

3 Tribefinder in action

In this section, we provide some examples of the use of the *Tribefinder* system, empirically validating its accuracy. As mentioned above, we are confident that firms may find *Tribefinder* useful for marketing. While a firm *a-priori* knows what kind of customers it wants to reach through its marketing activities, the *ex-post* results may not be as expected. *Tribefinder* thus offers a simple instrument to assess the alignment between the expected and actual characteristics of a *brand's virtual tribe*, which identifies the network of heterogeneous Twitter users that share an interest in this specific brand. In this way, it might reveal that particular tribes have become (unintentionally) attracted by marketing actions, which may in turn become a possible source of innovation for the firm.

To empirically test *Tribefinder's* accuracy, we selected four firms, brands, or key individuals (hereafter: *brands*) for each tribe category, whose target customers' (or audience's) characteristics fit with those of the tribes. We then identified and analyzed, using *Tribefinder*, the tribes of the users that tweeted about these brands, to measure their tribal affiliation and verify its congruence with the brand image. The results are presented below, divided into the three tribal macro-categories.

3.1 Alternative reality

In this section, we provide the percentage tribal affiliations for brands that specifically target *fatherlanders* (i.e., CNN, Fox News, MSNBC News, Politico), *nerds* (i.e., Apple, Microsoft, SpaceX, Star Wars), *spiritualists* (i.e., Dalai Lama, Paolo Cohelo, Osho, YogaWorks), and *treehuggers* (i.e., Greenpeace, Patagonia, PETA, WWF). On the vertical axis the percentage of analyzed Twitter users that fall into the specific tribe is reported.

Fig. 2. Tribal affiliations of Twitter users in the virtual tribes of four *fatherlander* brands.

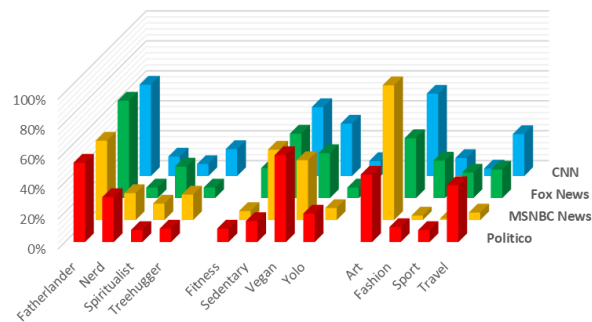


Fig. 3. Tribal affiliations of Twitter users in the virtual tribes of four *nerd* brands.

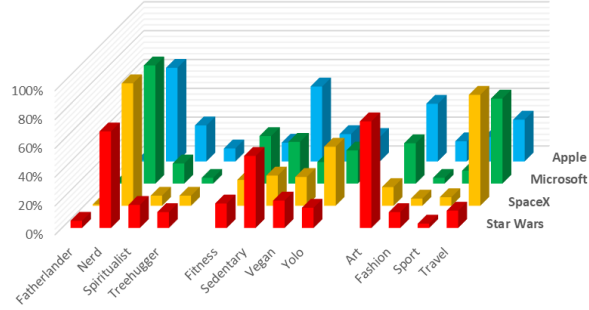


Fig. 4. Tribal affiliations of Twitter users in the virtual tribes of four *spiritualist* brands.

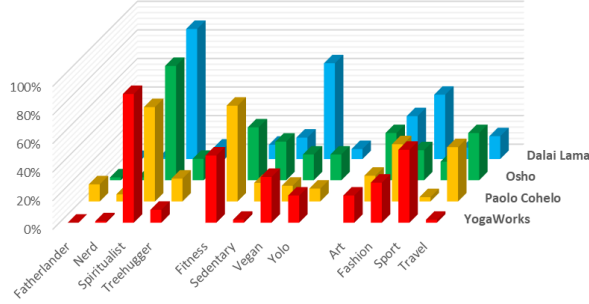


Fig. 5. Tribal affiliations of Twitter users in the virtual tribes of four *treehugger* brands.

