Opening Fields : A Methodological Contribution to the Identification of Heterogeneous Actors in Unbounded Relational Orders

Mohamed Benabdelkrim, Clément Levallois, Jean Savinien, Céline Robardet

M@n@gement

ORIGINAL RESEARCH ARTICLE

Opening Fields: A Methodological Contribution to the Identification of Heterogeneous Actors in Unbounded Relational Orders

Mohamed Benabdelkrim[1§], Clément Levallois[1*§], Jean Savinien[1§], and Céline Robardet[2]

[1]Quant Research Center, EM Lyon Business School, Lyon, France; [2]University of Lyon, INSA Lyon, CNRS, LIRIS UMR5205, Lyon, France

Abstract

Institutional scholarship studies how individuals coexist and interact with social structures. Organizations and inter-organizational relations within industries are a central focus of these studies. Hence, empirical research has so far largely relied on the observation of individual actors identified by their organizational attributes, and organizations identified by their industry characteristics. The flourishing of new types of social structures has sent an invitation to observe a broader range of actors beyond organizations *stricto sensu*, and to define the arena of interest beyond the boundaries of industry membership. However, in practice, these remain a favorite starting point of empirical investigations. In this article, we present a new method for the study of organizational fields that facilitates the identification of a large number and varied types of actors in a given field, provides a characterization of the relational structure of the field, and offers a content analysis on different sub-regions of the field. We test the method by replicating a previous study in the field of 'social impact of nonprofits', and show how it can contribute to operationalize mechanisms at play in the field. We conclude by noting that the principles of this method can extend beyond the dataset it is originally built on and facilitate a comparative approach to the study of fields. This contribution should enhance the value of the field as a theoretical construct by extending its operational reach.

Keywords: *Fields; Methodology; Curation; Classification; Online social networks*

Received: 18 June 2018; Accepted: 19 November 2019; Published: 30 March 2020

Organization studies experience some soul-searching regarding the central object of their investigations. The traditional focus on firm-centric organizations structured in industries is challenged by the explosion of network forms of organizations (Clegg, Josserand, Mehra, & Pitsis, 2016). In the digital age, a variety of new types of actors contribute to collective action, coordinating outside traditional organizational perimeters (McAfee & Brynjolfsson, 2017; Powell, Oberg, Korff, Oelberger, & Kloos, 2017). The concept of industry suggests a distribution of distinct roles (producers, distributors, and consumers) which have become more fluid in recent years (Furnari, 2020; Porter & Heppelmann, 2014). The relevance of organizations as a theoretical construct is put into question (Davis & Marquis, 2005), and the 'field' as an arena of heterogeneous actors (formal organizations or not) is found to offer an alternative point of theoretical focus, able to capture many types of collective actions (Zietsma, Groenewegen, Logue, & Hinings, 2017). Fields characterize unbounded local social orders: the web of relations between heterogeneous actors considering one another in their daily activities (Fligstein & McAdam, 2012; McAdam & Scott, 2005).

Regarding fields, we concur with Powell et al. (2017) that "a conceptual transition alone, however, does not suffice. We need new methods to accommodate a wider focus, which requires asking how to identify the members of nascent fields". "In order to capture the diversity and dynamism of an organizational field, the analyst must shift attention from the role of particular types of organizations to the interactions and relations among many participants" (Powell et al., 2017, p. 314). The authors offer an interesting methodological innovation in this direction; however, empirical studies in institutional scholarship

§These authors contributed equally to this work.
***Corresponding author:** Clément Levallois, Email: levallois@em-lyon.com

while embracing the field as a useful theoretical construct still tend to follow methodological procedures, which:

(1) Focus on firm-centric organizational forms, not giving its full due to the heterogeneity of actors and organizational forms that the notion of field invites to accommodate;
(2) Take industries or sectors and listings of their members as a starting point to delineate the field and identify its members. This tends to preset boundaries on the number and types of actors that will be amenable to observation;
(3) Adopt a definition of actors based on their attributes which tends to hide away the relational structure that the notion of field should have contributed to put into view.[1]

In this study, we make a methodological contribution to the empirical investigation of fields by providing a procedure that seeks to make progress on the following three fronts:

(1) The procedure identifies actors of a given field without imposing any precondition on their organizational form, industry, or sector membership (the analyst can add any of these conditions but they are not built-in).
(2) Actors identified as relating to the field are drawn from a very large pool of candidates which encompasses a much wider scope than the field itself, and no hard limit is set on the population size of the field. This removes limits set on the perimeter of a field deriving from the use of sector-based or industry-specific data sets.
(3) The identification of actors of the field under consideration is based on their relations to a couple of preselected key actors presumed to be central to the field. This puts the relational dimension of the field to the fore.

With these three features, the methodology we develop would 'open the field' by lifting some limiting conditions to the empirical investigation of local social orders.

The methodology rests on the exploitation of the informational value of individual acts of classification for identifying the actors of a field of interest. Classification schemes have long been identified as central to the structuration of fields either as acquired dispositions to differentiate and appreciate (Bourdieu, 1984 [1976]), artistic classification systems producing genres (DiMaggio, 1987), contests (Rao, 1994), and academic forms of classification such as examinations (Bourdieu, 1996 [1989]) or rankings shaping the identity of business schools and the field they form (Dubois & Walsh, 2017; Wedlin, 2007).

The 2000s have witnessed a multiplication of online platforms delivering a vast array of goods and services, and an associated proliferation of mechanisms for the classification and curation of this content. "In 2017, Netflix offered over 8,000 movie and television titles, Apple offered 2.2 million 'apps,' Amazon offered 33 million fashion-related items, Etsy offered 35 million craft-related items, and Spotify offered 30 million songs". This abundance has caused "a shift in the relative importance from those who create products to those who curate products" (Jansson & Hracs, 2018, p. 1603). Curation can be performed by intermediaries (Jansson & Hracs, 2018; Saxton & Ghosh, 2016) and by the users of the platforms themselves. These classification acts have lasting and reinforcing effects:

> Classifications are tools in strategies of inclusion and exclusion: whom to relate to and whom to isolate. They symbolize and consolidate patterns of inclusion and exclusion because they transform them into identities, which are taken for granted later on. In this perspective, classifications reinforce patterns of relations, which reinforce the classifications thereupon. (De Nooy, 2003, p. 323)

When considered in the aggregate, these individual acts of curation could amount to 'social curation': the accumulation of personal acts of curation, far from creating a cacophony of categories and diverging judgments, could reveal local orders, products of the 'collective rationality' of their constituents (DiMaggio & Powell, 1983). An emerging literature in consumer culture, media, and communication is showing that social curation is "based on the culturally shared or collective understandings (ideas, norms, and values) that give meaning to and thus regulate the activity […], [individuals engaging in curation] reflect the social and cultural context in which they perform the activity" (Villi, Moisander, & Joy, 2012, p. 492). Collectively, a series of users who create lists of their peers are akin to *social curators of local orders*. They reveal and mirror prevalent perceptions on how different actors group into distinct webs of relations. This mechanism – curation acts by observers or participants leading to the characterization of a local order – lays the foundation of the method we develop to identify actors in fields. While the mechanism is of general applicability, we developed our approach using Twitter as a prime ground of investigation because it includes such a curation device. For this reason, our methodological contribution is presented with Twitter as a prime use case.

In the rest of the article, we proceed by presenting the specificities of the Twitter data set before detailing the procedure leading to the identification of actors in a given field.

[1] See Wooten and Hoffman (2017) for a list of empirical studies on fields, the vast majority of which fits this characterization.

We then demonstrate the effectiveness of the method by replicating and expanding on a previous study. We conclude by highlighting limits and perspectives for further research.

The case of Twitter lists: Engines for social curation

Twitter is a social media platform created in 2006, which enables its 326 million users to write and publish short messages ('tweets'). Twitter is an example of 'big data' application, with an average of 500 million tweets sent per day in 2014.[2] To make it manageable for any Twitter user to read a selection of tweets in this enormous stream of publication, a basic curation device provided by the platform allows users to choose which other users to 'follow', with effect to display the tweets of these users, not the rest. Academics have relied on this feature to infer communities of users based on their follower/followee connections (see, e.g., Menichinelli, 2016) but with limited generalizability because of strict restrictions of access to this 'who follows whom' type of data imposed by Twitter.

