

Opening fields: a methodological contribution to the identification of heterogeneous actors in unbounded, relational orders.

Keywords: fields, methodology, curation, classification, online social network.

Forthcoming in [*M@n@gement*](#), January 2020.

Authors:

Mohamed Benabdelkrim*, emlyon business school, Lyon, France. benabdelkrim@em-lyon.com

Clément Levallois* †, emlyon business school, Lyon, France. levallois@em-lyon.com

Jean Savinien*, emlyon business school, Lyon, France. savinien@em-lyon.com

Céline Robardet, University of Lyon, INSA Lyon, CNRS, LIRIS UMR5205, F-69621 France.
celine.robardet@insa-lyon.fr

* contributed equally

† corresponding author

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 paper, we present a new method for the study of organizational fields which 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 of the field of “social impact of nonprofits” (Powell, et al., 2017), 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, facilitating 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.

Opening fields: a methodological contribution to the identification of heterogeneous actors in unbounded, relational orders.

INTRODUCTION

Organization studies experience some soul searching regarding the central object of their investigation. The traditional focus on firm-centric organizations structured in industries is challenged by the explosion of network forms of organizations (Clegg, et al., 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, consumers) which have become more fluid in recent years (Porter & Heppelmann 2014; Furnari 2020). The relevance of organizations as a theoretical construct is put into question (Davis & Marquis, 2005), and the “field” as 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, et al., 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.” (314). Powell, et al. (2017) offer an interesting methodological innovation in this direction, however empirical studies in institutional scholarship, while embracing the field as a useful theoretical construct, still tend to follow methodological procedures which:

- focus on firm-centric organizational forms, not given its full due to the heterogeneity of actors and organizational forms that the notion of field invites to accommodate,
- take industries or sectors and listings of their members as a starting point to delineate the field and identify its members, which tends to preset boundaries on the number and types of actors that will be amenable to observation,
- 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 which seeks to make progress on these three fronts:

- the procedure identifies actors of a given field without imposing any pre-condition on their organizational form, industry or sector membership (the analyst can add any of these conditions, but they are not built-in).

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

- 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 datasets.
- the identification of actors of the field under consideration is based on their relations to a couple of pre-selected 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 the purpose of 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), artistic classification systems producing genres (DiMaggio, 1987), contests (Rao, 1994), academic forms of classification such as examinations (Bourdieu, 1996), or rankings shaping the identity of business schools and the field they form (Wedlin, 2007; Dubois & Walsh, 2017).

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 8000 movie and television titles, Apple offered 2.2m ‘apps’, Amazon offered 33m fashion-related items, Etsy offered 35m craft-related items and Spotify offered 30m songs”. This abundance has caused “a shift in the relative importance from those who create products to those who curate products”. (Jansson & Hracs, 2018: 1603).

Curation can be performed by intermediaries (*ibid.*, Saxton & Ghosh, 2016) and also 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: 323).

When considered in the aggregate, these individual acts of curation can amount to “social curation”: the accumulation of personal acts of curation, far from creating a cacophony of categories and diverging judgments, can 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 & Joy, 2012: 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 will be presented with Twitter as a prime use case.

In the rest of the paper, we proceed by presenting the specificities of the Twitter dataset, before detailing the procedure which leads to the identification of actors in a given field. 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 2014². 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 subtopics, 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

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>

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 public or private. 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 can 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 can 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 of time.⁴ Fetching user profiles or list memberships "on the fly" from the Twitter

⁴ In the following, we will use interchangeably "Twitter account", "user profile", "user account" and "account" for short to designate a person's or organization's account on Twitter. An API or Application Programmatic Interface

API (when the procedure is launched) would lead to running times lasting weeks or months. Taking the preliminary step to collect user profiles, lists and list memberships allows our procedure to run in minutes, not days.

In Spring 2017, we started collecting public Twitter account profiles and 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)⁵. At the same time, we collected 3,692,097 Twitter lists (also stored on ElasticSearch), 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. The Redis database is 29Gb (which is very large given that it contains only numeric ids) and the ElasticSearch database is 296Gb large.

Methodology: a 6-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 can be summarized as:

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

allows distant computers to connect and exchange information. Major service providers like 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 the different endpoints of the Twitter API is available at <https://developer.twitter.com/en/docs/basics/rate-limits.html>

⁵ The data collection process was not continuous. 2017 is indicated merely as the origin of the project. In practice, a couple of weeks of data collection suffice to harvest the data in the volumes reported here.

2. for each seed, a network of similar Twitter users is identified based on the fact that 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 visualisation 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 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 can decide 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 of the field of interest can then become the new focus of the 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 publicly available (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 code written in Java for this study (for the code used for 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 will be sensitive to the selection of the 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 will be mitigated by the fact that since several seeds are selected, the associated groups of actors will 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 is composed of organizations, individuals, or other combinations of actors” (Davis & Marquis, 2005: 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 they each relate to different kinds of actors expected to populate the field. This will help capture the field in its diversity.
4. The seeds should be specific to the field of interest, to the largest extent possible: since they will 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 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 the seeds

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

together in the same lists. Our methodology remains blind to the reasons leading to the inclusion of two Twitter accounts in a 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 two steps:

1. The lists to which the seed belongs are collected (this information is in our dataset). 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 list⁶. 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.

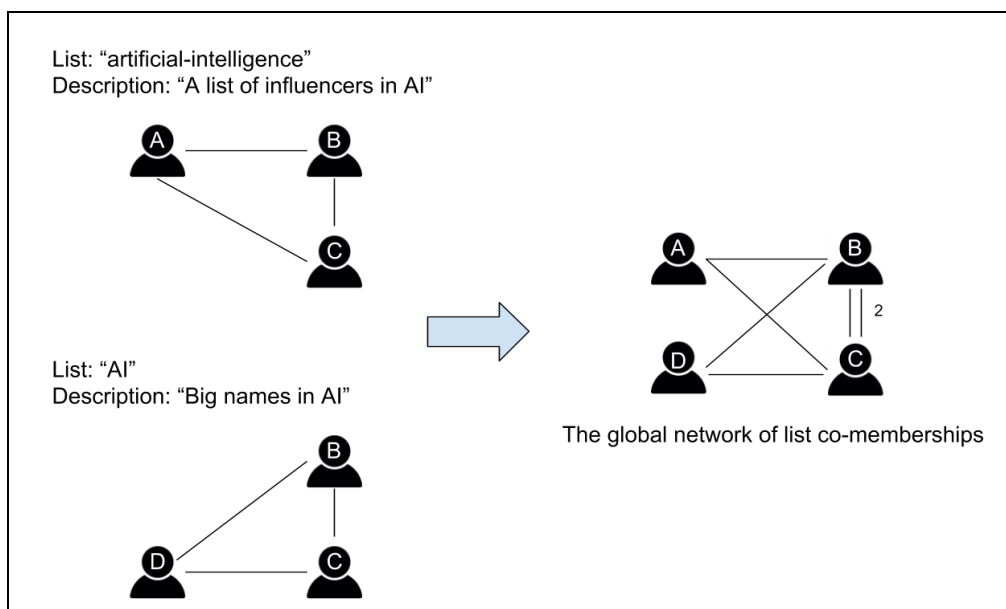


Figure 1. We consider the lists of which the seed user (here, “B”) is a member. Here, B is a member of lists “artificial-intelligence” and “AI”. We collect all members of these two lists (A,

⁶ Lists with thousands of members tend to be created by 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.

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 2 Twitter accounts belong to 2 lists in common, the weight of their connection is equal to 2, etc.). Here, B and C belong to 2 lists in common, so the weight w of their connection equals 2. If we set $k = 2$ (see main text), then only actors B and C and the link between them will be included in the field.

Through trial and error, we set a minimum threshold of $k = 3$: two Twitter accounts need to belong to at least 3 lists in common for a relation between them to be materialized. 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 can 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 a weight $w \geq k$ reflecting the number of lists they are jointly members of.