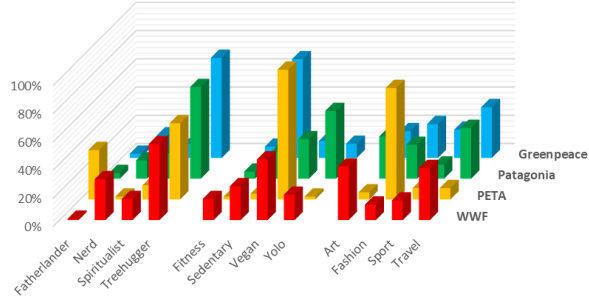
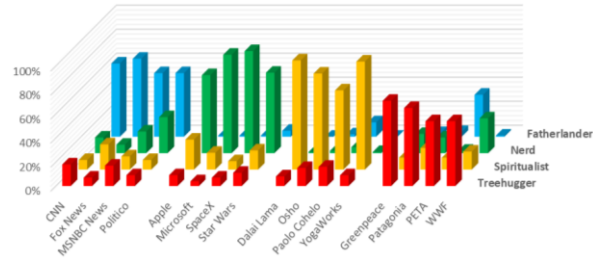


Fig. 2 shows that the majority of the Twitter users in the virtual tribes of the four selected *fatherlander* brands correctly fall into the *fatherlander* tribe. Moreover, as individuals typically belong to several tribes (Bauman 2000), looking at other tribal macro-categories, these users are also *sedentary* or *vegan* (depending on the brand) and interested in *art* (e.g., those tweeting about MSNBC News).

The correct functioning of the *Tribefinder* system becomes even clearer when looking at *nerd* brands (Fig. 3), which mostly attract *nerd* individuals. The same holds for *spiritualist* brands (Fig. 4), whose Twitter users are *spiritualist* as well. Confirming the validity of our system, Dalai Lama related individuals properly fall into the *vegan* tribe, while those associated with YogaWorks also belong to the *fitness* tribe. For *treehugger* brands (Fig. 5), the corresponding Twitter users are accurately classified as *treehuggers*.

Fig. 6. *Alternative reality* tribal affiliations of Twitter users belonging to *alternative reality* brands' virtual tribes.



Finally, Fig. 6 provides a concise view of the results presented above. Specifically, it shows the tribal affiliations of the aforementioned brands' virtual tribes only referring to the tribal macro-category of the analysis (i.e., *alternative reality*). From Fig. 6, a direct correspondence between brands' types and individuals' tribal affiliations is clearly visible. For instance, *nerd* brands attract *nerd* Twitter users.

3.2 Lifestyle

In this section we present the average tribal affiliations of the Twitter users engaged with brands that specifically target different *lifestyles*: *fitness* (i.e., Adidas, CrossFit, Nike, Peloton), *sedentary* (i.e., GrubHub, InstaCart, PizzaHut, Seamless), *vegan* (i.e., Beyond Meat, Impossible Foods, PETA, WWF), and *yolo* (i.e., Alpinestars, GoPro, Monster Energy, Rockstar Energy).

Fig. 7. Tribal affiliations of Twitter users in the virtual tribes of four *fitness* brands.

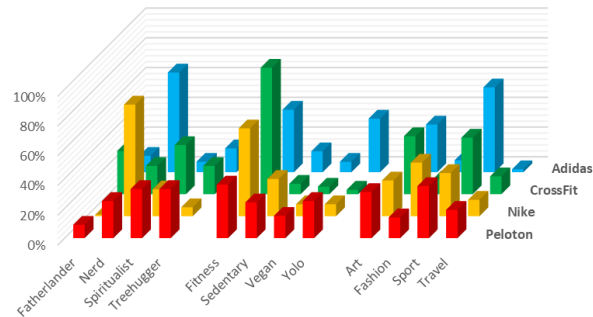


Fig. 8. Tribal affiliations of Twitter users in the virtual tribes of four *sedentary* brands.

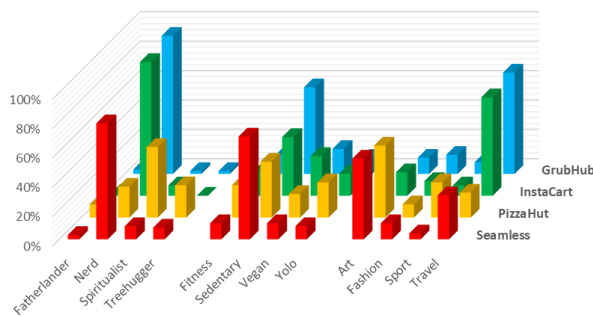


Fig. 9. Tribal affiliations of Twitter users in the virtual tribes of four *vegan* brands.

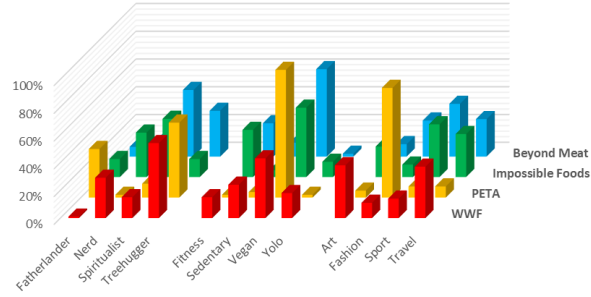


Fig. 10. Tribal affiliations of Twitter users in the virtual tribes of four *yolo* brands.