Another device offered by Twitter to facilitate the curation of tweets is a feature called a 'list'.[3]

Lists can be created by any user on Twitter, and they are used to group other users (and the tweets they publish) in a convenient way. This feature facilitates the categorization of users in sub-topics, which makes for an easier curation of content. Lists are characterized by a name (25 characters long maximum) and an optional description chosen by the creator of the list. Once a user has created a list, it can add to it any Twitter user (up to 5,000 users per list). The consent of the user is not required to add it to the list. Lists can be private or public. While the number of private lists is not known, there are enough public lists available for a large-scale analysis: as of 2011, close to 90 million public lists could be identified (Sharma, Ghosh, Benevenuto, Ganguly, & Gummadi, 2012).

At the aggregate level, lists happen to offer an unintended service: taken together, they are akin to a curation device effectively delineating millions of Twitter users in different groups and their associated topics. A study in computer science has shown that lists reveal rich and diverse sets of highly specialized and focused topical groups, spanning a variety of niche topics, at scale. This informational value could be leveraged to identify the topics of expertise of a given individual by scanning the names of the lists it is a member of. Lists also help identify top experts for a given issue by counting the Twitter users most frequently added to lists related to the topic (Bhattacharya et al., 2014). In this study, we leverage the informational value of lists in a novel way for the purpose of identifying the actors of a field.

Data collection

The point of collecting data and storing it in a database, rather than accessing it directly from Twitter when performing the different steps of the methodology detailed below, is to speed up computation time. Access to Twitter data in large volume is conducted through Twitter's API, which is throttled – meaning that only a limited amount of Twitter data can be retrieved over a given period.[4] Fetching user profiles or list memberships 'on the fly' from the Twitter API (when the procedure is launched) would lead to running times lasting for weeks or months. Taking the preliminary step to collect user profiles, lists and list memberships allow our procedure to run in minutes, not days.

As of January 2019, 53,150,075 full user profiles were collected and stored in ElasticSearch (a database specialized in the storage, indexing and querying of textual records). At the same time, we collected 3,692,097 Twitter lists, also available through the Twitter API. We also store information on the memberships of users in these lists, using a Redis database which is very efficient at storing and querying key-values.

Methodology: A six-step procedure for the identification of actors in fields

The objective of this procedure is to identify actors of a field without imposing restrictions on the type or number of actors to be considered while putting the relational structure and contents of the field in full view. The procedure relies on the classification of Twitter users in lists by fellow users, and is summarized as follows:

(1) Pick a small number of Twitter accounts ('seeds') which should be actors with a high relevance and visibility in the field of interest.

2. https://www.prnewswire.com/news-releases/twitter-announces-third-quarter-2018-results-300737803.html https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html

3. https://help.twitter.com/en/using-twitter/twitter-lists

4. In the following, we use interchangeably 'Twitter account', 'user profile', 'user account', and 'account' for short to designate a person's or organization's account on Twitter. An Application Programmatic Interface (API) allows distant computers to connect and exchange information. Major service providers such as Twitter create APIs to facilitate and control access to their data in higher volumes, with more precise queries and at greater speed than a human could possibly download via a website. The documentation of the Twitter API is available at https://developer.twitter.com/en/docs/api-reference-index. We used Twitter4J to connect to the Twitter API. The table of 'rate limits' on different endpoints of the Twitter API is available at https://developer.twitter.com/en/docs/basics/rate-limits.html

(2) For each seed, a network of similar Twitter users is identified because they are members in the same lists.
(3) 'Denoising' of the seed networks: Removing Twitter users which tend to be less strongly connected to the seed than the average.
(4) The networks obtained from each seed are merged: The outcome is a larger network, providing a vision of the actors populating the field and its surroundings.
(5) Visual exploration: Using an algorithm developed in the field of information visualization and network analysis, the network is laid out as a map to facilitate its interpretation.
(6) Sub-regions of the field are identified based on the patterns of connectivity between actors (actors densely connected with each other form a sub-region). Content analysis is performed on each sub-region to identify the topics of interest characterizing the actors of this sub-region. The analyst decides which sub-region(s) characterize the field, and which other sub-regions are better qualified as neighbors to the field. Sub-regions deemed to capture the field of interest can then become the new focus of analysis, while surrounding regions are ignored in subsequent steps.

The analyst iterates on step 6 until all sub-regions in view are judged to be constitutive of the field of interest.

Steps 2 to 4 are performed with custom code written in Java and Python for the purpose of this study, and made available publicly (See Appendix I). Step 5 is conducted with Gephi, an open-source desktop software for graph visualization and exploration. Step 6 is conducted with Gephi (for the identification of sub-regions) and with custom code (for the content analysis), See Appendix I.

Selection of 'seed' Twitter accounts

The departing point for the identification of actors of a field is the selection of a small number of Twitter accounts which are deemed to be core participants in the field (see Powell et al., 2017 for a similar seeding procedure). The outcome of the analysis would be sensitive to the selection of seeds, since they determine the discovery of similar Twitter accounts through an iterative process (see next step). Sensitivity to seed selection is an assumed choice in the design of the method. It allows for the reproducibility of the procedure (select identical seeds to reproduce results), and preserves room for human expert judgment and exploration (select different seeds to explore the same field from a different vantage point). In any case, sensitivity is mitigated by the fact that since several seeds are selected, the associated groups of actors would scan the field of interest from many angles, leading to a global overview relatively independent from a single seed (see below for a sensitivity analysis on a given case).

Twitter accounts used as seeds can be of different sorts: individuals, organizations, brands, etc. This richness of types of actors fits well with the diversity of actors in fields, which include organizations but not only: we can remain "agnostic about whether it comprises organizations, individuals, or other combinations of actors" (Davis & Marquis, 2005, p. 337). While there is a great degree of freedom at this step, a number of guidelines should be followed:

(1) The selection of seeds should be made by a group of experts of the field of interest: Participants or observers of the field with a sufficiently broad outlook that they can arrive at a consensus on a group of core participants.
(2) The seeds should be at least moderately active on Twitter, or enjoy some notoriety so that other Twitter users would actually include them in the lists they create (from experience, a rough estimate would be that a seed should belong to at least a few dozens of lists).
(3) The seeds should be 'spread apart', meaning that they should not be too similar in their identities so that each of them relates to different kinds of actors expected to populate the field. This helps to capture the field in its diversity.
(4) The seeds should be specific to the field of interest to the largest extent possible: Since they condition the discovery of the rest of the field through their similarities with other Twitter accounts, seeds which have unfocused, multifaceted roles would lead to a discovery of a corresponding large diluted network. For instance, choosing Arnold Schwarzenegger (@Schwarzenegger) as a seed for the analysis of the US Republican Party is not judicious as he would probably not only lead to connections with Republicans but also with Hollywood actors and users with interests in bodybuilding. Similarly, while picking an organization as a seed for the analysis of an industry (say, perfumes), the Twitter account of the dedicated branch should be chosen (e.g. @CDiorParfums for Dior perfumes), rather than the Twitter account of the group (@Dior), to enhance the focus.

Discovery of Twitter accounts related to seeds

Each seed is used to identify Twitter accounts which relate to it. Two Twitter accounts are said to be related if they are registered in a number of lists in common. In other words, two Twitter accounts are connected if several third parties (the creators of lists) had a reason to register them together in the

