How to Choose Community Detection Methods in Complex Networks

Cécile Bothorel, Laurent Brisson, Inna Lyubareva

How to Choose Community Detection Methods in Complex Networks

The Case Study of Ulule Crowdfunding Platform

Cécile Bothorel[1], Laurent Brisson[1], Inna Lyubareva[2]

[1]IMT Atlantique, LUSSI Department, UMR Lab-sticc, 29238 Brest Cedex 3, France
[2]IMT Atlantique, LUSSI Department, LEGO EA2652, 29238 Brest Cedex 3, France

Abstract Discovering community structure in complex networks is a mature field since a tremendous number of community detection methods have been introduced in the literature. Nevertheless, it is still very challenging for practitioners to choose in each particular case the most suitable algorithm which would provide the richest insights into the structure of the social network they study. Through a case study of the French crowdfunding platform, Ulule, this paper demonstrates an original methodology for the selection of a relevant algorithm. For this purpose we, firstly, compare the partitions of 11 well-known algorithms. Then, bivariate map based on hub dominance and transitivity is used to identify the partitions which unveil communities with the most interesting size and internal topologies. These steps result in three community detection methods relevant for our data. Finally, we add the socio-economic indicators, meaningful in the framework of the crowdfunding platform, in order to select the most significant algorithm of community detection, and to analyze the cooperation patterns among the platform's users and their impact on success of fundraising campaigns.
In line with previous socio-economic studies, we demonstrate that the social concept of homophily in online groups really matters. In addition, our approach puts in light that crowdfunding groups may benefit from diversity.

Keywords Social Networks Analysis, Community Detection, Choice of method, Complex Networks, Online cooperation, Crowdfunding

1. Introduction

In recent years socio-economic research on online groups and communities often proposes to extend the traditional approach and to encompass social network analysis modelling relationships by edges in the graph. The joint analysis of the two types of data—socio-economic and network structure—makes possible to provide important insights on the group functioning and to reveal properties of a social network that are not immediately obvious, e.g. the existence of sub-networks or communities operating within the global network.

For instance, in the context of crowdfunding platforms, many studies focus on the directly observable interactions among the participants of individual projects and show their role for the success of fundraising campaigns [14, 1, 27]. However, few empirical studies focus on relational structures which are *not explicitly* stated in the context of crowdfunding platform [11] and question whether relational circles can go beyond individual projects, by broadening the initial social capital of projects' leaders, and form a platform-level cross-project social network. Does the participation of the members of this cross-project network guarantee higher success rates of crowdfunding campaigns? These questions have strong managerial and economic implications for the platforms: *Could this networking lead to the formation of a core group with an intense participation? Should the platforms further develop the mechanisms for cross-project networking? How should project leaders surround themselves to give themselves every chance of success?* The study of these questions necessitates an appropriate methodology based on the detection of *implicit* communities. Therefore, the question arises of how to choose an appropriate community detection algorithm, relevant to the particular context.

As a matter of fact, many community detection methods exist. They have different ways to divide a network into multiple subsets of densely connected nodes, and hence result in different community structures. They may rely on different notions of community, and even with the same notion (eg. Newman defines a community as a "group of vertices with a higher-than-average density of edges connecting them" [17]), algorithms may optimize different objective functions or use different heuristic to get efficient implementations. The authors of the methods compare their efficiency in terms of computational properties, such as complexity, as well as in terms of validation metrics: modularity, rand index, conductance, etc [9], describing resulting partitions quality. Although, as demonstrated in some comparative surveys [8,9,12], choosing a community detection method is still a problematic task, especially if we are not aware of the underlying mechanisms.

Some recent papers have attempted to provide guidance on algorithm selection using criteria such as the mixing parameter of a network, computation time, or overlap with a simulated community structure [8,27]. Others provide more intuitive criteria such as the size of the communities [7] or descriptive network-oriented metrics on structural patterns within communities [5,6]. N. Smith *et al.* argue that such criteria should not be applied alone, but in conjunction with business-oriented objectives: the “best” method depends on the context, on the research question, i.e. on how the communities will be used [24].

The current paper addresses this problem. It proposes an original methodology to guide practitioners in their choice of methods in connection to a specific research question. We first pre-select some candidate methods, and qualify their results, i.e. the resulting partitions, through qualitative characteristics. Then, final users may compare the few pre-selected partitions with comprehensive measures and finally select the most relevant one regarding the research question.