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 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 seed⁷. We simulate a large number of 2-step random walks starting from the seed and weighted by the strength of the relation (the strength is the number of lists two users are jointly members of). Then, we compute for each actor the number of times a random walk has passed through it. We remove those actors which belong to less paths than the average number of visits of such paths per actor. The result is a reduced group made 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**).

⁷ 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.

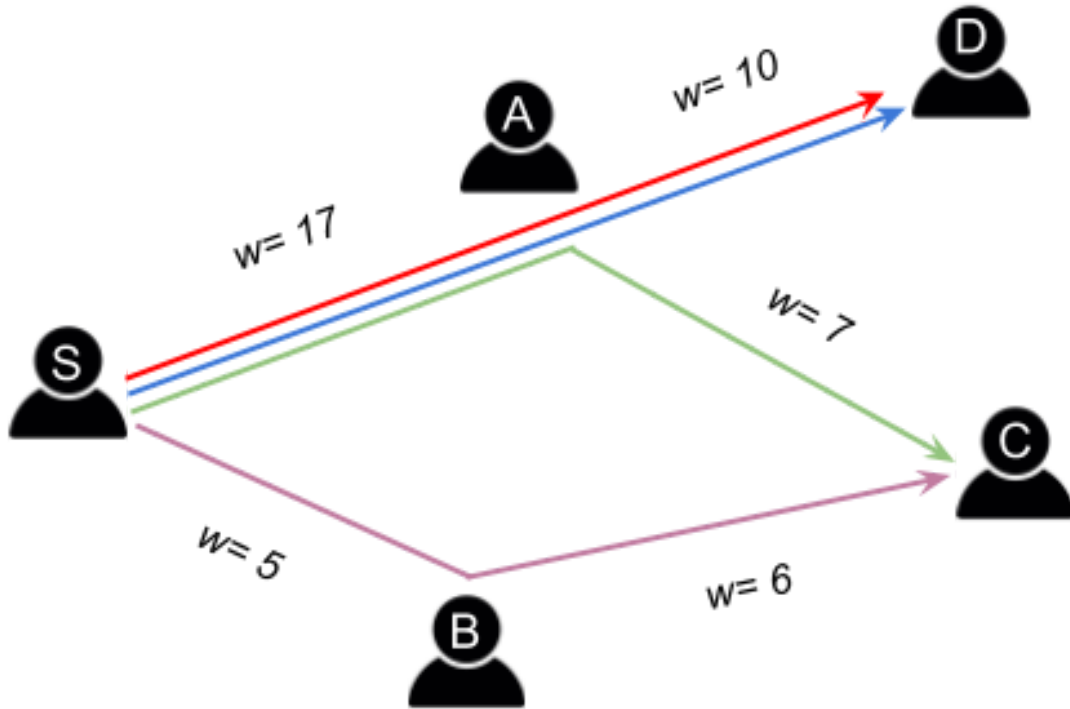


Figure 2. An illustration of the denoising process. **S** is the seed Twitter account. **A** and **B** are actors with lists in common with the seed. **C** and **D** are actors with lists in common with actors **A** and **B**. Colored straight lines are paths of length 2 (“two hops”) starting from **S**. w is the number of lists in common between any two actors. **A** sits on three paths of length two from the seed. **B** belongs to one path, **C** belongs to two paths, **D** belongs to two paths. The average number of paths is two. In the denoising process, **B** will be 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.

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 network have a strong probability to be 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.

Visual representation

The resulting network can take a pictorial form of representation, following the principle that the visual exploration of datasets is an efficient tool for the detection of complex, unanticipated relations and patterns (Tukey, 1977), especially for large unstructured datasets (LaValle, Lesser, Shockley, Hopkins & Kruschwitz, 2011). Beyond their usefulness to their 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 which improves significantly on previously developed packages (such as Pajek and UCINET’s NetDraw) for the visualisation and exploration of large networks (see Heijmans, Heuver, Levallois & van Lelyveld, 2016 for a comparison).

- The network is “flattened” in two dimensions to be represented as a map. Actors are positioned following the logic that connected actors will tend to get close to each other, while actors without a connection will 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).
- The size of the actors can be scaled to represent the intensity score of the actor.

- 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 which they are jointly members of. 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 which tend to be relatively more connected to one another, than to the rest of the actors. 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 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 the short profile descriptions (“Twitter biographies”) (255 characters or less) which users write to describe themselves on Twitter (this information was recorded in our database at the time of the collection of Twitter user profiles, see above). We conduct a series of classic text cleaning operations (such as removal of punctuation signs) and more complex operations drawn from quantitative content analysis (which tends to merge with computational linguistics) (Mitkov, 2005), specifically:

- language detection on the 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 one.
- we detect sequences of words in user profiles, in order to preserve multiword expressions which can carry richer meaning than isolated terms⁸.

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

⁸ 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, then 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 3 terms. “consultant for nonprofits” could then be counted as a single entity and its frequency in a text can 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”, “for nonprofits” in the example above).

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” (ibid.: 308). The authors start by drawing a list of 36 “seeds”: core participants in the field, engaged in the 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...) thanks to a qualitative assessment of the 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 broken down in 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) identify 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 dataset 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 of nearly 55 million Twitter profiles, though it contains a minority of existing active Twitter accounts (estimated at 326 million accounts in late 2018⁹), 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 re-use the 36 actors they seeded

⁹ <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

their analysis with. The full list of seeds is not shared in the study, except for 7 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* (ibid.: 315). We use these 7 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 two hours to run on a moderately powerful server (4 cores, 64Gb of RAM). 127,740 Twitter lists contributed to this result, in the sense that a list contributed if it includes two members which appear 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 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% to 50% of accounts identified by Powell, et al (2017). This shows a saturation at probably 3 to 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 7 seeds to get a sense of their placement relative to the entire network (see **Figure 3**).

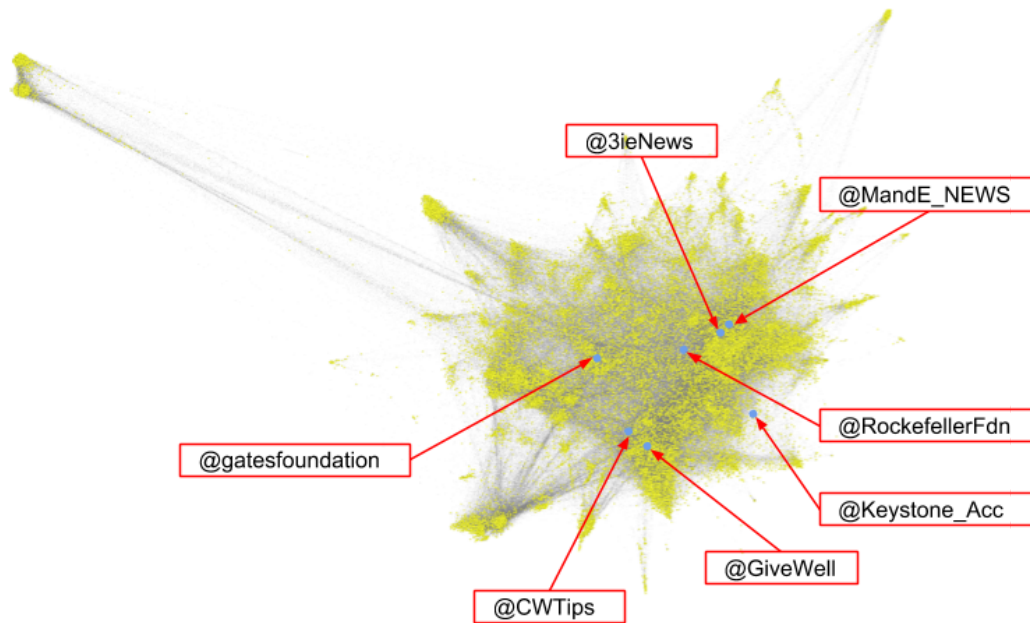


Figure 3: The 30,575 actors identified by our procedure (connected by 1,098,110 relations). These actors were identified by using 7 seed accounts mentioned in Powell, et al. (2017) which relate to the issue field of “social impact”.