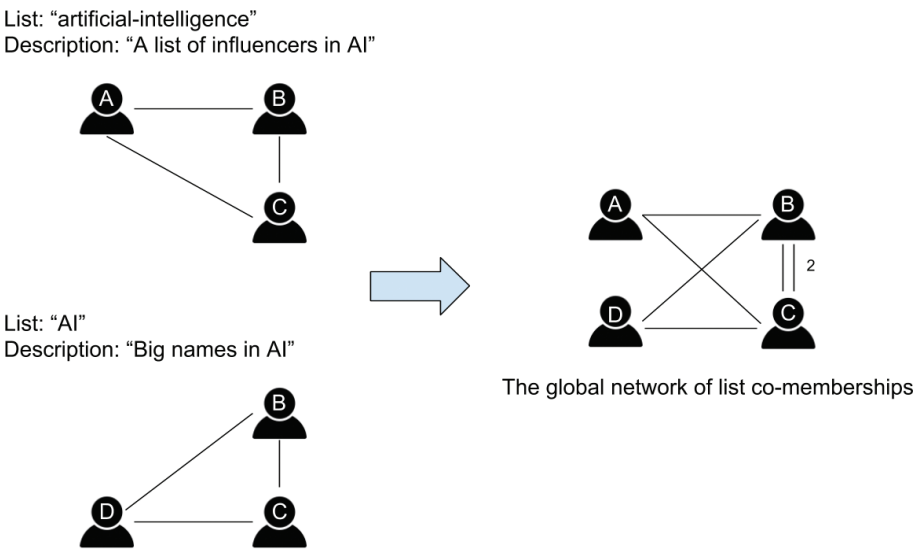

Figure 1. Using lists to discover Twitter users related to seed accounts. We consider the lists of which the seed user (here, 'B') is a member. Here, B is a member of lists 'artificial-intelligence' and 'AI'. We collect all members of these two lists (A, B, and C for the list 'artificial-intelligence' and B, C, and D for the list 'AI'). We consider that a connection exists between any two Twitter accounts if they belong to the same list. If they belong to several lists in common, then the weight ('strength') of their connection increases (if two Twitter accounts belong to two lists in common, the weight of their connection is equal to 2, etc.). Here, B and C belong to two lists in common, so the weight w of their connection is equal to 2. If we set $k = 2$ (see the main text), then only actors B and C and the link between them will be included in the field

same lists. Our methodology remains blind to the reasons leading to the inclusion of two Twitter accounts in the same list. However, if any two Twitter accounts are repeatedly picked together in lists by third parties, the odds are that they relate in some sense. We proceed in the following two steps:

(1) The lists to which the seed belongs are collected (this information is in our data set). Only lists with less than 500 members are considered, as we established through trial and error that lists with a larger membership did not have a strong specificity in relation to the topic of the list.[5] Then all the Twitter users who are members of these lists are examined. These Twitter users are considered further in the analysis if they have at least k lists in common with another Twitter account (see Figure 1), k being a parameter set by the analyst.

Through trial and error, we set a minimum threshold of $k = 3$: two Twitter accounts need to belong to at least three lists in common for a relation between them to materialize. A lower threshold would add noise (two users might be included in a couple of lists in common, without this reflecting a strong similarity of interests so that connecting them would be misleading), while a larger threshold would unnecessarily ignore meaningful connections. Also, any user with less than k lists in common with the seed user will not be included. Some seeds belong to a large number of lists, which could lead to the discovery of a very large number of users (up to millions). For this reason we cap the number of lists retrieved per user (default value: 100 lists), keeping in priority the lists the user has in common with the seed and also cap the number of users included at this step (default maximum value: 10,000 users).

(2) We repeat this procedure for each Twitter account n identified in the previous step. The lists which n belongs to are identified, and the members of these lists are collected (default capping parameters at this stage are 100 lists per user and 100,000 users). These members need to have at least k lists in common with n to be included. A relation is formed between n and each of these members, with the weight $w \geq k$ reflecting the number of lists of which they are joint members.

The result is a network with two rings of users around the seed user (see Appendix I for the code). The first ring contains Twitter users with at least k co-memberships in lists with the seed. Similarly, members of the second ring share at least k lists

[5] Lists with thousands of members tend to be created by the so-called 'bots', not humans. A bot is a computer program running autonomously, following the instructions given to it by its designer. A bot could create a list and follow the instruction to 'include in the list any Twitter account which mentions a given hashtag in their tweets'. Lists curated this way tend to include a large number of Twitter accounts with no strong meaningful relations between them.

in common with a member of the first ring. The identities of the users included in the network, and the patterns of their relations, should carry valuable information about the field from which the seed was selected, since the users are connected by relations representing at least *k* numbers of lists in common.

Denoising the seed networks

The actors identified from the seed and their relations exhibit different patterns of connectivity: some actors are more distant than others from the seed. Actors which are densely connected to the seed are more likely to be relevant to the field because dense connections mean that a relatively larger number of creators of lists have placed the seed Twitter account and these other Twitter accounts in the same lists. At this step, we aim at keeping the actors most connected to the seed in the results, and removing the actors which are less connected to the seed.[6] We simulate a large number of two-step random walks starting from the seed and weighted by the strength of the relation (the strength is the number of lists of which two users are joint members). Then we compute for each actor the number of times a random walk has passed through it. We remove those actors that belong to less paths than the average number of visits of such paths per actor. The result is a reduced group made up of actors which should have the strongest connections with the seed (see Figure 2). Each actor is characterized by an 'intensity' score measuring its distance to the seed: the previous computation is equivalent to computing the traffic of a stationary random walk from the seed (see Appendix A).

Merging the actors identified from each seed, into a single network representing the entire field

Seeds were selected as departing points for the discovery of the actors in the field and their connections. *n* seeds lead to the identification of *n* seed networks, which can now be merged into a single network representing the field of interest. Networks produced from each seed should presumably have many actors in common: since the seeds were chosen to belong to a common field of interest, their networks have a strong probability of overlapping. Hence, we expect that joining the networks of all seeds should produce a connected network (networks should not be 'islands' separated from each other). We define an 'intensity' score for each actor in the global network, corresponding to the sum of the values of the intensity scores in each seed network where this actor was present. Hence, the resulting intensity score for an actor represents a measure of its proximity to the seeds.

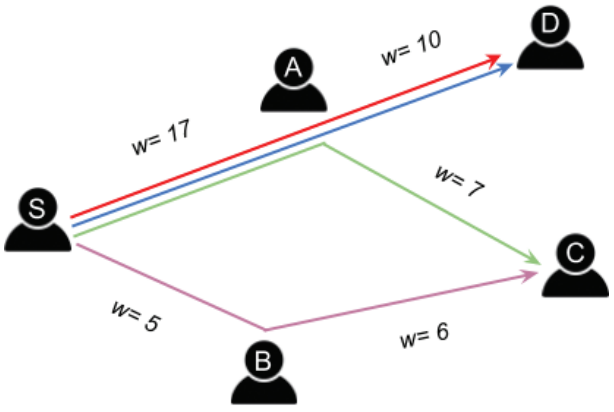

Figure 2. An illustration of the denoising process. **S** is the seed Twitter account. **A** and **B** are actors with lists in common with the seed. **C** and **D** are actors with lists in common with actors **A** and **B**. Colored straight lines are paths of length 2 ('two hops') starting from **S**. *w* is the number of lists in common between any two actors. **A** sits on three paths of length 2 from the seed. **B** belongs to one path, **C** belongs to two paths, and **D** belongs to two paths. The average number of paths is two. In the denoising process, **B** is removed as it belongs to less paths than the average. **C** and **D**, even if they are further of **S** than is **B**, are not removed in the process

Visual representation

The resulting network can take a pictorial form of representation following the principle that the visual exploration of data sets is an efficient tool for the detection of complex, unanticipated relations and patterns (Tukey, 1977), especially for large unstructured data sets (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Beyond their usefulness to direct users, visualizations travel farther while keeping compact and intact the matter they purport to report about ('immutable mobiles'; Latour, 1986; Maire & Liarte, 2018). The representation is performed with Gephi (Bastian, Heymann, & Jacomy, 2009), a software that improves significantly on the previously developed packages (such as Pajek and UCINET's NetDraw) for the visualization and exploration of large networks (see Heijmans, Heuver, Levallois, & van Lelyveld, 2016 for a comparison):

(1) The network is 'flattened' in two dimensions to be represented as a map. Actors are positioned following the logic that connected actors tend to get close to each other, while actors without a connection spread apart. The computation of the position for all pairs of actors is performed by the ForceAtlas2 algorithm implemented in Gephi (Jacomy, Venturini, Heymann, & Bastian, 2014).
(2) The size of the actors can be scaled to represent the intensity score of the actor.

[6] See Powell et al. (2017) for a similar step in their procedure, where they 'remove noise' from the list of actors they initially collect. In practice, they rely on the qualitative assessment of five members of their research team to remove irrelevant actors.

(3) The name of each actor (the Twitter account) can be displayed directly on the graph, or can be inspected by clicking on the actor.

Taken together, these visual cues provide a view on the relative relevance of the actors for the field (size of the actor), on their relation (distance between any two actors on the map), and the structure of the field (number, size, and relative positioning of sub-regions).