The case of the Ulule crowdfunding platform is used as example. In this case study, we suggest that some *non-directly observable* communities may exist within the cross-project social network. However, as in many exploratory studies, we have no preliminary information about their number or structures. Our goal is to discover these communities, to describe their internal organizations and to explore whether there is a relationship between particular communities’ organization and the success of fundraising campaigns. This last element is the central research question of the case study.

This paper is organized as follows: In Section 2, we introduce the Ulule platform and its cross-project social network. In Section 3, we present a methodology which permit us to select, among 11 methods, three most relevant and convergent algorithms to discover the communities. In Section 4, by introducing additional socio-economic attributes of communities, we choose the most suitable method and discuss the relative success of different communities for fundraising campaigns. Last Section concludes.

2. Ulule crowdfunding platform

2.1. Ulule platform and its network

Crowdfunding represents a model of participatory financing, which has been used by an increasing number of companies, associations and individuals since the early 2010s. Its principle consists for a project "pitched" by its creator, to collect money from a large public. Thanks to numerous interaction mechanisms between participants (comments, news, promotion systems, etc.), these platforms play the role of facilitators of social capital.

To better understand the networking role of crowdfunding platforms, we use empirical data for the period 2010-2016 from the Ulule platform—one of the main crowdfunding platforms in France and Europe, which pays particular attention to the strengthening of its proper platform's community[1]. After data cleaning, we keep 19,544 projects, of which 11,900 were successfully funded, and 7,644 failed. They globally attracted 876,758 contributors, who contributed a total of 47.75 million euros.

2.2. Graph of co-contributions

While 99.7% of Ulule contributors are one-time funders, 0.3% of them are not only very active by contributing to more than three different projects, but they regularly "meet", in each fundraising campaign, the same participants also contributing to these projects. We suggest that these active contributors, involved in a least 3 projects with at least 1 other contributor, are candidates to form the cross-project platform's network. On average, they contribute to 14 different Ulule projects, for an average total amount of over 600 euros per contributor. With these active Ululers, we define a non-oriented graph of co-contributions, in which each edge (u,v) means that the users u and v have contributed to 3 or more projects together. There are 469 connected components. All of them contain less than 10 nodes, except the largest one. With its 2,081 nodes and 4,749 edges, this giant component proves that the social network, transverse to individual projects, really exists at the Ulule platform.

The members of the cross-project network share projects with 4.56 other members (average degree). We find a power law distribution of degrees (the estimated exponential coefficient is 2.27) which is a common property in online social networks. 25% of the members have a degree greater than 4, the maximum degree being 199. There are therefore contributors who co-finance projects with really important number of different contributors: 24 of them have more than 50 neighbors in the graph.

We notice that the distribution of the Ulule's thematic categories according to whether or not projects are financed by the members of the graph reveals interesting information. The improvement in the success rate of campaigns of projects

[1] https://www.ulule.com/about/ulule/

belonging to the graph is observed for all categories (28.6% overall improvement), especially for Games, Comics, Technology and Publishing. Indeed, these domains have a strong social component in the production and consumption of goods. These findings confirm our interest in studying communities of the active Ululers.

We find out that members of the obtained social graph vary a lot in terms of their social activity (node's degree, clustering coefficient and centralities) and contribution behavior (number of projects funded, average amount of contributions, specialization rate which quantify the variety of thematic categories addressed (projects' category) and their similarities with the neighboring nodes).

Multiple correspondence analysis on these attributes, followed by a hierarchical bottom-up classification, leads to 5 clusters of Ulule's active contributors. A detailed version is available [15], but for space reasons, we sum up the 5 profiles as follows:

- The Sponsors (18) are the Ululers with a high degree and betweenness centrality. They may be considered as facilitors. Involved in 140 projects (in average), they are very active on various topics (not specialized).
- The Followers (653) are the users who arrive late, during the second half of the campaigns. They prefer very big projects (which average objective is more than 17k euros). Exhibiting a high closeness centrality, they frequently co-contribute with other contributors, in particular the Sponsors.
- The Precursors (538) are characterized by an early arrival in projects and especially before all their neighbors.
- The Collaborative Specialists (368), very highly thematically specialized, have also a very high clustering coefficient indicating a strong cohesion of links between neighbors. This may show a strong solidarity between some Ululers whose financing decisions are often collective, even if their fields of specialization are not necessarily identical. They don't contribute to a very large number of projects and are not attracted by the size of the projects.
- The Specialists (504) are not high contributors as well in terms of number of projects, nor in volume of contributions. They are not particularly highly connected as well, but they are however passionate about very specific themes, the same themes than their neighbors, demonstrating therefore social homophily.