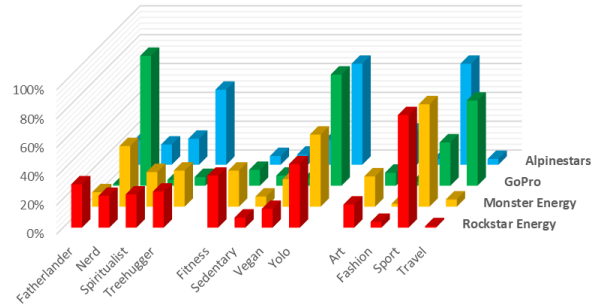
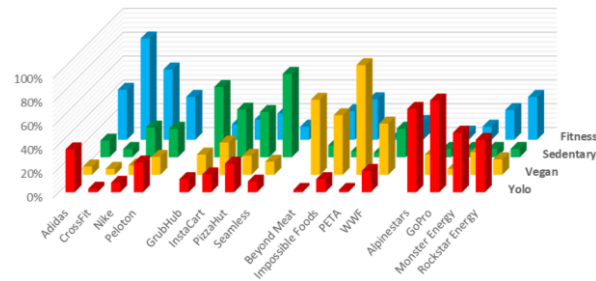


Fig. 7 shows that Twitter users in the virtual tribes of *fitness* brands properly fall into the *fitness* tribe; this is especially true for those related to the CrossFit brand. These brands also coherently attract individuals belonging to the *sport* tribe. Moreover, it is worth noting that Nike users are also *nerds* and *fashion* individuals. Regarding *sedentary* brands (Fig. 8), among the tribes in the tribal macro-category of *lifestyle*, their Twitter users are on average categorized as *sedentary*. Nevertheless, the strongest classifications emerge when looking at the tribal macro-categories of *alternative reality* and *recreation*. For instance, the great majority of users tweeting about the brands GrubHub, PizzaHut, and Seamless are *nerds*; at the same time, those interested in GrubHub and InstaCart belong to the *travel* tribe, while those related to PizzaHut and Seamless associate with the *art* tribe. The classification of the individuals tweeting on the four *vegan* brands (Fig. 9) is in line with the characteristics of these brands, and the same holds true for *yolo* brands (Fig. 10). Specifically referring to the latter, other relevant tribal affiliations emerge. For instance, GoPro Twitter users are also *nerds* and interested in *travels*, while individuals tweeting about Alpinestars, Monster Energy, and Rockstar Energy clearly fall into the *sport* tribe.

In Fig. 11 the tribal affiliations of the selected brands' tribal macro-categories are shown (i.e., *lifestyle*). Fig. 11 clearly shows that brands succeed in attracting Twitter users belonging to the tribe that best represents the brand (e.g., *fitness* individuals tweet on *fitness* brands).

Fig. 11. *Lifestyle* tribal affiliations of Twitter users belonging to *lifestyle* brands' virtual tribes.



3.3 Recreation

Finally, in this section we provide the same analyses for *recreation*-oriented tribes. In this case, we selected brands specifically targeting the following recreational activities: *art* (i.e., Guggenheim, Metropolitan Museum, Museum of Modern Art, Smithsonian), *fashion* (i.e., Chanel, Dior, Gucci, Louis Vuitton), *sport* (i.e., Broncos, Chicago Bulls, Nascar, National Football League), and *travel* (i.e., Delta, Lonely Planet, National Geographic, Southwest).

Fig. 12. Tribal affiliations of Twitter users in the virtual tribes of four *art* brands.

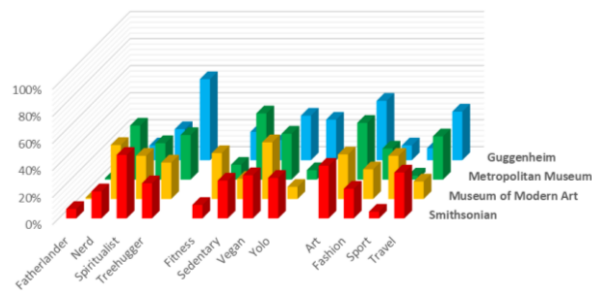


Fig. 13. Tribal affiliations of Twitter users in the virtual tribes of four *fashion* brands.