Identifying sub-regions in the field with network and content analysis

The steps of the procedure followed so far have resulted in a list of actors which entertain some relation to the seeds, in reason of the number of lists of which they are joint members. A fine characterization of field should, however, go beyond and explore the relational structure of the field – who relates to whom? Can the field be decomposed in sub-regions, and how are they positioned relative to one another? The identification of sub-regions consists in delineating groups of actors that tend to be relatively more connected to one another than to the rest of the actors. Interestingly, this definition of sub-regions in network analysis converges with a definition of organizational fields, which includes the condition that "participants interact more frequently and fatefully with one another than with actors outside the field" (Scott, 1995, p. 56). This step can be performed with the 'Louvain' algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), which relies on connection patterns between pairs of actors to proceed to the identification of sub-groups of actors. This algorithm leaves open the number of sub-regions to be detected (the analyst does not predetermine the final number of sub-regions the algorithm must identify). A parameter allows for tuning how selective the algorithm should be for a group of connected actors to be identified as a sub-region. We do not modify the default value of this parameter. The algorithm partitions the entire network in disjoint sub-regions, guiding the analysis on the structure of the field. The number and relative sizes of the sub-regions inform on the heterogeneity of the field: is it made up of many separate sub-regions or, at the other extreme, is it a densely connected whole? The exploration of these questions is facilitated by the visual representation of sub-regions: actors of the network (Twitter accounts) belonging to the same sub-regions can be painted with the same color, which draws the field into a patchwork of sub-regions.

The identity of each sub-region can be further explored with content analysis performed on short profile descriptions ('Twitter biographies') (255 characters or less) which users write to describe themselves on Twitter. We conduct a series of classic text-cleaning operations (such as removal of punctuation signs) and more complex operations specifically drawn from quantitative content analysis (which tends to merge with computational linguistics; Mitkov, 2005):

(1) Language detection on profile descriptions (is the Twitter profile of a user written in English, French, etc.). This improves the efficacy of text-cleaning and also offers the possibility to identify the most frequent terms *per language*, *per sub-region*, which preserves a view on the diversity of languages in a sub-region, instead of considering only the most represented ones.
(2) We detect sequences of words in user profiles in order to preserve multi-word expressions which can carry richer meaning than isolated terms.[7]

These operations prepare for a final step, which is conceptually straightforward but gives rich results: which terms appear most often in the biographies of the Twitter accounts in a given sub-region? To illustrate the usefulness of the detection of sub-regions, content analysis, and of the methodology in general, we apply it to an empirical case drawn from a published study.

Gauging the methodology: A test case

To test the merits of this method, we use it to explore the issue field of 'social impact' in the US context, studied recently by Powell et al. (2017).

Powell et al. (2017) use a website-crawler technology to draw a network of organizations active in the issue field of 'social impact' (mostly related to nonprofit organizations), which is in a state of proto-institutionalization. The authors draw a parallel between the "highly fluid system" that is an organizational field in its early stages of formation, with the "flow of ideas and concepts between disparate domains", which is the hallmark of the World Wide Web, "allowing broad and open access to multiple sources of information" (Powell et al., 2017, p. 308). The authors start by drawing a list of 36 'seeds': core participants in the field, engaged in the

[7] These sequences of words are called '*n*-grams'. We develop an example of an n-gram to illustrate the usefulness of the notion. If considering only single terms within a text, the expression 'consultant for nonprofits' would be processed and turned into 'consultant, for, nonprofits': the three terms would be considered independently from each other and the sequence of the three terms would be lost in the analysis. Considering tri-grams (n = 3) allows the text-processing algorithm to consider the frequency of sequences of terms, up to three terms: 'consultant for nonprofits' could then be counted as a single entity and its frequency in a text could be measured. Introducing *n*-grams in the analysis is not trivial to implement, as a number of language-specific rules must be introduced in the algorithm to rule out frequent but irrelevant *n*-grams to appear in the results (e.g. 'consultant for' and 'for nonprofits' in the above example).

assessment of the performance of nonprofits. Then the crawler (a computer program) collects the outgoing hyperlinks from the respective websites of these 36 participants in two iterations. After removing noise (irrelevant websites such as Google, software providers, newspapers, etc.), due to a qualitative assessment of 1,394 websites produced by the crawl, the resulting sample comprises 369 entities (with 32 unidirectional connections on average), which they categorize along their institutional forms, with nonprofit organizations being divided into subcategories (foundations, social movements, etc.). The resulting map offers a list of actors and a view on their relations. Based on the structure of linkage (outdegree, indegree, and reciprocal degree), the organizations populating the field are categorized in three roles in relation to social impact: proselytizing (large outdegree relative to indegree), convening (relatively large indegree), and strengthening (large number of reciprocal links relative to incoming links). We now proceed to mapping the same field with the methodology laid out above.

Mapping the field of social impact of nonprofits

Powell et al. (2017) identified 369 actors in the field of social impact. We start by identifying the Twitter accounts of these 369 actors to evaluate the overlap with the actors that our method will identify. Lacking access to the original data set listing these 369 actors, we referred to Figures 4a–d of the published study where the names of the actors appear, without being all legible. Through visual inspection, we could retrieve 254 names out of 369 appearing in the figures. For these 254 names, we identified by manual search the corresponding websites and the 227 corresponding Twitter accounts (e.g. the name 'sunfoundation' referred to the website https://sunlightfoundation.com/, from which we matched the Twitter account @sunfoundation). This shows that for the field of 'social impact' at least, Twitter is a communication channel used by virtually all actors. We then queried our database for these 227 Twitter accounts: 100% of them were present. This signals that the database we created, though it contains a minority of existing active Twitter accounts (estimated at 326 million accounts in late 2018[8]), is correctly focused on accounts of at least a moderate visibility.

Step 1 of our methodology for identifying the participants to a field and their relations consists in selecting seeds relevant to the field. Since the web-crawling methodology of Powell et al. (2017) is similarly based on a seeding mechanism, we would naturally reuse the 36 actors with which they had seeded their analysis. The full list of seeds is not shared in the study, except for seven seeds mentioned to illustrate the diversity of actors chosen as seeds: *3ie Impact*, *Charity Watch*, *GiveWell*, *Rockefeller*, *Gates*, *Keystone Accounting* and *Monitoring and Evaluation News* (Powell et al., 2017, p. 315). We use these seven actors as seeds.

We run steps 2 to 4 of the procedure, which produces a list of 30,575 actors and 1,098,110 connections between them: to compare with the 369 actors and close to 4,800 connections identified by Powell et al. (2017). This operation takes less than 2 h to run on a moderately powerful server (4 cores, 64 Gb of RAM). This resulted in 127,740 Twitter lists, in the sense that a list contributed if it includes two members appearing in the final list of actors. These lists were authored by 88,184 different Twitter users, 3,633 of which are present among the list of actors of the field.

Among the 227 accounts of the Powell et al. (2017) study that we could identify and match in our database, 125 (55%) are included in the map produced by our procedure. A sensitivity analysis (see Appendix B) shows that initiating the procedure with any two seeds drawn out of the seven would suffice to discover 40–50% of accounts identified by Powell et al (2017). This shows a saturation at probably 3–4 seeds: no new account is discovered by adding new seeds from a similar topic. Conversely, removing the actors discovered through any single seed does not remove any actor from the entire network, meaning that every account (actor) in the result is discovered by at least two seeds. This is an indication of the robustness of procedure to alternate selections of seeds for a given field.

Step 5 of the procedure consists in drawing a visualization of the aggregate network. In this first visualization, we color and label the seven seeds to get a sense of their placement relative to the entire network (Figure 3).

Next, plotting the 125 Twitter accounts corresponding to the actors identified by Powell et al. (2017) gives a sense of similarities between the two methods (Figure 4):

We observe that the accounts from Powell et al. (2017) lay near the center of the network and in a given region (south of Figure 4), with a few of them at the periphery. The Force Atlas layout algorithm tends to place densely connected actors near the center of the visual, but it does not follow that actors in the periphery are less relevant to the field. It might be that the core/periphery structure denotes differences in roles, for example, with generalists in the center and specialists at the periphery. We investigate these questions in step 6 of our procedure, where the detection of sub-regions and content analysis assist in understanding the inner structure and identification of the field.