These findings will be mobilized in the Section 4 to refine the final choice of the community detection method in connection with the case study research question.

3. A methodology to choose community detection methods

There are many approaches to perform community detection based on different paradigms, including cut, internal density clustering, stochastic equivalence, flow models, etc [9]. The purpose is not to provide an exhaustive overview here. We refer the reader to surveys like [9, 23] to get details about these different approaches. In this work we focus on well-known methods, apply them to our case-specific graph and show how different the partitions produced can be, making the choice of a method non obvious.

We carefully kept a large variety of approaches as summarized in Table 1. While Edge betweenness is based on edge centrality detection in order to split networks into several communities, Louvain and Fast greedy optimize modularity by iteratively folding nodes into meta-nodes. Spectral method is also based on modularity, but identifies the community structure by finding leading eigenvectors corresponding to largest eigenvalues of a modularity matrix. Some approaches are based on a dynamic distance: for Walktrap, if two nodes are in the same community, the probability that a random walker will move from one to another in only a few movements is very high (notion of trap) and consequently the distance is low. Conclude combines a similar random walk-based distance to agglomerate nodes, and the local optimization of modularity inspired by the Louvain method selects iteratively the best agglomeration. Infomap relies on finding a configuration that maximizes the compression of random walks represented by an encoded binary sequence. Inspired by epidemic spreading mechanisms, a more basic and direct distance is used by the Label propagation and its variant SLPA, where a node should belong to the same community as most of its immediate neighbors. In the same vein, concepts have also been borrowed from theoretical physics with the Spin glass model which may be seen as an alternative to modularity maximization; the idea is to consider nodes as spin states and to minimize the energy of the configuration of spins to reach a stable state. Finally we can cite a statistical inference approach, DCSBM which uses stochastic blockmodels to infer the likeliness that a given observed network (and its latent block structure) is generated from a compatible model, and then suggest the most likely set of model parameters.

Methods are often compared with classic quantitative measures from information theory domain, such as NMI, the Normalized Mutual Information [3], which evaluates their agreement to arrange nodes into similar clusters. Recent studies introduced more intuitive, and very simple quantitative metrics such as the size of the communities [7] where methods are similar if they produce similar community size distribution. In [5], the authors propose metrics dedicated to graph analysts. They describe the communities with structural measures, in order to quality the communities from a topological point of view. Each community is qualified with well-known organizational patterns, such as star-based structures, cliques (this will be detailed later, in Figure 3).

In this paper, we propose to practitioners a methodology based on both these quantitative and qualitative metrics. We show how to use them in order to compare the methods; through the case study from Ulule, and its specific business question—which communities lead to successful fundraising—, we show how to select the most relevant method:

- The initial step is to run community detection algorithms. We used 11 methods described in Table 1 and 11 partitions (where each node belongs to only one community) were obtained.
- Step 1 aims at the choice of a subset of methods. We compare the partitions with validation metrics. Since we have no preliminary knowledge on the platform's communities, we are looking for the methods making consensus findings in order to build a robust foundation for subsequent steps.
- As proposed in the previous literature, Step 2 brings an additional intuitive criterion, the size of communities, to refine the subset of consensual methods.

- Step 3 characterizes the consensual partitions with qualitative measures that are relevant for the current business-oriented problem. As our case study focuses on organizational patterns involving nodes, we will use bivariate maps based on graph structural indicators, such as the hub dominance and the clustering coefficient.
- The last step introduces specific “business” indicators related to the current problem to finally make the choice decision (section 4).

3.1. Initiate the selection with consensual partitions

After the initial step, that is run the 11 methods with their default parameters), we obtain 11 partitions. In the first selection step, we compare how nodes are arranged into clusters: we compute the Normalized Mutual Information (NMI) often used in community detection because it allows the comparison of partitions even where nodes are assigned to a different number of clusters. We apply its normalized variant with values in [-1, 1] which is popular in the field of community analysis [3]. The Figure 1b shows how our 11 partitions are astonishingly similar from a NMI point of view where all scores are positive, ranging from 0,26 to 0,76, with a large majority above 0,5. Eight among them demonstrate scores greater than 0.6 when compared to each other, with a very consensual group consisting of four methods: Edge Betweenness, SLPA, Fast greedy and Walktrap with scores higher than 0.7. There are three slightly different methods: Spectral, Label Propagation and Spin Glass that produce more specific partitions, which are all different from each other. While this difference may be easily explained by the fact that Spectral and Spin Glass implement inherently distinctive mechanisms, the results of Label Propagation, being a variant of SLPA, are quite surprising.