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

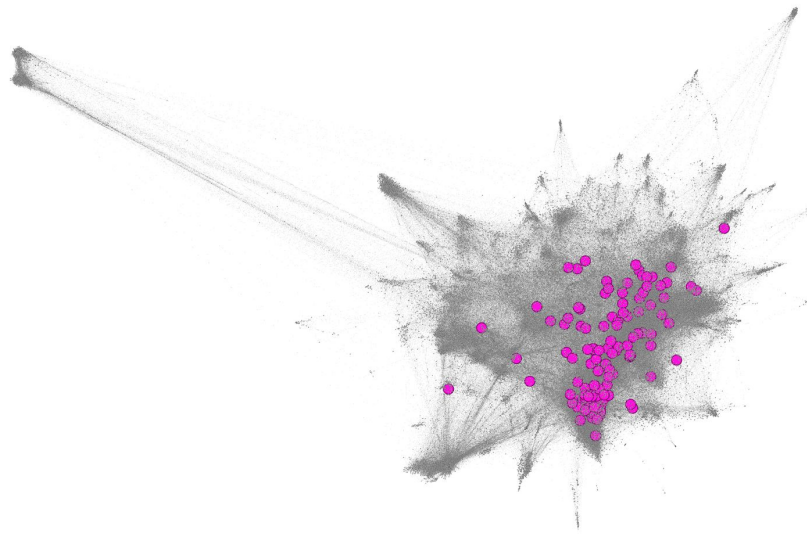


Figure 4: The 125 actors in the field of social impact retrieved from Powell, et al. (2017) are colored in pink on the map.

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 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, e.g. 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 identity 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 sub-regions shown in different colors, with the most relevant sub-regions and the result of their content analysis shown in **Figure 5**. A complete list of sub-regions and their 10 most frequent terms can be found in **Appendix D**.

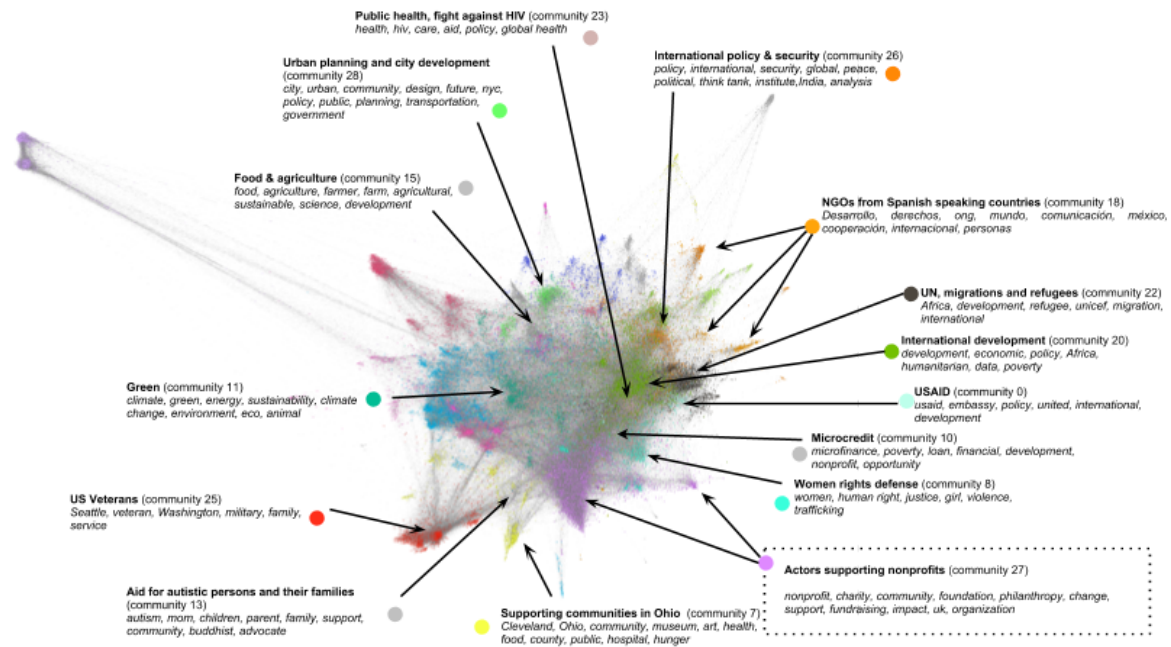


Figure 5. Sub-regions detected by the Louvain algorithm, shown in different colors. The names of the sub-regions (in bold) are inferred by the authors, from the list of most frequent terms in the sub-region (italicized). Among the 30 sub-regions, we point to those which have a close relation to the field 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.

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”) seem 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, which each focus on a type of issue which nonprofits seek to alleviate, actors in this sub-region pursue the goal to help other organizations achieve their goals. This resonates with the definition of an *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).

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... they support. A reason for this distinction might be that this second kind of actors are concerned 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 to 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**).

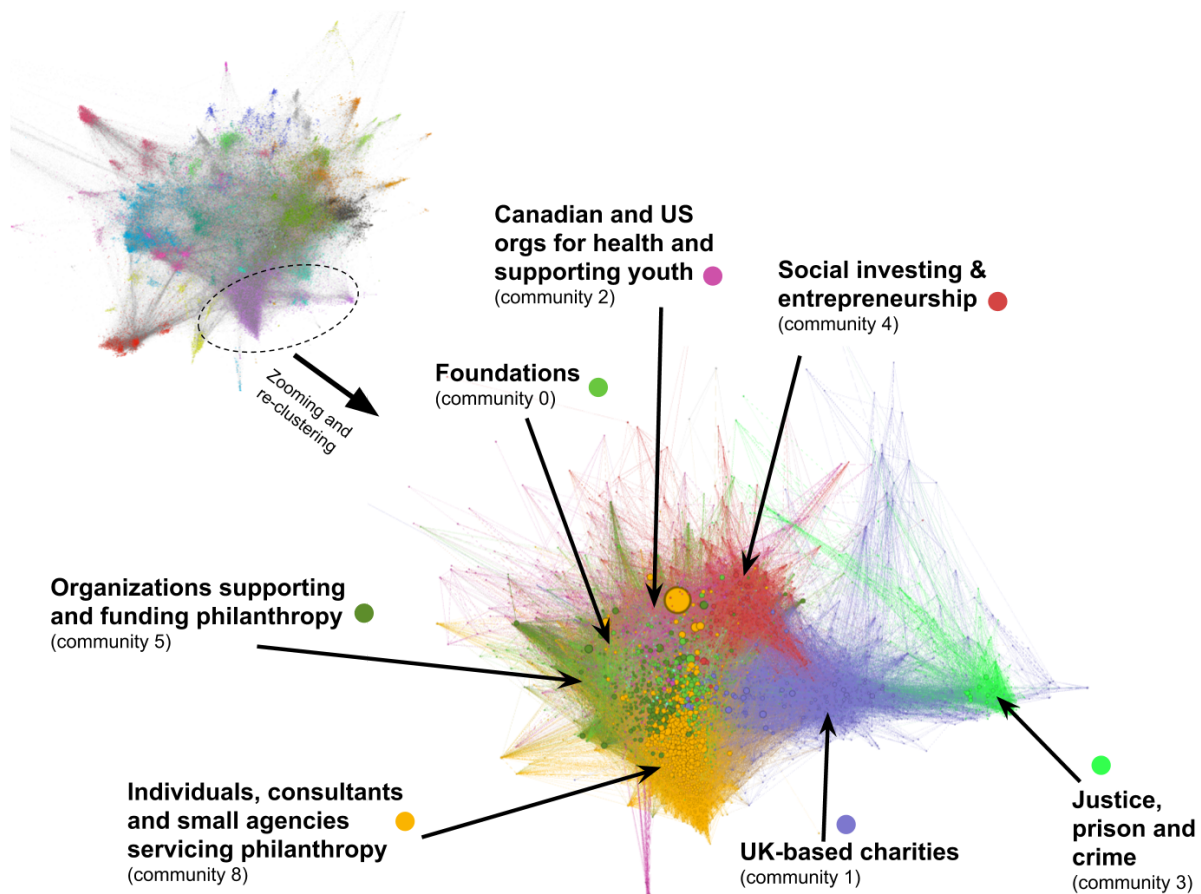


Figure 6: Iterating on step 6 of the procedure, we zoom in the sub-region 27 and ignore the rest of the map. Sub-region 27 is focused on actors engaged in supporting nonprofits. The names of the sub-groups within sub-region 27 (in bold) are inferred by the authors, from the list of most frequent terms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