Running the Louvain algorithm on the network produces 30 sub-regions. For each sub-region, we perform a content analysis to arrive at a list of the 10 most frequent terms used in the Twitter biographies of the actors in the sub-region in order to identify the key topics characterizing each sub-region (see Appendix I for a link to the code). We draw the map with

[8] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

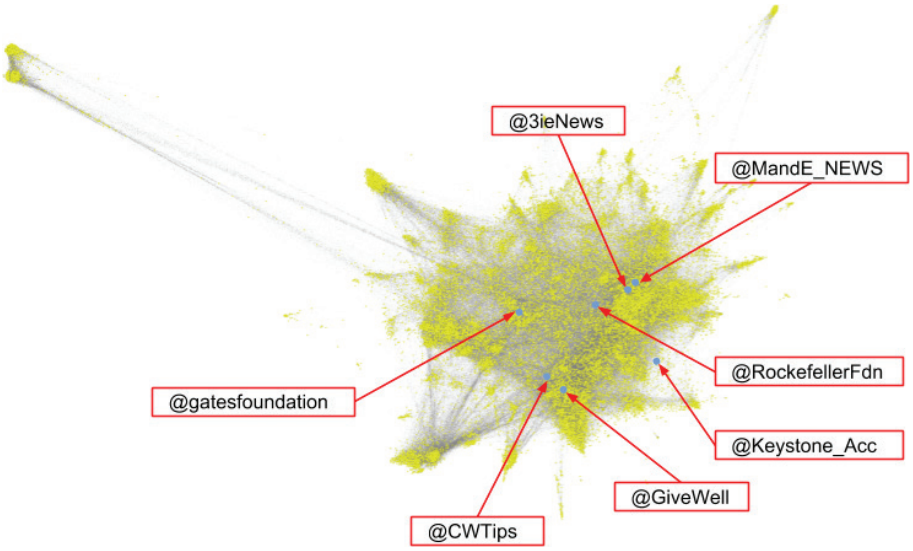

Figure 3. The 30,575 actors identified by our procedure (connected by 1,098,110 relations). These actors were identified by using seven seed accounts mentioned in Powell et al.'s (2017) study which relate to the issue field of 'social impact'

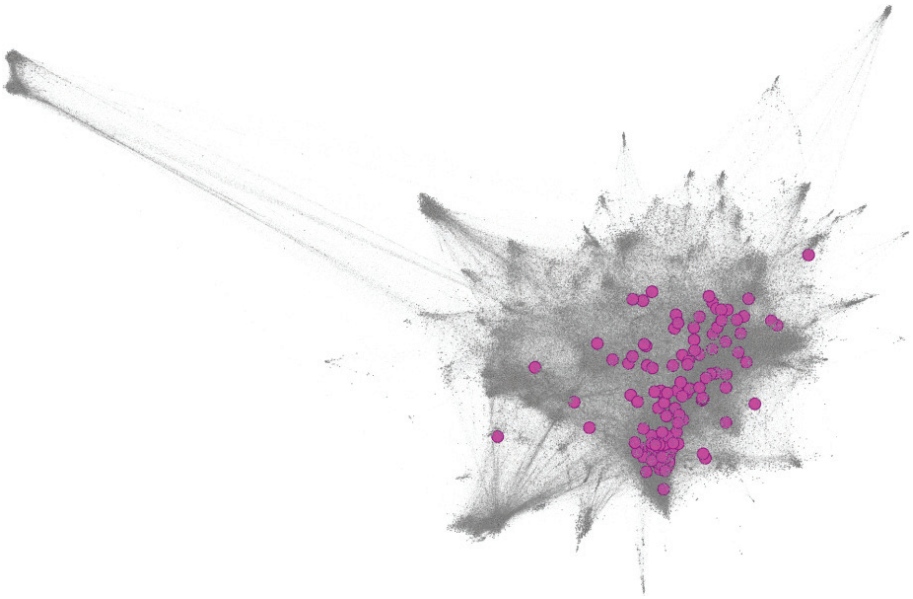

Figure 4. The 125 actors in the field of social impact retrieved from Powell et al. (2017). Colored in pink on the map

sub-regions shown in different colors, with the most relevant sub-regions and the result of their content analysis shown in Figure 5 (see Appendix C for an oversized version). A complete list of sub-regions and their 10 most frequent terms can be found in Appendix D.

This analysis invites to question the identity and boundaries of the issue field 'social impact of nonprofits'. One of the sub-regions (sub-region '27') seems directly relevant to the issue: 'Actors supporting nonprofits'. The biographies of the Twitter accounts of this sub-region refer to 'nonprofit', 'charity', 'community', 'foundation', 'philanthropy', 'change', 'support', 'fundraising', 'impact', 'uk', and 'organization'. Contrary to the rest of the sub-regions of the map, each of which focuses on the type of issue that nonprofits seek to alleviate, actors in this sub-region pursue the goal to help other organizations to achieve their goals. This resonates with the definition of an

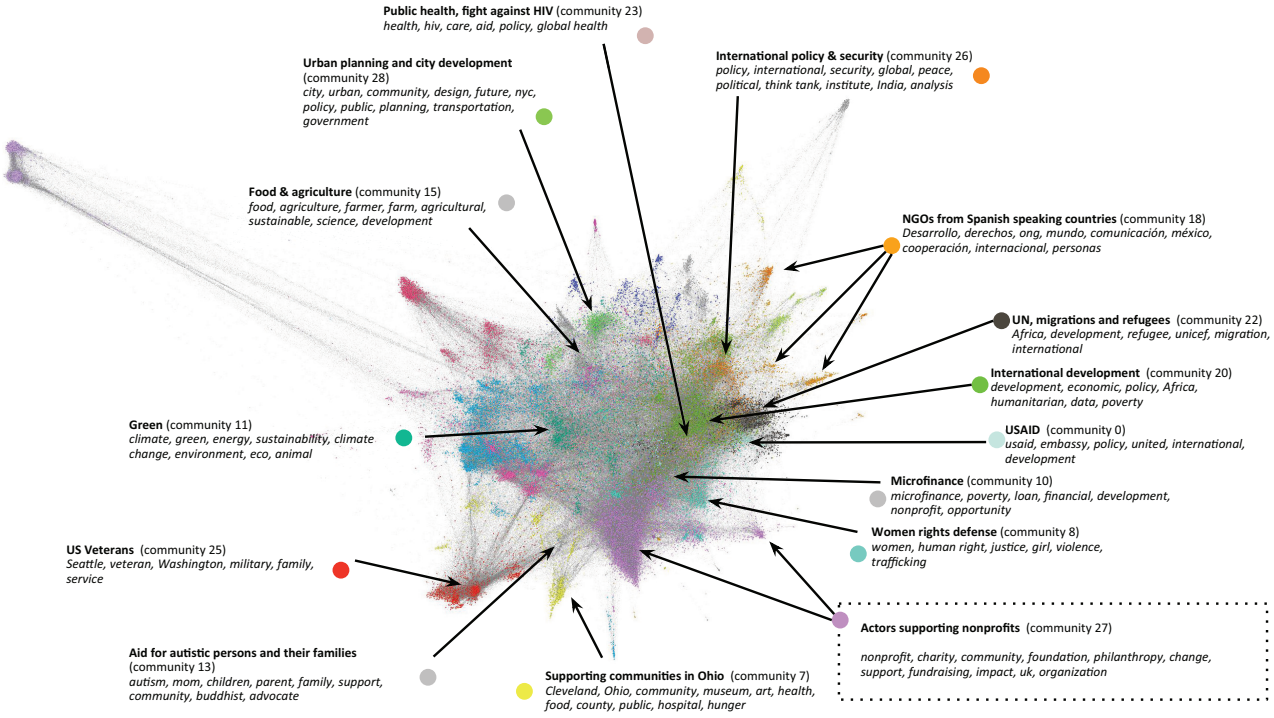

Figure 5. Sub-regions detected by the Louvain algorithm, shown in different colors. The names of the sub-regions (in bold) are inferred by the authors, from the list of the most frequent terms in the sub-region (italicized). Among the 30 sub-regions, we point to those which have a close relation to the field of interest: 'social impact of nonprofits'. Sub-region '27' is the largest sub-region (11, 74% of the actors of the entire network) and the closest to an *a priori* characterization of the field, according to the results of the content analysis. See Appendix C for a full-size version of the visual

interstitial issue field, which emerges "when an issue arises in society that people care about across several (and sometimes a broad spectrum of) social groups" (Zietsma et al., 2017, p. 401).