Method	**Approach**	**Reference**
Louvain	Multilevel modularity	Blondel et al. (2008) [2]
Fast greedy	Modularity optimization	Clauset et al. (2004) [4]
Spectral	Vector partitioning	Newman and Girvan (2004) [18]
Spin glass	Energy model	Reichardt and Bornholdt (2006) [21]
DCSBM	Stochastic blockmodels	Karrer and Newman (2011) [13]
Walktrap	Dynamic distance	Pons and Latapy (2005) [19]
Conclude	Dynamic distance	Meo et al. (2014) [16]
Edge betweenness	Edge centrality detection	Girvan and Newman (2002) [10]
Infomap	Information compression	Rosvall and Bergstrom (2008) [22]
Label propagation	Topological closeness	Raghavan et al. (2007) [20]
SLPA	Topological closeness	Xie and Szymanski (2011) [25]

Table 1 : A summary of community detection methods used to study community structure.

As an intermediate result, on the basis of the NMI scores, Edge Betweenness, SLPA, Fast greedy, and Walktrap converge on their clustering task. To put it differently, knowing a random node’s affiliation in Edge Betweenness partition, in our example, gives a high probability to successfully deduce its membership in the 3 other partitions. Louvain is also quite close to this group and as it is frequently used, we don't want to discard it right now.

These 5 methods reach a quite good consensus. We argue that in exploratory studies, as the current one, where practitioner has no a priori knowledge about communities that she wants to analyze, it is important to identify such robust clustering, demonstrating an agreement between different methods. This step is determinant for the further exploration of the research question.

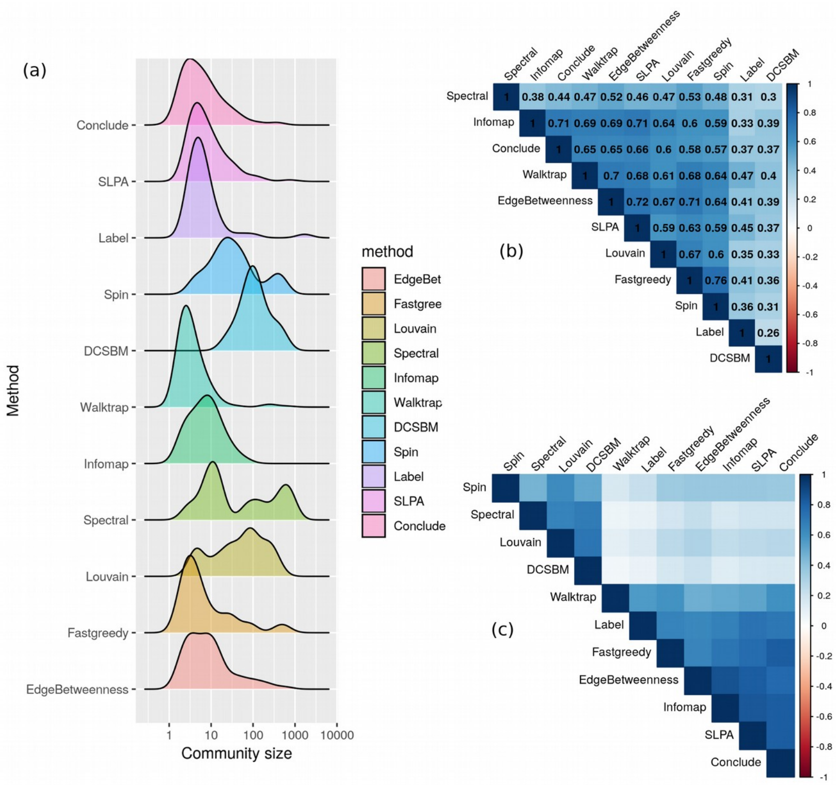

Figure 1: The similarity between community detection methods in terms of (a) Community size distributions, (b) NMI, (c) size fitting quality,

3.2 The size of communities as qualitative choice

We propose then to complete the NMI analysis by adding information about the size of the discovered communities (Figure 1a). One can notice two types of partitions. Some communities are large with tens or sometimes thousands of members. Other partitions, on the contrary, exhibit small (2 or 3 members) to medium-sized communities (around 10-20 nodes). These two classes of methods are depicted in Figure 1c, where the similarity score compares the distribution of sizes of communities (score introduced in [6]).