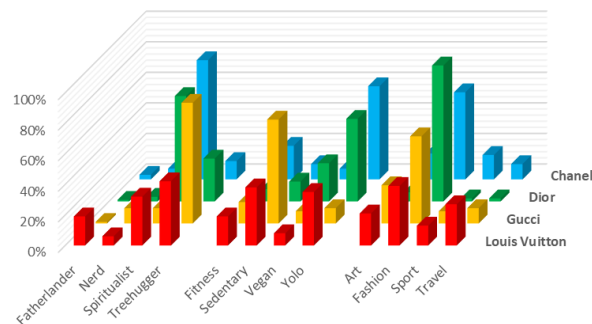


Fig. 14. Tribal affiliations of Twitter users in the virtual tribes of four *sport* brands.

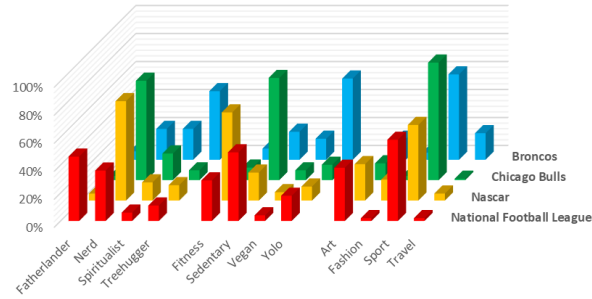
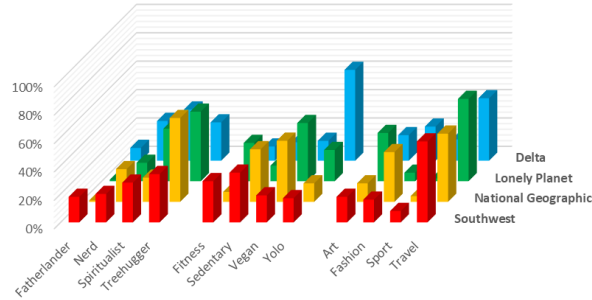
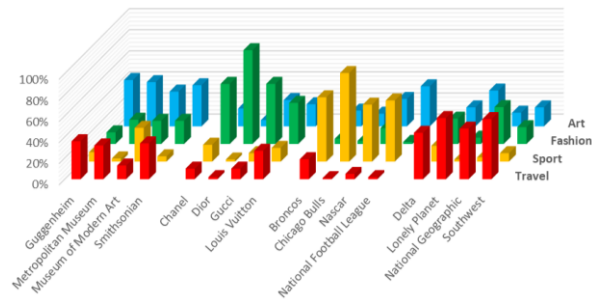


Fig. 15. Tribal affiliations of Twitter users in the virtual tribes of four *travel* brands.



Regarding *art* related brands (Fig. 12), while the majority of Twitter users appears to belong to the corresponding tribe, results are not as clear cut as for other brand categories (with the exception of the *sports* tribe that is significantly less well represented). However, this result is reasonable as the brands we choose are likely to attract *travelling* individuals, who are also interested in *fashion*. A more clear classification emerges when analyzing *fashion* (Fig. 13) and *sport* (Fig. 14) brands. The *fashion* tribe affiliation indeed predominates among individuals tweeting about *fashion* brands; a case in point is Dior, indeed 89% of individuals tweeting about Dior belong to the *fashion* tribe. The same trend exists for *sport* brands, as the dominant tribal affiliation in the macro-category of *recreation* is *sport*. In this case, also the affiliations regarding the other two tribal macro-categories seem to be reasonable; for instance, users that tweet about Chicago Bulls are mainly *sedentary* and *nerd* individuals. Also for *travel* related brands the *Tribefinder* system works well (Fig. 15) as the majority of Twitter users tweeting about these brands are classified as members of the *travel* tribe.

Fig. 16. *Recreation* tribal affiliations of Twitter users belonging to *recreation* brands' virtual tribes.



Similar to the previous tribal macro-categories, Fig. 16 shows a synthesis of the results with reference to *recreation* tribes. We find again good correspondence between the type of brand and the tribal affiliations of the individuals tweeting about the brand. The clearest results are those regarding *fashion* and *sport* brands, where the great majority of Twitter users fall into the *fashion* and *sport* tribe, respectively. The classification is somewhat less clear for *art* and *travel* related brands, this result likely depends on the brands' characteristics.

4 Further validation of *Tribefinder* results

Section 3 presented an intuitive validation of the *Tribefinder* results. To additionally verify the accuracy of our classification algorithm, two independent annotators manually assigned tribal affiliations to 500 Twitter users randomly extracted from a generic database of tweets covering different topics. The inter-rater agreement between their independent classifications, measured by means of Cohen's Kappa, was high (greater than 0.80). The two annotators then met to find an agreement on discordant cases. Their tribe allocations were subsequently matched with those produced by *Tribefinder*. The analysis of confusion matrices produced good results in terms of accuracy and Kappa statistic (see Table 2).