This sub-region should not be reified as the sole locus for actors engaged in the field of social impact. A comparison with the data from Powell et al. (2017) shows that it contains only 55 (44%) of the actors identified in their study, the others being spread in other sub-regions (see Appendix E). This suggests that a large proportion of actors engaged in supporting the performance and impact of nonprofits operate within a field of their own (their immediate neighbors being actors engaged in a similar pursuit), while the rest are surrounded by the organizations, individuals, media outlets, etc., they support. A reason for this distinction might be that this second kind of actors are concerned not only with measuring or fostering social impact but are also 'impact makers' themselves in a given area.

Examining the profile descriptions of each of the 125 actors (see Appendix F) retrieved from Powell et al. (2017) confirms this differentiation in roles, with most of the profiles included in sub-region '27' self-identifying as supporters of nonprofits and the rest being supporters *and participants* in remedying a given issue.

Sub-region '27', which focuses on 'actors supporting nonprofits', comprises 55 profiles identified by Powell et al. (2017). However, the total size of sub-region '27' is 3,589 members, which suggests that the field of 'actors supporting nonprofits' is potentially rich of different sub-groups, each with different identities and logics of action. Following the next steps of our methodology, we now iterate by focusing on sub-region '27' only, ignoring the rest of the map. We launch a new round of sub-region detection on it (with identical parameters), then we run the same routines for content analysis to characterize each of these sub-groups which altogether constitute sub-region '27' (see Figure 6).

This iteration gives a finer view on sub-region '27' and the sub-groups it is made of. In particular, it is now visible that this sub-region comprises distinct sub-groups which follow two unrelated logics. Sub-groups '1', '2', '3', and '4' gather actors according to topical or geographical dimensions, while sub-groups '0', '5', and '8' are indeed directly related to helping nonprofits achieve social impact. These sub-groups (which comprise, respectively, 538, 243, and 931 members) give a

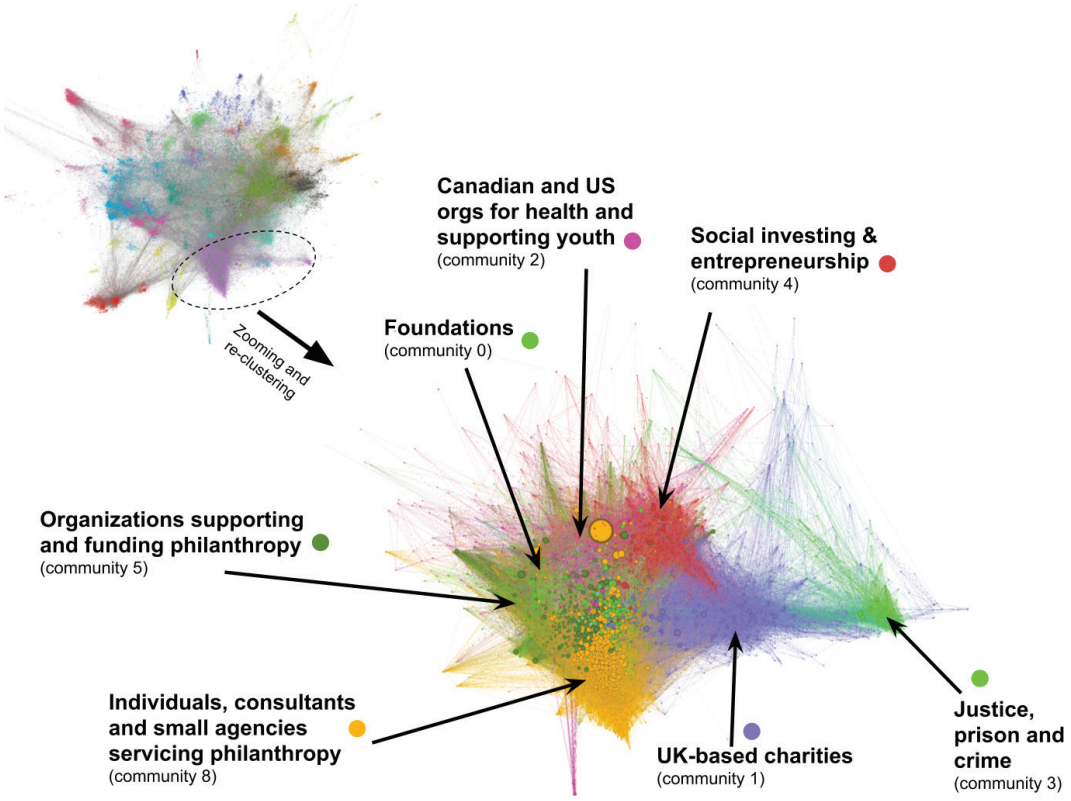

Figure 6. Using lists to discover Twitter users related to seed accounts. Iterating on step 6 of the procedure, we zoom in the sub-region 27 and ignore the rest of the map. Sub-region 27 and ignore the rest of the map. Sub-region 27 is focused on actors engaged in supporting nonprofits. The names of the sub-groups within sub-region 27 (in bold) are inferred by the authors from the list of the most frequent terms in the sub-group (full report on the size and content analysis of sub-groups is available in Appendix G)

coarse view of the following three types of actors forming the social space of 'social impact of nonprofits':

(1) Sub-group '0': a group of nonprofits comprising US donors (grant makers, foundations, funds, etc.), which is remarkably homogeneous. These actors are mostly organizations and garner a sizeable attention (average number of followers: 9,437; median: 3,602).
(2) Sub-group '5': a heterogeneous group of (mostly) for-profit organizations supplying resources and services to nonprofits. They are very visible on social media (average number of followers: 21,092; median: 5,628).
(3) Sub-group '8': a heterogeneous group of (mostly) individual consultants, media contributors, and private agencies offering their services to charities and nonprofits. This group is the largest one; however, it has a smaller footprint (average number of followers: 12,070; median: 2,457).

The map is the first sizeable contribution of the methodology presented in this study: it allows for the identification of a very large number of actors, without being bounded by pre-determined actors' types or attributes. It identifies sub-regions in the field (and the sub-groups inside them), leaving room for interpretation as to what are the boundaries of the field. It offers a view on the relations between actors, and the content analysis helps understand how actors 'cluster' around specific topics within the field.

We are well aware that such maps are several layers of abstractions away from the underlying field under investigation: Digital traces on Twitter do not reflect social interactions in the offline world, and the search for "structures should not be left to the computer program, but instead should be consistent with a theoretical view of the underlying social processes" (Fligstein & McAdam, 2012, p. 197). In the final section, we discuss the social processes which underpin the field operationalized as an arena of heterogeneous actors related through their co-memberships in lists. Before this, we meet the challenge in a different way by presenting how the map is not a flat and passive device but lends itself to the conceptualization of mechanisms of use to explore problems relevant to the field (Davis & Marquis, 2005).

Design of mechanisms to identify roles in the field

Powell et al. (2017) focus on the question of the emergence of a field. Actors vying for a position in a field in the process of coalescing might adopt different strategies, depending on their identity and the kind of role they see for themselves in the field. Conversely, these actors are not in full control of their positioning, given that other actors pursue their own strategies which impact their neighboring environment. To navigate this moving space, actors engage in the tactics of 'soft power': "organizations engage in activities that enable them to influence the development and design of new institutional arrangements" (p. 311). These actions can be characterized by the following three mechanisms (Powell et al., 2017, p. 311):

(1) *Proselytizing* of information and championing alternative visions
(2) *Convening* to create spaces for exchange among dissimilar participants
(3) *Strengthening* as a means to fund and support the adoption of new practices and attract converts.

In the empirical context of Powell et al.'s (2017) study of social impact, each of these mechanisms is operationalized in terms of the signature they leave in the network of hyperlinks between actor's websites. The methodology that we have developed, based on a network of social media accounts rather than websites, affords new ways to operationalize these mechanisms, and suggests that new mechanisms might be identified as well.

Transposing the mechanisms of *proselytizing*, *convening*, and *strengthening* to the network of actors that we traced is straightforward. While our network is undirected, contrary to the web of links in Powell et al. (2017) (incoming and outgoing links from one website to another), the Twitter accounts on the map include their number of followers and the number of accounts they follow ('friends'), which provide directionality. When these bidirectional relations are examined in combination with the sub-region structure of the network, which qualifies the social space into neighborhoods with distinct identities, we can refine these three mechanisms operationally and identify new ones, such as the following:

(1) 'Conforming': the strategy for an organization to interact with peers *from its sub-region rather than outside of it*, with effect to reinforce and stabilize the identity and boundaries of the field.
(2) 'Interfacing': the role of an organization which communicates with actors outside of its field while keeping a voice and influence in its field. The effect would be to act as a conduit for external information and influence in the field leading to its enrichment, and also to contribute to position the field relative to its neighbors. The empirical study of this mechanism would be a useful contribution to the study of the relations between fields, a domain that has remained underresearched (Furnari, 2016) (see Appendix H for the operationalization of these mechanisms).