Regarding large communities, Louvain, Spin glass and Spectral produce an interesting variety of large to very large communities, whereas DCSBM only produces huge ones. We observe the same disparity when we focus on partitions with smaller groups. SLPA and Walktrap produce a lot of communities with approximately 10 nodes, without notable variety of sizes. On the contrary, Fast greedy's and Edge Betweenness partitions have a more flat profile in Figure 1a: most of their communities are small to medium, but some of them have also more than 100 nodes.

Therefore the consensual methods—Louvain, Edge Betweenness, SLPA, Fast greedy, and Walktrap—produce different distributions of sizes of communities, we have to choose whether we give priority to small-medium communities or to large ones. This example demonstrates that whereas in some contexts, the size criteria may be enough to make a choice, such information is not always sufficient. Taking into account that our research question focuses on the communities' forms and their efficiency for the fundraising campaigns, we add topological indicators in order to differentiate the partitions from the organizational perspective.

3.3. Structural classification of consensual partitions

In order to characterize organizational patterns within communities, which is our business-oriented objective here in the case study, we propose the use of structural measures applied to communities such as internal link density, average centrality of nodes, average degree. Such indicators are interesting to be combined in bivariate map [6] to describe structural patterns. For example, ploting a bivariate map with the mean out degree fraction (meanODF) paired with its standard deviation (stdODF) allows to explore different situations regarding the openness of communities and the cooperation between groups of Ululers [5]. However, Hub dominance and Transitivity are particularly relevant when one considers internal patterns of organization like cooperation, because their combination leads to well-known patterns depicted on Figure 3:

- Hub dominance: Internal edges of a community can be distributed in various ways around its nodes, either concentrating around a few highly centralized nodes, or uniformly distributed over the nodes. The Hub dominance metric identifies the level of centralized organization around well-connected nodes. The higher this metric of a community, the more likely it has a hub-like structure. Hub dominance can be considered as a normalized version of degree centrality. High Hub dominance leads to the well-known star-based patterns as depicted on Figure 3.
- Transitivity: Very similar to the clustering coefficient [6], Transitivity reflects the probability that adjacent vertices of a vertex are connected. This metric is usually employed to evaluate modular structures (grids) or clique dominance in networks (Figure 3). For example, high Transitivity coupled with similar spheres of interests (or other attributes) among individuals often indicates the existence of social homophily, especially in online groups, also known as the proverb "birds of a feather flock together".

Figure 2 plots the communities in the Hub dominance vs Transitivity space. One can see that large communities are concentrated in the same area. The methods responsible of those large communities (e.g. Louvain) indeed produce a priori very few different structural patterns. With low Hub dominance and low Transitivity, most communities could be considered as "string-based" structures (Figure 3a). In order to detect whether or not cooperation exists within large communities, we should have to zoom-in to extract dense sub-zones, i.e. apply again a detection of community to each community, and then project the new smaller communities in our bivariate map.

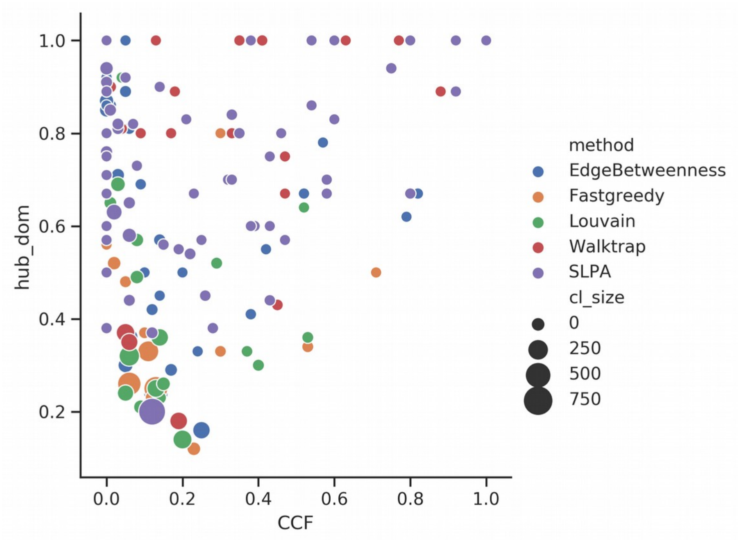

Figure 2: Structural description of communities in terms of Hub dominance and Transitivity (consensual methods)