This iteration gives a finer view on sub-region “27” and the sub-regions it is made of. In particular, it is now visible that this sub-region is composed of 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 coarse view of three types of actors forming the social space of “social impact of nonprofits”:

- sub-group “0”: A group nonprofits made of US donors (grantmakers, foundations, funds), which is remarkably homogeneous. These actors are mostly organizations and garner a sizeable attention (average number of followers: 9,437; median: 3,602).
- 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).
- 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, however it has a smaller footprint (average number of followers: 12,070; median: 2,457).

The map is a 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 actor’s 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 abstraction 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: 197). We will come back in the final section on 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 tactics of “soft power”: “organizations engage in activities that enable them to influence the development and design of new institutional arrangements”. These actions can be characterized by three mechanisms (ibid.: 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. (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 we 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 we traced is straightforward. While our network is undirected, contrary to the web of links in Powell, et al. (2017) (in- 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 3 mechanisms operationally and identify new ones, such as:

- "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.
- "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 which remains under researched (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 is 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. This data on list creators suggests that it is a large and diverse group which can be itself a subject of inquiry. In an iterative fashion, 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 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 dataset

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, like other social media data, is not statistically representative of the underlying population and phenomenon of interest that it is supposed to give an access to - in the same way that model organisms can 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 like 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, et al., 2019; Levallois, Smidts & Wouters, 2019). So 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 two important limits:

- 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 a consent provided by the 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.”¹⁰

In practice, developers can 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, Corcoglioni, & Giuliano, 2017), however the procedure developed here

¹⁰ <https://developer.twitter.com/en/developer-terms/agreement>

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.

- "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. While a powerful analytical device, it tends to turn a connection reconstructed between two users into a much stronger assumption of association between these individuals. This 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, co-memberships in teams) provide worthy data sources feeding into the 6-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, 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 can 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 datasets, such as the plaintiffs and defendants from the Westlaw environmental law database to study environmental protection in the U.S. (*ibid.*).

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 dataset 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 will be populated by the types of actors and roles included in the database (litigants, in *ibid.*). The tight coupling between the scope of the data source and the definition of the constituency of the field can question 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 dataset which is “issue agnostic”: users on Twitter are free to register, and pursue any kind of professional or personal types of connections and communications.¹¹ Hence, the constituents of an issue field identified in this dataset were not “pre constrained” in their identities or roles by the mere fact that they are included in the dataset. The dataset 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”, “French political scene”, and “Artificial Intelligence”, which could be conducted in short amounts of time.

Finally, this methodology can make contributions 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: 46). Organizations and individuals have different claims to the status of stakeholders, based on different relational attributes: power, legitimacy and urgency (Mitchell, et al., 1997). Our methodology is particularly well suited to help an organization chart its environment, detailed in different sub-regions. It can then help it

¹¹ Other datasets with similar characteristics exist and could be considered (e.g., Wikipedia).

pursue its strategic goals with regards 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.

Acknowledgements:

The paper 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. We acknowledge with thanks the remarks and suggestions by the Editor and three anonymous reviewers, which contributed to significantly improve the paper. The usual disclaimer applies.

Bios:

Mohamed Benabdelkrim is a doctoral student in data mining and network analysis. He is associated with the Data Mining and Machine Learning (DM2L) team in the computer science laboratory (LIRIS) at INSA Lyon. His research focuses on algorithms and methods for community detection in large temporal networks. He also works on possible applications of such methods on social networks for management and marketing purposes. Mohamed graduated in 2017 from INSA Toulouse with an engineering degree in applied mathematics and a specialization in statistics and machine learning.

Clément Levallois is Associate Professor at emlyon business school. He leverages data intensive methodologies to investigate how different communities form and interact, in science and in society at large. As a Chaired implid professor in data valuation since 2018, he conducts a research project on how to measure the value of data from an accounting and financial perspective.

Jean Savinien is Associate Professor of Data Science at emlyon business school in Lyon, France, and Associate Professor of Mathematics at the University of Lorraine, in Metz, France. His research focuses on statistical learning and network science.

Céline Robardet is a full professor at the National Institute of Applied Science in Lyon (France), member of the Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS, UMR 5205 CNRS) and coordinator of the Data Mining and Machine Learning team. She is particularly interested in: clustering analysis, pattern extraction under constraints and complex dynamic network analysis.

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., Zafar, M.B., Ganguly, N. & Gummadi; K.P. (2014). Deep Twitter Diving: Exploring Topical Groups in Microblogs at Scale. *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).
- Bourdieu, P. (1984) [1979]. *Distinction*. Cambridge, MA: Harvard University Press
- Bourdieu, P. (1996) [1989]. *The State Nobility: Elite Schools in the Field of Power*. 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.
- 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.
- 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.
- DiMaggio, P. (1987). Classification in art. *American sociological review*, 440-455.
- 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.

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.

Etter, M., Ravasi, D., & Colleoni, E. (2019). Social media and the formation of organizational reputation. *Academy of Management Review*, 44(1), 28-52.

Fligstein, N., & McAdam, D. (2012). *A Theory of Fields*. 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.

Furnari, S. (2020). Industry or field? The value of the field construct to study digital creative industries. In Pederson, J., et al., eds. *Technology and Creativity*, 63-86. Palgrave Macmillan, Cham.

Greenwood S, Perrin A, Duggan M (2016). Social media update 2016. *Pew Research Center*. Available at:

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.

Hoffman, A. J. (1999). Institutional evolution and change: Environmentalism and the US chemical industry. *Academy of management journal*, 42(4), 351-371.

- 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.
- 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*, 0308518X18777497.
- Lampel, J., & Meyer, A. D. (2008). Guest editors' introduction. *Journal of Management Studies*, 45(6), 1025-1035.
- Latour, B. (1986). *Visualisation and Cognition: Drawing Things Together* in Kuklick, H., ed. *Knowledge and Society Studies in the Sociology of Culture Past and Present*, Jai Press vol. 6, 1-40.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21.
- Maire, S., & Liarte, S. (2018). Building on visuals: Taking stock and moving ahead. *M@n@gement*, 21(4), 1405-1423.
- Levallois, C., Smidts, A., & Wouters, P. (2019). The emergence of neuromarketing investigated through online public communications (2002-2008). *Business History* (online), 1-40.
- Menichinelli, M. (2016). *Mapping the Structure of the Global Maker Laboratories Community through Twitter Connections*. In Levallois C., et al., eds. *Twitter for Research Handbook*, 47-62. Lyon: Emlyon Press.
- McAdam, D., & Scott, W. R. (2005). *Organizations and Movements*. In *Social Movements and Organization Theory*, 4-40. Cambridge, U.K.: Cambridge University Press.