Conclusion: Limits and perspectives

This study offers a method to identify actors of a field and their relations. It allows for a multiplicity of types of actors to be considered, and does not rely on a logic of sector or industry membership, or actor's attributes, to be considered as a potential candidate for inclusion in the field. This contribution comes with a number of limits, which we identify with possible remedies, with perspectives for further work.

Who curates the curators?

It is established that the population on Twitter is not representative of the general population (Greenwood, Perrin, & Duggan, 2016), and the sub-sample of creators of lists on Twitter is probably also specific in many ways. However, this population can be researched, as their profile and activities on Twitter are publicly available information. It is beyond our scope to conduct a full-fledged study of list creators on Twitter. We can simply note that in the test case presented, among the 88,184 list creators, which contributed to define the field with their curation work, 4% of them are included in the field itself, and they represent close to 13% of the membership of the field.

The process of constructing the view of a field depends on list curators providing the primary material for this exercise. The data on list creators suggest that it is a large and diverse group which can itself be a subject of inquiry. In an iterative manner, it would be possible to re-run the algorithmic procedure with the same seeds but factoring in only the lists created by authors previously identified as members of the field. Alternatively, the Twitter biographies and other attributes of the list creators could be examined so that only the actors from a given geographical origin, or certain types of actors (e.g. individuals vs. organizations) would get their lists used in the algorithmic procedure. These suggestions point to the fact that the method we have presented in this study is designed to facilitate and enrich the study of fields, but it can also stimulate research in the direction of the phenomenon of classification underpinning the maintenance of fields, which is itself a complex social process.

Using Twitter (and other social media) as a data set

The data source used to illustrate the methodology is Twitter data, specifically Twitter user profiles and list memberships as accessed through the Twitter public API. Twitter data, similar to other social media data, are not statistically representative of the underlying population and phenomenon of interest that they are supposed to give an access to – in the same way that model organisms could be unrepresentative of their taxa (Tufekci, 2014). More profoundly, there is an ontological gap between maps created from digital traces, and the world they attempt to represent: what happens on Twitter does not entertain a one-to-one relationship with the offline world. We acknowledge this absence of strict correspondence, but we also believe that there is no air-tight separation between 'offline fields' and 'their digital representation', at least since the rapid development of user-based content creation on the web ('Web 2.0'). Social media such as Twitter have considerably changed the nature of the relation between 'the media' and 'what the media reports about', in the direction of field actors participating more actively in the media representations of their field to the point of blurring the distinction between actors who produce in the field, and those who disseminate (Etter, Ravasi, & Colleoni, 2019; Levallois, Smidts, & Wouters, 2019). Hence, while the map is not the territory, maps remain useful abstractions to advance our understanding of fields, especially when a direct access to the constituents of a field is impractical or costly.

A number of ethical boundaries prevent a number of use cases to be developed. We identify the following two important limits:

(1) User profiling and targeting. The method can be used for profiling individuals related to any topic of interest based on a small sample of individuals with a proven involvement in the topic. Even with the consent provided by users to make their identity and messages public on Twitter, individuals might still hold a reasonable expectation not to get their implicit personal attributes discovered through data analysis. Twitter's Developer agreement states that:

> [Twitter data may not be used by] any entity for the purposes of conducting or providing surveillance, analyses or research that isolates a group of individuals or any single individual for any unlawful or discriminatory purpose or in a manner that would be inconsistent with our users' reasonable expectations of privacy.[9]

In practice, developers could access content through the API without Twitter being able to examine *ex ante* the purpose and use of the data being collected, making the enforcement of these terms relatively ineffective. In the face of the spread of methods for user profiling (Piao & Breslin, 2017), individuals can develop tactics to prevent the discovery of their implicit attributes on social media (Nechaev, Corcoglioniti, & Giuliano, 2017); however, the procedure developed here would be immune to these countermeasures as it does not rely on the features of an individual's profile but on characterizations made by third parties which the individual does not control.

(2) '*Connection is not association*'. In the procedure laid out above, connections between any two individuals are not constructed from a purposeful, mutual relation initiated by them: they are inferred. A link between two individuals is defined through a decision made by multiple third parties – creating lists and ascribing users to these lists, hence tracing a connection between these users. From this connection, it can be tempting to conclude that these individuals are associated in some ways. Actually, this assumption of an association might be considered irrelevant, false, or even harmful to the users under consideration. Our method (and any method founded on co-memberships) should be read and used with these caveats in mind.

Within these limits, the method presented in this study offers several potential venues for future research.

Research perspectives

In principle, the method can be decoupled and used independently from Twitter data. The procedure, consisting in identifying relevant connections between individuals starting from a small set of seeds, can be applied to a variety of data sources which are not necessarily bound to social networks or online sources. For instance, co-memberships in professional associations, conference co-organization, and co-attendance, or co-contributions made by individuals (such as co-purchase, co-authorships, and co-memberships in teams) provide worthy data sources feeding into the six-step algorithmic procedure outlined above. These extensions open interesting venues for research – for example, the study of 'field configuring events' could benefit from such an approach (Lampel & Meyer, 2008).

We also believe that the method laid out here is a useful building block for the empirical investigation of issue fields specifically (Hoffman, 1999). The conceptualization of issue fields opened up or revived the study of organizational fields as groups of actors engaged in interactions around a given topic,

[9] https://developer.twitter.com/en/developer-terms/agreement

rather than relying on shared affiliations or identity-based criteria for inclusion or exclusion. This definition of issue fields suggested that records of interactions pertaining to a particular issue could be helpful to identify the constituents of the field (agents engaged in these interactions are the actors of the field), and assist in drawing the structure of the field (patterns of interactions reveal the underlying organization of the field). These records can be retrieved from judiciously selected data sets, such as the plaintiffs and defendants from the Westlaw environmental law database to study environmental protection in the United States (Hoffman, 1999).

However, even when it relinquishes the focus on an industry, the empirical investigation of issue fields presents at least two difficulties: first, relying on a data set which is 'field-specific' to identify the constituency of an issue field defeats the purpose of an open-ended, exploratory empirical investigation of the field. By construction, the issue field is populated by the types of actors and roles included in the database (litigants, in Hoffman, 1999). The tight coupling between the scope of the data source and the definition of the constituency of the field questions the external validity of the results: can the field of environmental protection be anything else than traversed by legal battles between its constituents when the database used for its characterization is a record of litigations? Second, field-specific data sources are exactly this: *specific*. Studying different fields requests each time renewed efforts to identify, secure access, code, and analyze a different empirical material. This leads to slow progress in empirical research, and makes comparative analysis prohibitive.

In this study, we contributed to the empirical exploration of issue fields by removing these two obstacles. We leverage a digital data set which is 'issue agnostic': users on Twitter are free to register and pursue any kind of professional or personal types of connections and communications.[10] Hence, the constituents of an issue field identified in this data set were not 'pre-constrained' in their identities or roles by the mere fact that they are included in the data set. The data set is of an international and multilingual dimension, spanning many types of actors and registering their inter-relations. This general dimension decreases the cost of exploring a large number of fields, and should foster comparative studies. For instance, unpublished studies by the authors using this methodology include the investigation of the fields of 'Formula One' and 'Artificial Intelligence' which were conducted in short amounts of time.

Finally, this methodology can contribute to areas of research in strategic management. In particular, it could prove useful in the discovery or confirmation of stakeholders engaged with an organization, that is, "any group or individual who can affect or are affected by the achievement of an organization's goals" (Freeman, 1984, p. 46). Organizations and individuals have different claims to the status of stakeholders based on different relational attributes: power, legitimacy, and urgency (Mitchell, Agle, & Wood, 1997). Our methodology is particularly well suited to help an organization chart its environment, detailed in different sub-regions. It could then help it pursue its strategic goals with regard to gaining, maintaining, or repairing legitimacy by trying to fit or change its environment, and doing so by adapting their strategies of action and communication to each sub-region (Suchman, 1995). This would open up the perspective for this methodology to be used not only as a device for exploration and explanation of a field's logic but also as a tool for reconfiguration.