Conversely, Edge Betweenness, Walktrap and Fastgreedy which produce small and medium-sized communities seem to generate various directly observable types of organizations. Indeed the points are distributed in 3 of the 4 areas of our bivariate map in Figure 2. We especially find a lot of groups in the upper part of the map. This means that their members are organized around hubs, but in 2 different manners. When the Transitity is high, we find some clique-based organizations (Figure 3d), where Ululers are (nearly) all contributing with each other to common projects. When the Transitivity is low, the organizations are mimicking star-based structures, with a very high centralization (Figure 3c). Ululers in these groups are less involved in horizontal cooperation, but seem to follow influencers (Ululers with large degrees) who concentrate common projects with a lot of poorly connected contributors.

With these new insights, Edge Betweenness and Walktrap seem to be very good candidates: (i) they belong to the consensual group of methods previously shown, producing a robust partition; and (ii) they offer various internal organizational forms. Louvain, without offering diversified topological structures, demonstrates however interesting properties of high Hub dominance in its large communities. For this reason, we propose to keep Louvain to further explore the typology of its communities.

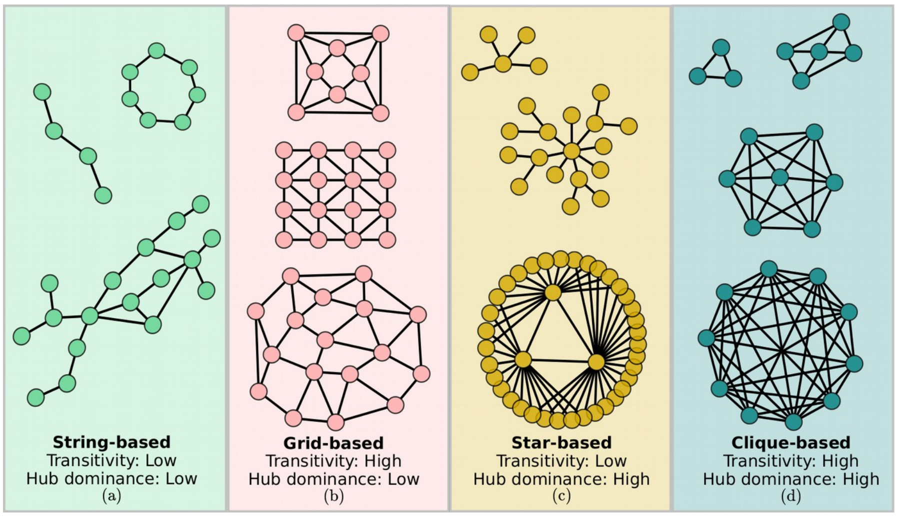

Figure 3: A categorization of internal community structure according to two topological property

To summarize, Table 3 describes the obtained communities using several conventional network metrics. First, the three partitions do not appear very distinguishable with regards to these traditional indicators. Interestingly, it seems that the size of the groups does not really influence their average degree. Moreover, the average closeness centrality and the average density also remain stable regardless of the algorithm used. Second, this confirms the need for more qualitative measures, such as the bivariate map that we proposed here above, to cover the specificities of different community detection methods and to be able to choose among them.

Method (Communities count)	**Member count (mean)**	**Degree (mean)**	**Clustering coefficient (mean)**	**Betweenness (mean)**	**Closeness (mean)**
Louvain (22)	90.48	3.63	0.26	2,871.45	0.24
Edge Bet. (72)	28.90	2.96	0.20	2,234.84	0.23
Walktrap (167)	12.39	2.29	0.15	1,695.03	0.22

Table 3: Conventional topological metrics to describe communities: number of communities generated (in brackets) and average statistics calculated on the whole set of generated communities.

4. Introduction of case-specific data and typology of communities

At this step, in accordance with the objective alignment approach [24], we introduce additional information on the contributors' profiles, specific to the current study, which was presented in the Section 2. This helps us to compare classifications of communities, generated by three partitions, and to choose one relevant community detection method. Depending on particular context and available data, other socio-economic indicators may be mobilized for the communities' classification.

4.1. Final selection of the method

To classify the communities of Louvain, Edge Betweenness and Walktrap methods, regarding the distribution of contributors' profiles, the following clustering methods were used: (i) a principal component analysis (2 dimensions) followed by an agglomerated hierarchical clustering (Euclidean distance, Ward method of variance minimization), (ii) the K-means method and (iii) a decision tree.