- McAfee, A., & Brynjolfsson, E. (2017). *Machine, Platform, Crowd: Harnessing our Digital Future*. WW Norton & Company.
- 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.
- Mitkov, R. (Ed.). (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.
- Nechaev, Y., Corcoglioniti, F., & Giuliano, C. (2017). Concealing interests of passive users in social media, *CEUR WORKSHOP PROCEEDINGS* (Vol. 1939). CEUR-WS.
- Piao, G., & Breslin, J. G. (2017). Inferring user Interests in microblogging social networks: A survey. *arXiv preprint arXiv:1712.07691*.
- 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., & Kloos, K. (2017). Institutional analysis in a digital era: Mechanisms and methods to understand emerging fields. In Krücken G., et al., eds. *New Themes in Institutional Analysis*, 305-344. 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(S1), 29-44.
- Saxton, G. D., & Ghosh, A. (2016). Curating for engagement: Identifying the nature and impact of organizational marketing strategies on Pinterest. *First Monday*, 21(9).

- Scott, W. R. (1995). *Institutions and Organizations: Ideas, Interests, and Identities*. London: Sage.
- Sharma, N., Ghosh, S., Benevenuto, F., Ganguly, N. & Gummadi, K. P. (2012). Inferring who-is-who in the Twitter social network. *Proceedings of the 2012 ACM workshop on online social networks*, 55-60.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610.
- 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 : Addison-Wesley.
- Villi, M., Moisander, J., & Joy, A. (2012). Social curation in consumer communities: Consumers as curators of online media content. *ACR North American Advances*.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (Vol. 8). Cambridge University Press.
- 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.
- Wooten, M., & Hoffman, A. J. (2017). Organizational fields: Past, present and future. In Greenwood R. et al., eds. *The Sage handbook of organizational institutionalism* (2nd edition), 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.

Appendix A

Let $G=(V,E)$ be a graph, with a set of vertices V , and a set of edges E . The adjacency matrix A of the graph G is the matrix whose (i,j) -th entry is equal to the number of edges in E linking vertex i to vertex j in V . In the Twitter graph presented in this paper, the edges are undirected, thus A is a symmetric matrix ($A_{ij} = A_{ji}$). It follows that the k -th power of A has (i,j) -th entry equal to the number of paths of length k linking vertex i and vertex j . And the number of paths of length two from a reference vertex s (the seed) and visiting a vertex i can be computed as:

$$NV_s(i) = A_{si}^2 + A_{si} \sum_j A_{ij}$$

This number of paths is interpreted as the number of visits of a stationary random walker, traveling two steps from the seed s . (In the above formula, the first term corresponds to the number of times the random walk ends at i , the second term equals the number of times the random walk passes through i and goes further to a neighbour j of i .)

To denoise the network, we select those vertices around a seed s which are visited by the random walk more than the average number of visits of the random walk per vertex (the mean \overline{NV}_s of $NV_s(i)$ in the above formula). We then normalise the resulting number of such visits in units of standard deviations (of visits) σ_s :

$$nv_s(i) = \max \left(\frac{NV_s(i) - \overline{NV}_s}{\sigma_s}, 0 \right)$$

In other words, the vertices of the seed network associated with seed s are those vertices i for which $nv_s(i)$ is non-negative.

In this paper, the network of a field of interest is build from the seed networks of several seeds.

The *intensity score* $IS(i)$ of a vertex i in a field of interest, is the sum of the normalised number of visits of random walks from each of the seeds:

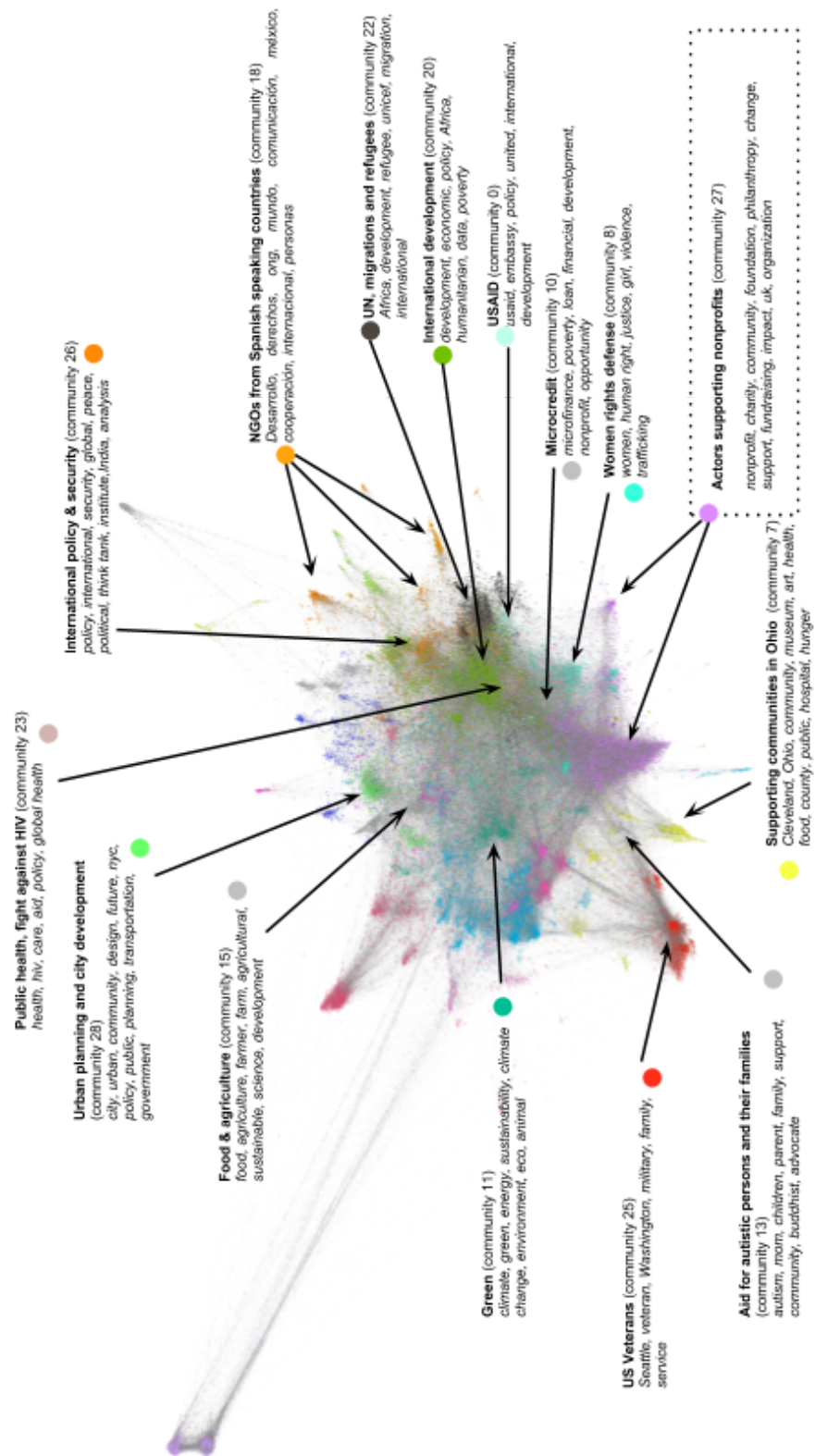
$$IS(i) = \sum_s nv_s(i)$$

Appendix B - sensitivity analysis.

	rockefellerfdn	MandE_NEWS	gatesfoundation	GiveWell	Keystone_Acc	CWTips
3ieNews	13458 (44%) - 109 (48%)	9251 (30.25%) - 91 (40%)	15995 (52%) - 101 (44%)	11754 (38%) - 114 (50%)	9787 (32%) - 113 (50%)	11655 (38%) - 110 (48%)
rockefellerfdn		13282 (43%) - 107 (47%)	17576 (57%) - 105 (46%)	13452 (44%) - 115 (51%)	12276 (40%) - 113 (50%)	13593 (44%) - 111 (49%)
MandE_NEWS			15438 (50%) - 92 (41%)	11042 (36%) - 112 (49%)	8514 (28%) - 101 (44%)	10570 (35%) - 97 (43%)
gatesfoundation				14961 (49%) - 102 (45%)	14298 (47%) - 93 (41%)	14876 (49%) - 86 (38%)
GiveWell					9475 (31%) - 106 (47%)	10428 (34%) - 102 (45%)
Keystone_Acc						8672 (28%) - 84 (37%)
<p>How to read this table. When @rockefellerfdn and @gatesfoundation are the only two seeds used to cover the field of "social impact", it results in a network made of 17,576 actors, which is 57% of the actors in the network that is obtained when the 7 seeds are used. It also discovers 105 of the 225 accounts (46%) of the Powell, et al. (2017) study.</p>						

Appendix C.