Table 2. Tribefinder classification accuracy.

Tribal macro-category	Classification accuracy	Kappa statistic
Alternative reality	81.2%	0.731
Lifestyle	68.8%	0.573
Recreation	69.8%	0.580

5 T-SNE visualization of tribe members

As a last step in our analysis, we present the T-SNE visualization approach, which can also be used to have an idea of the quality of the tribe classification generated by the *Tribefinder* system. T-SNE (Maaten and Hinton 2008) is a popular method for visualizing high dimensional data. In contrast to classic dimensionality reduction methods like PCA (Jolliffe 2011), which are mainly concerned with preserving large pairwise distances between the data points, T-SNE successfully captures much of the local structure of the high dimensional representation, while also giving an idea about the global structure such as the existence of clusters (Maaten and Hinton 2008). Using this technique, it is possible to visualize the members of the tribes and visually inspect the quality of their tribe assignments by verifying if the individual tribe members are clustered closely together.

We illustrate the power of our approach by adding new tribes to the ones illustrated in previous sections (shown in sections 5.4 to 5.6) and visually inspecting if they form cohesive clusters. New tribes can be created by selecting hundreds of “tribe leaders” for a topic, and computing their common vocabulary through deep learning (De Oliveira and Gloor 2018).

In our approach we selected a subset of tribes – according to the macro-categories or by individually selecting some of the tribes – for which we wanted a T-SNE visualization. We gathered all the twitter users associated with the tribes. The next step was to fetch the 200 most recent tweets of each of the gathered twitter users and tokenize the tweets content for further processing. The tokenization included getting rid of stop words, interpunctuation, URLs, unnecessary whitespace and tokens which were too short.

Having the tokens for each user and thereby also the tokens of the tribes, which is just the collection of all the tokens of its members, we calculated the tf-idf scores (Salton and McGill 1986) for the tokens for the individual users as well as for the tribes. At the same time, we calculated the unigram probability, i.e. the word probability, of all the tokens that we have encountered. The tf-idf scores can be used to define and restrict the vocabulary for further analysis: we selected from each tribe 200 distinct tokens sorted by their importance according to the tribe’s tf-idf. The constructed vocabulary defined which tokens have been then embedded to generate user vectors.

We used a pretrained fastText (Bojanowski et al. 2016) word embedding to embed the individual tokens into 300 dimensional vectors. An advantage of using fastText embeddings is its capability to use subword information. This allowed us to obtain embedding vectors for compound words, which in other cases are often not part of the embedding vocabulary. An example of such compound words are hashtags (e.g. #photooftheday), which often contain useful information about a tweet.

Using the embedding we got a collection of vectors for each user. Our goal was to represent each user with a single vector, which can be used for the visualization with T-SNE. To combine the collection of vectors into a single representative vector we tried different methods like for instance summing and weighted average (White et al. 2015) using td-idf scores.

The best results were achieved with the approach described in the work by Arora et al. (2016). The idea was to first aggregate the vectors by weighting them with their corresponding unigram probability and then summing them up. We then ended up with a single vector for each user. We stacked these vectors together into a matrix and factorized it using SVD. This gave us the eigenvectors of our user vector matrix. We then proceeded by subtracting from each user vector the first eigenvector. Intuitively we can think of subtracting the most common properties of all the users of the tribes.

After this step we used T-SNE to reduce the dimensionality of the vectors down to only two dimensions in order to visualize them in scatter plots. We then plotted

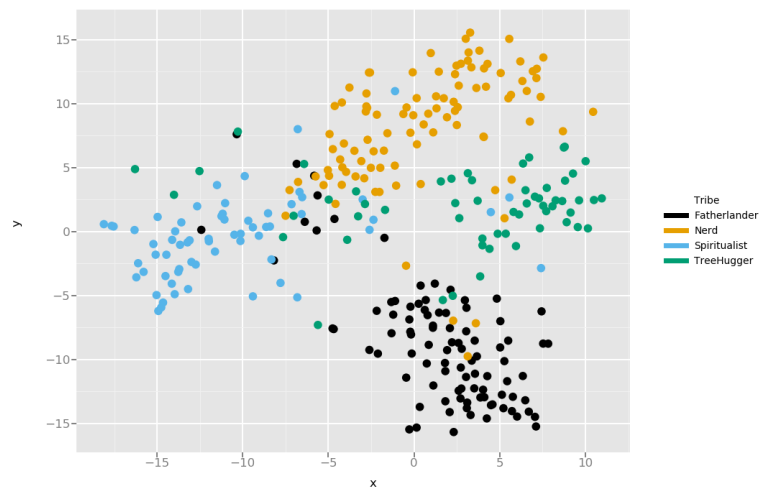
each of the two-dimensional user vectors and colored them according to their assigned tribe. It is important to note that in the steps described above, we only used the tribe information of a user at the end to assign the respective color in the scatter plot.