The decision tree produces clusters of communities (Families in Figure 4) almost identical to the K-means method (completeness = 0.964; Adjusted Rand Index : ARI = 0.854). Proximity to clusters produced by the component analysis is not obvious when considering these measures (completeness = 0.513; ARI = 0.398), but nevertheless, the distribution of the types of contributors remains relatively close.

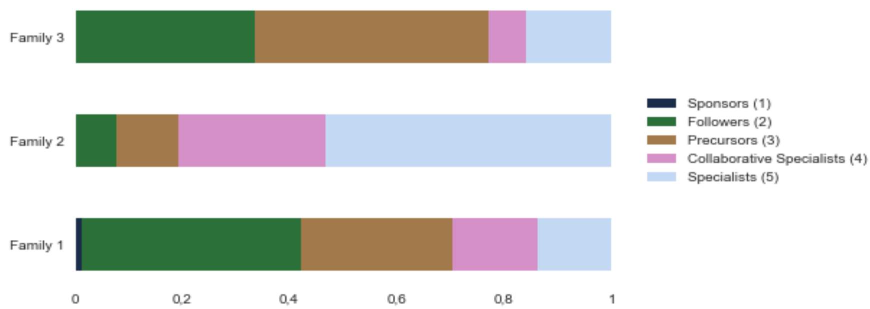

Figure 4: Typology of communities (Edge Betweenness, Decision Tree).

We observe similar distributions of contributors' profiles across the different community families when we compare the 3 community detection methods and their clusters of communities. The same variety of community forms can be seen:

- Family 1 (with the Sponsors): very large balanced communities composed of all the profiles. They have the particularity of having attracted the Sponsors and a lot of Followers.
- Family 2 (Specialists): communities very clearly dominated by Specialists, whether collaborative or not.
- Family 3 (Followers, Precursors): small and even micro communities rather dominated by Precursors and to a lesser extent Followers.

Therefore, regardless the clustering method and the community detection method, the three afore-mentioned families of communities are clearly detected. The only distinguishing feature is the representation of each community family depending on the community detection method. Edge Betweenness proposes the most balanced clusters (Table 4), while Louvain produces mainly Family 1 and 2 items, and not surprisingly, most of the 167 (small) communities produced by Walktrap are from Family 3.

On this basis we choose the Edge Betweenness partition, composed of 72 communities of various sizes, exhibiting substantial examples of each family (see the exact distribution of its Ululers' profiles in Figure 4).

Family	**Community (count)**	**Member count per community (mean)**	**Member contribution (mean in €)**	**Project goal (mean in €)**	**Project count per member (mean)**	**Project count per community (mean)**
1	8	161.5	41.4	8,050	14.4	469.5
2	28	19.1	47.7	8,501	13.3	124.8
3	36	7.0	41.3	9,589	13.7	42.5

Table 4: Sizes of communities and their contributions by Family (Edge Betweenness, Decision Tree).

4.2. Community structures and projects success

Once we know the community detection method, it is possible to describe the communities and their success rates in fundraising campaigns.

Table 5 uses relational (clustering coefficient, the average of the betweenness and closeness centralities of the members, mean degree) and socio-economic attributes (volume of funding, number of interactions via comments, thematic specialization and project success rate) in order to characterize collaboration rates and organization of communities..

Family	**Member count per community (mean)**	**Project count**	**Shared project count (mean)**	**Theme count per community (mean)**	**Comment count per project (mean)**	**Community Clustering Coefficient (mean)**
1	161.5	1,642	4.3	13.4	73.2	0.25
2	19.1	781	3.7	4.1	137.3	0.32
3	7.0	547	3.4	5.8	176.0	0.15

Table 5: Communities characteristics

The three families of communities differ crucially in terms of collaboration. On crowdfunding platforms, it can take different forms: sharing projects within a community (weight of links in the Ulule social network), cohesion of members of a community around the same projects (clustering coefficient), or communication via a feedback system (comments). As presented in the Table 5, each family of communities in the case of the Ulule platform favors one of these forms.

The combination of two collaborative aspects—the weight of links in the graph (average number of shared projects) and the clustering coefficient—can inform us about community organization. For example, in Family 1, the fact that the average number of shared projects is very high but the members are poorly connected to each other (low clustering coefficient) highlights a highly centralized organization around a few central actors. We probably find here the communities with low Transitivity and high Hub dominance (probably star-based communities as seen previously). The Sponsors may play the structuring roles of these communities.