Figure 5 in oversized version.



Appendix D: List of the 28 sub-regions identified in the network for the field of “social impact”, which comprised more than 10 members (sub-regions “5”, “6”, “12”, “17” were left out as they did not meet this threshold). For each sub-region, the 10 most frequent terms used in the profiles of the members of each sub-region are listed. In the rare case when a sub-region comprises several linguistic sub-regions (most often, English and a second one), the two languages are detected as such and the most frequent terms of each are treated separately.

sub-region 0 members: 590 % of total: 1,93	Language: en usaid x 55, embassy x 52, policy x 41, united x 33, international x 31, office x 25, privacy policy x 24, democracy x 24, america x 22, affair x 22, development x 21	
sub-region 1 members: 521 % of total: 1,70	Language: en national x 55, resistance x 42, unofficial x 41, reporter x 33, editor x 32, political x 28, team x 27, politics x 27, host x 27, fact x 25, post x 22	
sub-region 2 members: 737 % of total: 2,41	Language: pt oficial x 47, contato x 34, tv x 26, apresentadora x 25, atriz x 23, jornalista x 23, vida x 23, br x 21, instagram x 19, brasil x 17, facebook x 16	Language: en tv x 8, producer x 7, instagram x 7, actress x 6, da x 5, gmail com x 5, gmail x 5, music x 4, atriz x 4, actor x 4, director x 4
sub-region 3 members: 1481 % of total: 4,84	Language: en science x 458, space x 250, nasa x 110, stem x 101, education x 92, open x 90, astronomy x 75, editor x 72, writer x 69, engineering x 63, scientist x 63	
sub-region 4	Language: en ministry x 52, republic x 48, president x 30,	Language: es economista x 38, chile

members: 566 % of total: 1,85	minister x 26, government x 16, prime minister x 16, office x 13, english x 12, mfa x 9, latest x 7, international x 7	x 35, república x 34, colombia x 25, economía x 25, diputado x 23, presidente x 23, ministerio x 19, desarrollo x 17, director x 15, nacional x 15
sub-region 7 members: 1412 % of total: 4,62	Language: en cleveland x 178, ohio x 160, community x 128, museum x 111, art x 109, health x 100, food x 86, county x 81, public x 80, hospital x 76, hunger x 74	
sub-region 8 members: 1421 % of total: 4,65	Language: en women x 202, human x 141, human right x 110, community x 95, justice x 79, girl x 79, national x 70, organization x 68, global x 63, violence x 55, trafficking x 51	
sub-region 9 members: 828 % of total: 2,71	Language: en education x 227, teacher x 145, school x 128, learning x 115, educator x 90, student x 69, teaching x 48, policy x 46, learner x 43, public x 43, national x 41	
sub-region 10 members: 190 % of total: 0,62	Language: en microfinance x 51, financial x 35, poverty x 25, community x 16, organization x 16, loan x 15, global x 15, financial inclusion x 14, development x 14, nonprofit x 14, opportunity x 13	
sub-region 11 members: 1468 % of total: 4,80	Language: en climate x 221, green x 181, energy x 132, sustainability x 128, sustainable x 105, climate change x 101, environment x 95, global x 90, environmental x 90, eco x 71, animal x 70	

sub-region 12 members: 1006 % of total: 3,29	Language: en food x 137, health x 106, fashion x 45, chef x 40, healthy x 39, writer x 39, recipe x 34, live x 33, fitness x 31, online x 30, lover x 28	
sub-region 13 members: 261 % of total: 0,85	Language: en autism x 115, mom x 31, with autism x 28, children x 25, parent x 22, family x 19, support x 18, community x 16, buddhist x 16, autistic x 16, advocate x 16	
sub-region 14 members: 948 % of total: 3,10	Language: en financial x 88, industry x 82, global x 77, leading x 75, finance x 64, service x 64, company x 52, bank x 51, investment x 47, management x 46, maritime x 45	
sub-region 15 members: 503 % of total: 1,65	Language: en food x 156, agriculture x 103, farmer x 75, farm x 62, agricultural x 51, ag x 48, global x 46, sustainable x 37, science x 35, development x 33, community x 32	
sub-region 16 members: 3222 % of total: 10,54	Language: en speaker x 139, ceo x 115, director x 108, design x 101, writer x 100, community x 97, consultant x 95, editor x 86, book x 82, web x 81, coach x 79	
sub-region 18 members: 749 % of total: 2,45	Language: en car x 17, toronto x 12, event x 9, team x 8, canada x 8, owner x 8, entertainment x 7, music x 7, service x 7, online x 7, local x 7	Language: es desarrollo x 37, derecho x 35, ong x 31, mundo x 26, i x 25, comunicación x 25, méxico x 22, cooperación x 21, internacional x 20, personas x 20, los derecho x 19

sub-region 19 members: 181 % of total: 0,59	Language: en irish x 44, view x 35, journalist x 30, ireland x 29, rte x 24, editor x 23, rté x 23, presenter x 20, radio x 17, correspondent x 17, dublin x 15	
sub-region 20 members: 3448 % of total: 11,28	Language: en development x 514, global x 334, economic x 309, international x 268, policy x 236, africa x 206, humanitarian x 200, data x 194, director x 182, view x 170, poverty x 169	
sub-region 21 members: 476 % of total: 1,56	Language: en sociology x 75, philosophy x 58, professor x 52, anthropology x 52, university x 49, book x 41, journal x 35, anthropologist x 33, sociologist x 33, science x 31, study x 29	
sub-region 22 members: 1559 % of total: 5,10	Language: en africa x 174, development x 117, refugee x 116, water x 111, human x 85, unicef x 79, office x 78, global x 77, view x 74, migration x 73, international x 72	
sub-region 23 members: 1454 % of total: 4,76	Language: en health x 611, global x 208, hiv x 167, care x 139, aid x 135, policy x 134, global health x 115, community x 98, healthcare x 96, service x 91, public health x 89	
sub-region 24 members: 130 % of total: 0,43	Language: en girl x 6, demi lovato x 5, justin x 5, fun x 5, ♥ x 5, music x 4, hey x 4, jonas x 4, miley x 3, profit x 3, non profit x 3	
sub-region 25	Language: en seattle x 260, veteran x 109, washington x 94, military x 79, family x 63, service x 60,	

members: 1068 % of total: 3,49	community x 51, reporter x 43, sport x 42, washington state x 41, event x 39	
sub-region 26 members: 1514 % of total: 4,95	Language: en policy x 174, international x 147, security x 140, global x 115, peace x 101, political x 99, tank x 96, think tank x 94, institute x 88, india x 84, analysis x 82	
sub-region 27 members: 3589 % of total: 11,74	Language: en nonprofit x 502, charity x 379, community x 374, foundation x 258, philanthropy x 234, change x 218, support x 206, fundraising x 199, impact x 195, uk x 193, organization x 172	
sub-region 28 members: 1023 % of total: 3,35	Language: en city x 259, urban x 145, community x 83, design x 81, future x 75, nyc x 73, policy x 69, public x 63, planning x 57, transportation x 50, government x 50	
sub-region 29 members: 212 % of total: 0,69	Language: en breaking x 15, analysis x 12, latest x 11, team x 10, global x 9, story x 9, source x 8, politics x 8, ceo x 7, editor x 6, source for x 6	

Appendix E

Spread of the 127 actors from Powell, et al. (2017) across the sub-regions identified in the network.

sub-region 4: 0,80%	sub-region 11: 4,00%	sub-region 23: 5,60%
sub-region 7: 1,60%	sub-region 12: 0,80%	sub-region 26: 6,40%
sub-region 8: 1,60%	sub-region 16: 0,80%	sub-region 27: 44,00%
sub-region 9: 2,40%	sub-region 20: 25,60%	sub-region 28: 1,60%
sub-region 10: 3,20%	sub-region 22: 1,60%	

Appendix F

Users from Powell et al. **inside** sub-region “27” (twitter handle followed by the first 80 characters of the Twitter profile description).