This gave us a visual indication if the tribe assignments of the individual users make sense, i.e. if they seem to cluster and are distinguishable from tribe members of other tribes. This visualization also allowed to identify outliers in the tribes. The information can help to improve the tribes and make them more distinct from each other.

In the next subsections we show some visualization results for the already existing macro-categories as well as for some individually defined subset of tribes.

5.1 *Alternative Reality*

Fig. 17. T-SNE Visualization for the *alternative reality* macro-category.



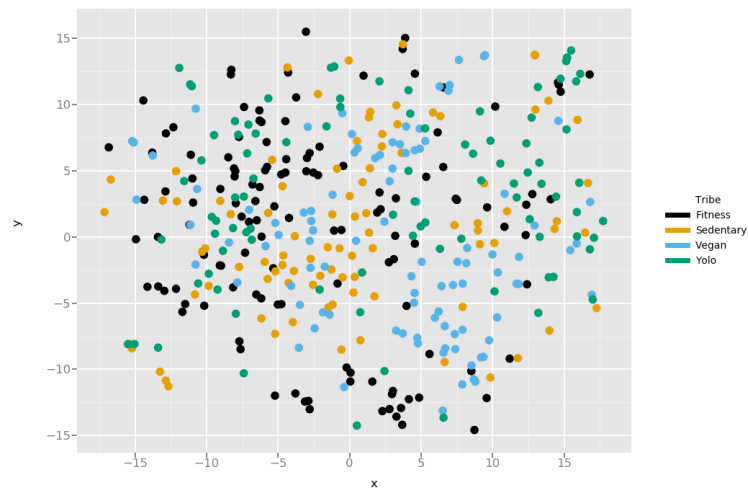
The plot shows that the four tribes cluster nicely. The cluster of the tribe *Fatherlander* seems to be a little more separate compared to the other tribes.

5.2 *Lifestyle*

In the plot below, we see that the tribes do not seem to be orthogonal to each other. Intuitively this makes sense since, for instance, there are Fitness-Youtubers, who

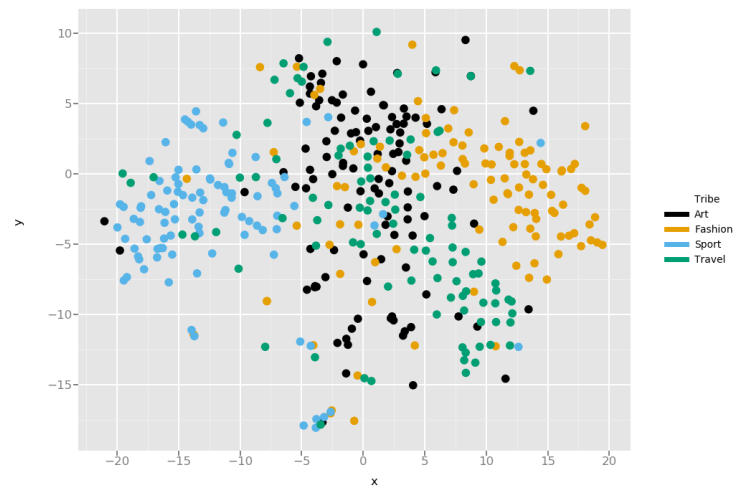
promote the vegan lifestyle and would therefore fit into the category *Fitness* as well as category *Vegan*.

Fig. 18. T-SNE Visualization for the *lifestyle* macro-category.



5.3 Recreation

Fig. 19. T-SNE Visualization for the *recreation* macro-category.



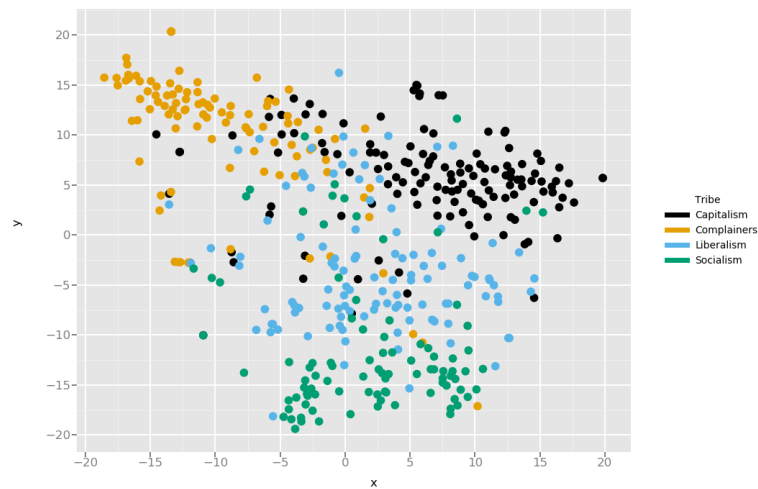
In the plot below, it is possible to easily identify clusters for the tribes *Sport*, *Travel* and *Fashion*. The tribe *Art* seems instead to be scattered among the others.