[10] Other data sets with similar characteristics exist and could be considered (e.g. Wikipedia).

Acknowledgments

The article benefited from discussions with the participants of the STORM seminar series at Emlyon Business School. Bernard Forgues provided generous advice at key junctures of the writing of this article. We acknowledge with thanks the remarks and suggestions by the editor and three anonymous reviewers, which contributed to significantly improve this article. The usual disclaimer applies.

References

Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media, 8*, 361–362.

Bhattacharya, P., Ghosh, S., Kulshrestha, J., Mondal, M., et al. (2014). Deep Twitter diving: Exploring topical groups in Microblogs at scale. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing,* 197–210.

Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. doi: 10.1088/1742-5468/2008/10/P10008

Bourdieu, P. (1984 [1976]). *Distinction.* Cambridge, MA: Harvard University Press.

Bourdieu, P. (1996 [1989]). *The state nobility: Elite schools in the field of power.* Cambridge, UK: Polity Press.

Clegg, S., Josserand, E., Mehra, A. & Pitsis, T. S. (2016). The transformative power of network dynamics: A research agenda. *Organization Studies, 37*(3), 277–291. doi: 10.1177/0170840616629047

Davis, G. F. & Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: Institutional fields and mechanisms. *Organization Science, 16*(4), 332–343. doi: 10.1287/orsc.1050.0137

De Nooy, W. (2003). Fields and networks: Correspondence analysis and social network analysis in the framework of field theory. *Poetics, 31*(5–6), 305–327. doi: 10.1016/S0304-422X(03)00035-4

DiMaggio, P. (1987). Classification in art. *American Sociological Review, 52*(4), 440–455. doi: 10.2307/2095290

DiMaggio, P. & Powell, W. W. (1983). The iron cage revisited: Collective rationality and institutional isomorphism in organizational fields. *American Sociological Review, 48*(2), 147–160. doi: 10.2307/2095101

Dubois, S. & Walsh, I. (2017). The globalization of research highlighted through the research networks of management education institutions: The case of French business schools. *M@n@gement, 20*(5), 435–462. doi: 10.3917/mana.205.0435

Etter, M., Ravasi, D. & Colleoni, E. (2019). Social media and the formation of organizational reputation. *Academy of Management Review, 44*(1), 28–52. doi: 10.5465/amr.2014.0280

Fligstein, N. & McAdam, D. (2012). *A theory of fields.* Oxford: Oxford University Press.

Freeman, E. (1984). *Strategic management: A stakeholder approach*. Boston, MA: Pittman.

Furnari, S. (2016). Institutional fields as linked arenas: Inter-field resource dependence, institutional work and institutional change. *Human Relations, 69*(3), 551–580. doi: 10.1177/0018726715605555

Furnari, S. (2020). Industry or field? The value of the field construct to study digital creative industries. In J. Strandgaard Pedersen, B. Slavich & M. Khaire, (Eds.), *Technology and creativity* (pp. 63–86). Cham: Palgrave Macmillan.

Greenwood, S., Perrin, A. & Duggan, M. (2016). *Social media update 2016.* Pew Research Center. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/11/10132827/PI_2016.11.11_Social-Media-Update_FINAL.pdf

Heijmans, R., Heuver, R., Levallois, C. & van Lelyveld, I. (2016). Dynamic visualization of large financial networks. *Journal of Network Theory in Finance, 2*(2), 57–79. doi: 10.21314/JNTF.2016.017

Hoffman, A. J. (1999). Institutional evolution and change: Environmentalism and the US chemical industry. *Academy of Management Journal, 42*(4), 351–371. doi: 10.5465/257008

Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One, 9*(6), e98679. doi: 10.1371/journal.pone.0098679

Jansson, J. & Hracs, B. J. (2018). Conceptualizing curation in the age of abundance: The case of recorded music. *Environment and Planning A: Economy and Space, 50*(8), 1602–1625. doi: 10.1177/0308518X18777497

Lampel, J. & Meyer, A. D. (2008). Guest editors' introduction. *Journal of Management Studies, 45*(6), 1025–1035. doi: 10.1111/j.1467-6486.2008.00787.x

Latour, B. (1986). Visualisation and cognition: Drawing things together. In H. Kuklick (Ed.), *Knowledge and society studies in the sociology of culture past and present* (Vol. 6, pp. 1–40). Greenwich: Jai Press.

LaValle, S., Lesser, E., Shockley, R. & Hopkins, M. S. , et al. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21.

Levallois, C., Smidts, A. & Wouters, P. (2019). The emergence of neuromarketing investigated through online public communications (2002–2008). *Business History*, 1–40.

Maire, S. & Liarte, S. (2018). Building on visuals: Taking stock and moving ahead. *M@n@gement, 21*(4), 1405–1423. doi: 10.3917/mana.214.1405

McAdam, D. & Scott, W. R. (2005). Organizations and movements. In G. F. Davis, D. McAdam, W. R. Scott & M. N. Zald (Eds.), *Social movements and organization theory* (pp. 4–40). Cambridge, UK: Cambridge University Press.

McAfee, A. & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future.* WW Norton & Company.

Menichinelli, M. (2016). Mapping the structure of the global maker laboratories community through Twitter connections. In C. Levallois, M. Marchand, M. Mata & A. Panisson (Eds.), *Twitter for research handbook* (pp. 47–62). Lyon: EM Lyon Press.

Mitchell, R. K., Agle, B. R. & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review, 22*(4), 853–886. doi: 10.5465/amr.1997.9711022105

Mitkov, R. (Ed.). (2005). *The Oxford handbook of computational linguistics.* Oxford: Oxford University Press.

Nechaev, Y., Corcoglioniti, F. & Giuliano, C. (2017). Concealing interests of passive users in social media. In *CEUR Workshop Proceedings* (*Vol. 1939*). CEUR-WS.

Piao, G. & Breslin, J. G. (2017). Inferring user interests in microblogging social networks: A survey. *User Modeling and User-Adapted Interaction, 28*(3), 277–329. doi: 10.1007/s11257-018-9207-8

Porter, M. E. & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review, 92*(11), 64–88.

Powell, W. W., Oberg, A., Korff, V. & Oelberger, C., et al. (2017). Institutional analysis in a digital era: Mechanisms and methods to understand emerging fields. In G. Krücken, C. Mazza, R.E. Meyer & P. Walgenbach (Eds.), *New themes in institutional analysis* (pp. 305–344). Northampton, UK: Edward Elgar Publishing.

Rao, H. (1994). The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895–1912. *Strategic Management Journal, 15* (Suppl. 1), 29–44. doi: 10.1002/smj.4250150904

Saxton, G. D. & Ghosh, A. (2016). Curating for engagement: Identifying the nature and impact of organizational marketing strategies on Pinterest. *First Monday, 21*(9). doi: 10.5210/fm.v21i9.6020

Scott, W. R. (1995). *Institutions and organizations: Ideas, interests, and identities.* London: Sage.

Sharma, N., Ghosh, S., Benevenuto, F. & Ganguly, N., et al. (2012). Inferring who-is-who in the Twitter social network. In *ACM SIGCOMM Computer Communication Review, 42*(4), 533–538.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. doi: 10.5465/amr.1995.9508080331

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *ICWSM, 14*, 505–514.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Villi, M., Moisander, J. & Joy, A. (2012). Social curation in consumer communities: Consumers as curators of online media content. In Z. Gürhan-Canli, C. Otnes & R. Zhu, (Eds.), *NA - Advances in Consumer Research* (Vol. 40, pp. 490–495). Duluth, MN: Association for Consumer Research.

Wedlin, L. (2007). The role of rankings in codifying a business school template: Classifications, diffusion and mediated isomorphism in organizational fields. *European Management Review, 4*(1), 24–39. doi: 10.1057/palgrave.emr.1500073

Wooten, M. & Hoffman, A. J. (2017). Organizational fields: Past, present and future. In R. Greenwood, C. Oliver, T. B. Lawrence & R. E. Meyer (Eds.), *The Sage handbook of organizational institutionalism* (2nd edn., pp. 131–147). London: Sage.

Zietsma, C., Groenewegen, P., Logue, D. M. & Hinings, C. R. (2017). Field or fields? Building the scaffolding for cumulation of research on institutional fields. *Academy of Management Annals, 11*(1), 391–450. doi: 10.5465/annals.2014.0052

Appendix A-I