@newprofit: Break through with New Profit. <https://t.co/mRBW5FCIRb>[...]

@NatlCouncilNPs: The National Council of Nonprofits is a trusted resource and advocate for Americ[...]

@Give2Asia: Give2Asia builds connections between donors and local charitable groups in Asia [...]

@PointsofLight: Points of Light inspires, equips and mobilizes people to take action that change[...]

@FordFoundation: Around the world, we make grants that support the visionary people and organizat[...]

@GreatNonprofits: Stories That Make a Difference[...]

@phijo: Your online source for nonprofit news, resources, jobs, and events. PJ is a prog[...]

@fdncenter: Foundation Center is the leading source of information about philanthropy worldw[...]

@DRKFoundation: We invest in early stage, high impact social entrepreneurs who are changing the [...]

@Alliancemag: The leading global source of independent and informed discussion of philanthropy[...]

@macfound: John D. & Catherine T. MacArthur Foundation supports creative people and effecti[...]

@NCVO: National Council for Voluntary Organisations - largest membership body for Engli[...]

@MooreFound: We foster path-breaking scientific discovery, environmental conservation, patien[...]

@ThirdSector: The UK's leading publication for everyone who needs to know what's going on in t[...]

@ImpactSP2: The Center for High Impact Philanthropy provides smart, actionable guidance to d[...]

@aapip: We are a member-supported national advocacy organization committed to building d[...]

@TheComNetwork: The Communications Network Foundations & nonprofits that communicate well are [...]

@PackardFdn: We work with partners around the world to improve the lives of children, familie[...]

@BlendedValue: Impact Investing and Entrepreneurship Thought Leader; Founding Director, REDF an[...]

@gpforg: Global Philanthropy Forum (GPF) aims to build a community of donors & social inv[...]

@peakgrantmaking: Member driven network leading innovation in grantmaking and effective practices [...]

@vpppartners: Venture Philanthropy Partners leads change to build a better future for vulnerab[...]

@pewtrusts: Driven by knowledge to solve today's challenging problems, improve public policy[...]

@CEPData: The Center for Effective Philanthropy focuses on the development of comparative [...]

@kresgefdn: We're working to expand opportunities through grantmaking and social investing. [...]

@MntrInstitute: Mining the leading edge, challenging assumptions, and redefining expectations, w[...]

@BridgespanGroup: A global nonprofit organization collaborating with mission-driven leaders, organ[...]

@HeronFdn: Investing in enterprise as if people matter, challenging the rules of philanthr[...]

@IrvineFdn: Expanding opportunity for the people of California.[...]

@npquarterly: Journalism for the nonprofit sector. Tweets on nonprofit trends, news, democrati[...]

@atlantic: The Atlantic Philanthropies is a limited life foundation dedicated to bringing a[...]

@MyForefront: Illinois' statewide association of nonprofits, foundations + advisors. Forefront[...]

@hiltonfound: We provide funds to #nonprofit organizations working to improve the lives of dis[...]

@philaction: Philanthropy Action: A Journal for Donors[...]

@PhilanthropyNY: We ARE the professional association of grantmakers in the NYC region. Please not[...]

@RGK_Center: Located in @TheLBJSchool, we turn students and practitioners into changemakers w[...]

@TidesCommunity: Accelerating the pace of social change. Amplifying our community's efforts to ac[...]

@siliconvalleycf: Silicon Valley community Foundation makes all forms of philanthropy more powerfu[...]

@Deborah909: Nonprofit technology strategist. I live to bring resources and needs together.[...]

@SOCAPmarkets: SOCAP connects global innovators who are dedicated to increasing the flow of cap[...]

@arabellaadvisor: We help individual, family, institutional & corporate philanthropists & investor[...]

@RockBrosFund: The Rockefeller Brothers Fund advances social change that contributes to a more [...]

@GivingUSA: The Giving USA Foundation publishes data and trends about charitable giving thro[...]

@gkccf: Greater Kansas City community Foundation | Greater Horizons: Simplify your givin[...]

@knightfdn: Knight Foundation supports transformational ideas that promote quality journalis[...]

@Socalgrantmaker: Southern California Grantmakers (SCG) is the leadership hub where 300 of our reg[...]

@svtgroup: Impact measurement, management & valuation mavens. When the world sees your soci[...]

@MeyerFoundation: Meyer supports organizations working to meet local community needs throughout th[...]

@kaboom: We're the national nonprofit dedicated to bringing balanced and active play into[...]

@Surdna_Fdn: The Surdna Foundation seeks to foster just and sustainable communities in the Un[...]

@nff_news: We make millions of dollars in loans to nonprofits and push for fundamental impr[...]

@CWphilanthropy: A philanthropy firm that knows management, and a management firm that knows phil[...]

@colsfoundation: We are the trusted philanthropic advisor to more than 2,500 individuals, familie[...]

@CharityNav: Empowering Donors to Make Informed Giving Decisions[...]

@bostonfdn: Greater Boston's community foundation: a partner in philanthropy, a major funder[...]

Users from Powell, et al. (2017) **outside** of sub-region "27" (twitter handle followed by the first 80 characters of the Twitter profile description).

@UNICEF: UNICEF promotes the rights and wellbeing of every child in 190 countries and ter[...]

@LuminaFound: Lumina is a private foundation committed to increasing the # of Americans w/high[...]

@anticorruption: We fight #corruption around the world with a movement spanning 100+ countries. F[...]

@commonwealthfnd: Affordable, quality health care. For everyone.[...]

@UNStats: Official Twitter account of the Statistics Division of the United Nations Depart[...]

@Keystone_Acc: Keystone helps #socialchange organisations improve their #impact and #performanc[...]

@gmfus: GMF strengthens transatlantic cooperation on regional & global challenges and op[...]

@UNAIDS: UNAIDS leads and inspires the world to achieve zero new #HIV infections, #zerodi[...]

@PathfinderInt: We champion sexual and reproductive health and rights worldwide.[...]

@MDRC_News: MDRC, a nonprofit, nonpartisan organization, develops and evaluates innovative e[...]

@NewAmerica: Think tank and civic enterprise renewing America in the digital age[...]

@McKinsey: Since 1926, the trusted advisor to the world's leading businesses, governments &[...]

@AEI: Cherish freedom? The power of enterprise? Opportunity for all? It's these core b[...]

@RANDCorporation: RAND helps improve policy and decisionmaking through research and analysis. RAND[...]

@UNESCO: Building peace where it starts - in the minds of men & women. Follow our Directo[...]

@CGDev: The Center for Global Development: Independent research + practical ideas for gl[...]

@FairTradeCert: Your Purchase Matters. Quality products, improving lives, and protecting the en[...]

@gatesfoundation: We work to help all people lead healthy, productive lives, focused on health, po[...]

@Refugees: The official account of #UNHCR. Follow us as we provide vital aid and protection[...]

@DalbergTweet: Official feed of Dalberg Global Development Advisors, a strategic advisory firm [...]

@PSIimpact: PSI is a global non-profit organization dedicated to improving the health of peo[...]

@MathPolResearch: Research for the public good, guided by the highest standards of quality, rigor [...]

@CanadaDev: Global Affairs Canada's account dedicated to international development. FR: @Dev[...]

@MicroCredSummit: From 1997-2016, the Microcredit Summit Campaign regularly convened leaders in #m[...]

@amprog: Dedicated to improving the lives of Americans through ideas and action. Tweets b[...]

@the_IDB: At the Inter-American Development Bank we work to improve lives in Latin America[...]