In sections 5.4 to 5.6 we illustrate the expressive power of the T-SNE algorithm with new tribes not yet included in the previous analysis.

5.4 Ideology

In this plot, the tribes *Socialism* and *Liberalism* are nicely clustered. The tribes *Capitalism* and *Complainers* both have sections where their members cluster, but in the middle of the plot there is some overlap with members of other tribes.

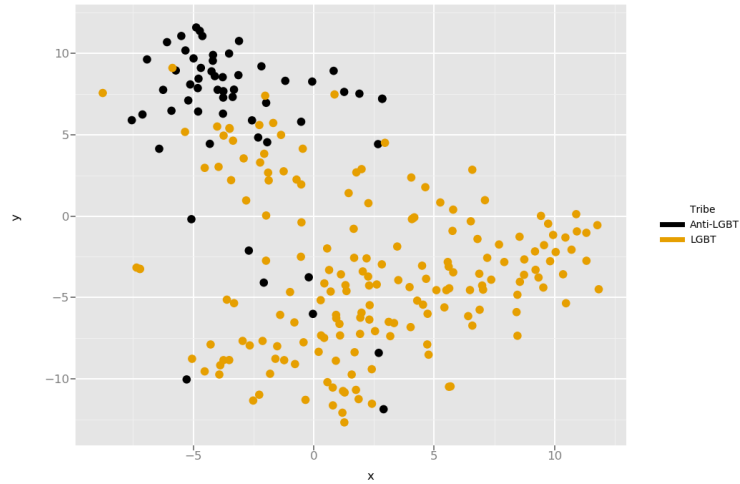
Fig. 20. T-SNE Visualization for the *ideology* macro-category.



5.5 LGBT vs Anti-LGBT

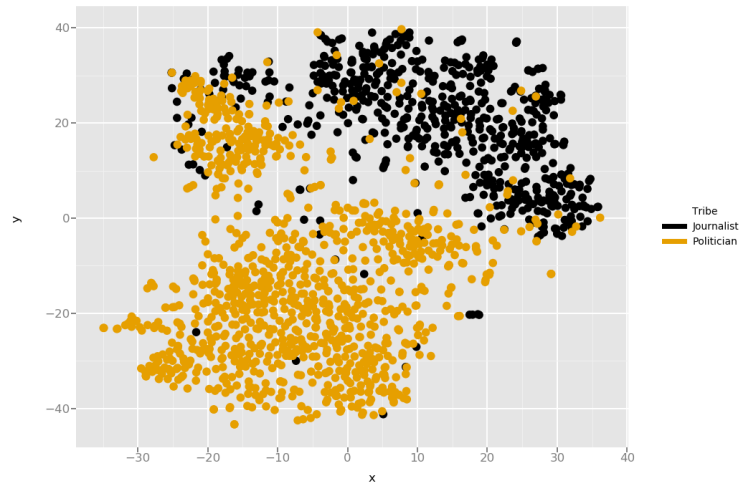
In this case, where we have two tribes opposing each other by their topic, we get for the most part a clear distinction. There are some members of both tribes, which do not seem to communicate the same way as their more aligned peers.

Fig. 21. T-SNE Visualization for the custom tribes *LGBT* vs *Anti-LGBT*.



5.6 Journalist vs Politician

Fig. 22. T-SNE Visualization for the custom tribes *Journalist* vs *Politician*.



In general, we get a nice separation between the two large tribes. On both sides there are some individual members which are in the cluster of the respective opposite tribe. For the *Politician* tribe there seem to be two clusters, the big one in the lower middle as well as the smaller one in the upper half on the left side.

6 Conclusion

In this paper we introduce *Tribefinder*, a novel system that is able to identify tribal affiliations of Twitter users. Leveraging tribal vocabularies, it analyzes an individual's words used on Twitter and categorizes her/him into tribes. We present its functionality for three specific tribal macro-categories (*alternative reality*, *lifestyle*, and *recreation*), which are taken as examples. *Tribefinder* can be easily extended to alternative tribal macro-categories depending on users' needs.