@FINCA: FINCA's mission is to alleviate poverty through lasting solutions that help peop[...]

@UNHABITAT: UN-Habitat, is the United Nations agency for human settlements. Improving the qu[...]

@GrameenFdn: Innovating for the world's poor. #microfinance #financialinclusion #socent #mhea[...]

@CeresNews: Sustainability nonprofit organization working with the most influential investor[...]

@MandE_NEWS: A news service focusing on developments in monitoring and evaluation methods rel[...]

@Deloitte: Sharing the latest news, research, events and more from Deloitte Touche Tohmatsu[...]

@WorldBank: The official World Bank Twitter feed. The World Bank's mission is to fight pover[...]

@BrookingsInst: Independent research and analysis on the most important policy issues in the wor[...]

@echoinggreen: 1987–2017: Celebrating the Power of 30 Years Supporting Social Innovation[...]

@CWTips: [...]

@AdvocatesTweets: Rights. Respect. Responsibility. Following an account does not indicate philoso[...]

@USAID: USAID works to end extreme global poverty and enable resilient, democratic socie[...]

@policylink: PolicyLink is a national research and action institute advancing economic and so[...]

@charitywater: We're bringing clean and safe drinking water to people in need around the world.[...]

@ClintonFdn: The Bill, Hillary & Chelsea Clinton Foundation transforms lives and communities [...]

@aeaweb: Evaluation Esoterica from the American Evaluation Association[...]

@cochranecollab: Cochrane: Trusted evidence. Informed decisions. Better health.[...]

@Ashoka: Ashoka is an international citizen-sector organization that is leading the way t[...]

@NextBillion: Exploring the connection between development and enterprise through analysis, ne[...]

@KauffmanFDN: Fostering economic independence by advancing education & entrepreneurship. RTs # [...]

@MSHHealthImpact: Management Sciences for Health helps leaders, health managers & communities buil [...]

@RnfrstAlliance: International nonprofit organization working to conserve biodiversity and ensure [...]

@unfoundation: Connecting people, ideas and resources with the United Nations. #2030Now [...]

@Accion: We're a global nonprofit dedicated to building a world where everyone has access [...]

@CarnegieEndow: The Carnegie Endowment for International Peace is the oldest international affai [...]

@_AfricanUnion: For an integrated, prosperous and peaceful Africa. Official page of the AU. [...]

@Asia_Foundation: Nonprofit international development organization committed to improving lives ac [...]

@mix_market: MIX is the premier source for financial inclusion data and analytics. [...]

@PATHtweets: PATH is an international nonprofit organization that transforms global health th [...]

@SusanGKomen: A force united by a promise to end breast cancer, funding groundbreaking researc [...]

@350: Join a global movement that's inspiring the world to rise to the challenge of th [...]

@GlobalFundWomen: Global Fund for Women is a champion for gender equality, standing up for women a [...]

@CIVICUSalliance: CIVICUS is a global network of civil society organisations and activists working [...]

@UNDP: UNDP helps empower lives & build resilient nations. Follow @ASteiner @PNUD @PNUD [...]

@MSF: Médecins Sans Frontières/MSF is an international, independent, medical humanitar [...]

@EndOvershoot: Global Footprint Network is a research organization changing how the world manag [...]

@dfat: Official account of Australia's Department of Foreign Affairs & Trade. Follow @s [...]

@UN_Women: UN Women is the #UN entity for #genderequality & women's empowerment. Executive [...]

@SkollFoundation: Driving large-scale change by investing in, connecting, & celebrating social ent [...]

@ADB_HQ: Guided by a vision of an Asia-Pacific region free from poverty, Asian Developmen [...]

@GRI_Secretariat: GRI is an international independent organization that has pioneered #sustyreport [...]

@AAMers: #AAM2017 #AAMSMJ #museuminclusion American Alliance of Museums. Champion Museums [...]

@Accenture: Follow us for updates on Accenture research, blogs, podcasts and more. Tweets by [...]

@OneWorldTrust: The One World Trust is a UK-based charity that conducts research, develops reco [...]

Appendix G

Size and relative share of sub-groups detected in sub-region 27.

sub-group 0: 538(14,99%)	sub-group 3: 224(6,24%)	sub-group 6: 5(0,14%)
sub-group 1: 757(21,09%)	sub-group 4: 474(13,21%)	sub-group 7: 52(1,45%)
sub-group 2: 365(10,17%)	sub-group 5: 243(6,77%)	sub-group 8: 931(25,94%)

For each sub-group with more than 10 members, the 10 most frequent terms used in the profiles of the members of each sub-region.

sub-group 0 members: 538 % of total: 14,99	Language: en foundation x 150, community x 119, philanthropy x 117, nonprofit x 88, change x 35, community foundation x 33, organization x 31, support x 31, network x 28, donor x 26, sector x 26
sub-group 1 members: 757 % of total: 21,09	Language: en charity x 251, uk x 142, support x 84, community x 59, children x 57, sector x 53, live x 50, family x 47, leading x 44, national x 41, change x 38
sub-group 2 members: 365 % of total: 10,17	Language: en community x 53, service x 49, cancer x 37, organization x 36, live x 30, nonprofit x 28, national x 27, foundation x 24, support x 24, poverty x 20, volunteer x 20
sub-group 3 members: 224	Language: en justice x 59, prison x 39, charity x 30, policy x 28, crime x 27, offender x 23, criminal justice x 22, uk x 22, support x 22, community x 20, family x 19

% of total: 6,24	
sub-group 4 members: 474 % of total: 13,21	Language: en impact x 90, change x 47, global x 42, community x 39, investing x 38, investment x 35, socent x 34, impinv x 34, capital x 27, ceo x 26, entrepreneurship x 25
sub-group 5 members: 243 % of total: 6,77	Language: en nonprofit x 49, community x 20, philanthropy x 16, online x 14, charity x 13, organization x 12, profit x 12, foundation x 11, support x 11, change x 11, non profit x 10
sub-group 7 members: 52 % of total: 1,45	Language: en impact x 8, global x 8, community x 5, engage x 4, csr x 4, nonprofit x 4, company x 4, giving tuesday x 4, corporate x 4, brand x 4, tuesday x 4
sub-group 8 members: 931 % of total: 25,94	Language: en nonprofit x 298, fundraising x 151, philanthropy x 81, consultant x 79, profit x 72, non profit x 63, organization x 62, change x 61, speaker x 60, community x 59, helping x 51

Appendix H

A comparison of the operationalization of three mechanisms characterization actions by organizations engaged in the proto-institutionalization of a field:

	Powell et al. (2017)	Network based on co-memberships in Twitter lists
Proselytizing	“If its outdegree is twice its indegree”	- if the number of accounts they follow <i>in their sub-region</i> is twice the number of their followers <i>within their sub-region</i> .
Convening	“If an organization’s indegree is larger than a threshold [...] we set this threshold to 99 [out of a maximum of 368]	- if the number of followers <i>in their sub-region</i> is above a threshold (e.g., a third of the sub-region members)
Strengthening	“those organizations whose bidegree is larger than 50% of their indegree”	- those organizations which follow, and are followed by, a number of Twitter accounts in their sub-region which is larger than 50% of the number of their followers within their sub-region.

Conforming		- if an organization tends to follow a large number of Twitter accounts from <i>its own sub-region</i> , relative to Twitter accounts it follows which are <i>outside</i> its sub-region.
Interfacing		- organizations following, and being followed by, Twitter accounts situated <i>outside</i> their sub-region, while being followed by a large number of members from their sub-region

Appendix I

1. Repository for the Java code used to create local networks from seeds:

<https://github.com/seinecle/issuefields>

A repository is also available in Python, with slight differences (does not include capping limits for number of lists and users, see main text):

https://github.com/MohamedBnbm/seed_expansion

2. Repository for the Java code used to identify the most common terms per sub-region:

<https://github.com/seinecle/TextMiningOnCommunities>