We are convinced that this system will be of value for both researchers and firms. The advent of the Internet and the diffusion of social networking platforms changed marketing paradigms (Burton and Soboleva 2011) and scholars are more and more advising firms to get rid of traditional marketing strategies (Addis and Podesta 2005; Canniford 2011), and to look for new solutions able to incorporate the essence of the tribes interested in the products or services they offer (Cova and Cova 2002; Moutinho et al. 2007). Tribe characteristics may indeed affect the success of both a marketing campaign and the firm itself (e.g., Holzweber et al. 2015). Overcoming the limits of traditional methodologies that have been used in the past to study tribes, *Tribefinder* allows scholars and practitioners to easily identify Twitter users' tribal affiliations and have a clear picture of their characteristics. The information gathered through this system thus potentially constitutes a foundation for future research - e.g. understanding how firms may rely on tribes as a strategic resource (Cova and Cova 2002) - as well as for firms to develop a better understanding of their brand's virtual tribes on Twitter, to measure the efficiency of their marketing campaigns, and to set up or adjust their marketing strategies.

Acknowledgements

We are grateful to Fabrizio Marini, for his help in data collection and tribe annotation.

References

- Addis M, Podesta S (2005) Long life to marketing research: a postmodern view
European Journal of Marketing 39:386-413
- Arora S, Liang Y, Ma T (2016) A simple but tough-to-beat baseline for sentence embeddings
- Bauman Z (1990) Thinking sociologically. Blackwell, Oxford
- Bojanowski P, Grave E, Joulin A, Mikolov T (2016) Enriching word vectors with subword information.

- Bringay S, Béchet N, Bouillot F, Poncelet P, Roche M, Teisseire M (2011) Towards an on-line analysis of tweets processing. Database and Expert Systems Applications. Springer, Heidelberg, Berlin
- Burton S, Soboleva A (2011) Interactive or reactive? Marketing with Twitter Journal of Consumer Marketing 28:491-499
- Canniford R (2011) How to manage consumer tribes Journal of Strategic Marketing 19:591-606
- Cova B (1996) What postmodernism means to marketing managers European Management Journal 14:494-499
- Cova B, Cova V (2002) Tribal marketing: The tribalisation of society and its impact on the conduct of marketing European Journal of Marketing 36:595-620
- Cova B, Pace S (2006) Brand community of convenience products: new forms of customer empowerment—the case “my Nutella The Community” European Journal of Marketing 40:1087-1105
- Cova B, White T (2010) Counter-brand and alter-brand communities: the impact of Web 2.0 on tribal marketing approaches Journal of Marketing Management 26:256-270
- De Oliveira JM, Gloor PA (2018) GalaxyScope: Finding the “Truth of Tribes” on Social Media. Collaborative Innovation Networks. Springer, Cham.
- Dionísio P, Leal C, Moutinho L (2008) Fandom affiliation and tribal behaviour: a sports marketing application Qualitative Market Research: An International Journal 11:17-39
- Garry T, Broderick AJ, Lahiffe K (2008) Tribal motivation in sponsorship and its influence on sponsor relationship development and corporate identity Journal of Marketing Management 24:959-977
- Goulding C, Shankar A, Canniford R (2013) Learning to be tribal: facilitating the formation of consumer tribes European Journal of Marketing 47:813-832
- Hamilton K, Hewer P (2010) Tribal mattering spaces: Social-networking sites, celebrity affiliations, and tribal innovations Journal of Marketing Management 26:271-289
- Holzweber M, Mattsson J, Standing C (2015) Entrepreneurial business development through building tribes Journal of Strategic Marketing 23:563-578
- Jolliffe I (2011) Principal component analysis. In: International encyclopedia of statistical science. Springer, Berlin, Heidelberg, pp 1094-1096
- Kozinets RV (1999) E-tribalized marketing?: The strategic implications of virtual communities of consumption European Journal of Marketing 17:252-264
- Maaten LVD, Hinton G (2008) Visualizing data using t-SNE Journal of Machine Learning Research 9:2579-2605
- Maffesoli M (1996) The time of the tribes. Sage, London
- Moutinho L, Dionísio P, Leal C (2007) Surf tribal behaviour: a sports marketing application Marketing Intelligence & Planning 25:668-690
- Richardson B (2013) Tribal marketing, tribal branding: An expert guide to the brand co-creation process. Springer, New York
- Salton G, McGill MJ (1986) Introduction to modern information retrieval.

- Taute HA, Sierra J (2014) Brand tribalism: an anthropological perspective *Journal of Product & Brand Management* 23:2-15
- White L, Togneri R, Liu W, Bennamoun M How well sentence embeddings capture meaning. In: 20th Australasian Document Computing Symposium, 2015. p 9