The communities in Family 2 are made up of members who share the same interests. These thematic groups, which are strongly connected and supportive, collectively take their decisions on project funding and contribute on average more than other communities.

Finally, in Family 3, communities are also structured around certain themes but without clear specialization. Members of these communities communicate a lot through the feedback system (comments).

4.3. Comparative analysis of success rates of families of communities

Since the thematic categories are not represented in a balanced way in the three Families of Communities, we cannot directly compare the respective effectiveness of community forms for the different themes. Nevertheless, if we look in more detail at the communities that have a 100 percent success rate for all our indicators (Table 6), there are 13 communities belonging to Families 2 and 3 (respectively 9 and 4 communities). These two Families have very close thematic choices.

These best communities are small in size, quite specialized and finance projects on average on 4 different themes. The projects are very different in terms of the objectives to be achieved (all types of amounts are represented), but less than 25 % of the projects have a larger than average scope, so we have rather modest projects.

Surprisingly, the 9 least successful communities in the graph also belong to Families 2 and 3. What can be noticed in the Table 7 is that the most and the least successful communities share quite similar characteristics (size, thematic specialization). What clearly differentiates the best performing communities is the clustering coefficient and the number of comments, which are significantly higher. In other words, the thematic specialization that characterizes Family 2 and partially Family 3, does not in itself guarantee success of crowdfunding projects. To reach significant economic performance of fundraising campaigns, they must be coupled with strong social involvement and cohesion of its members.

With regard to Family 1, characterized by a high centralization and a high thematic openness, we notice a relatively high success rate. Thematic diversity also attracts participants in communities.

	Shared project count (mean)	Comment count per project (mean)	Project goal (mean)	Member count (mean)	Theme count (mean)	Degree (mean)	Betweenness (mean)	Clustering coefficient (mean)
Success Rate 100% (13 communities)								
mean	11	**517**	8,826	6	4	2.84	2,126	**0.39**
std	6	1,120	6,132	3	3	1.06	1,356	0.35
Min	3	2	2,630	3	1	1.67	777	0
25%	6	21	4,408	3	1	2.00	1,419	0
50%	9	60	7,946	4	3	2,50	1,559	0.43
75%	15	123	10,807	6	7	3.25	2,078	0.67
max	22	3 840	25,019	**13**	9	4.85	5,869	0.88
Success Rate < 85% (9 communities)								
mean	11	234	9,126	4	4	2.06	1,796	**0.09**
std	10	534	6,271	2	3	0.85	721	0.17
min	5	15	2,202	2	2	1.50	693	0
25%	6	29	6,305	3	3	1.67	1,039	0
50%	6	35	7,006	3	3	1.80	2,078	0
75%	12	88	11,189	5	3	2,00	2,078	0.06
max	35	1,653	23,916	**8**	11	4.25	2,970	**0.44**
All (72 communities)								
mean	58	393	10,832	29	6	2.96	2,235	0.20
std	122	686	5,799	78	4	1.18	1,092	0.22
min	3	2	2,202	2	1	1.50	693	0
25%	8	33	6,894	3	3	2.00	1,558	0
50%	17	78	9,053	8	5	2.93	2,077	0.15
75%	36	381	14,915	13	8	3.63	2,728	0.30
max	789	3,840	27,555	579	15	6.46	5,869	0.88

Table 6 Communities with high success rates vs. communities with lower success rates. The indicators are averaged by community. Sub-table “All” shows the indicators for all 72 communities.

5. Conclusion

Community detection makes it possible to identify very diverse groups in a social network. This paper demonstrates a methodology to choose one relevant community detection algorithm, among 11 well-known ones, providing fruitful insights into the cooperation forms not directly observable on a crowdfunding platform.

The choice of a particular community detection method is not an easy or a neutral choice. As demonstrated in the paper, depending of partition methods, practitioners obtain a range of various community types, which will drastically change the final results of their analysis. This paper substantiates that an accurate way to choose one suitable method is a complex task. Especially in the context of exploratory studies it necessitates the combination of a range of techniques, e.g. in our case, partitions' similarities, qualitative criteria and structural indicators (string-based, star-based or clique-based organizations of communities).

In line with N. Smith et al. [24] this study substantiates that the choice of a method is determined by the research context and problematics. Additional techniques, specific data and indicators allow to narrow down the scope of available options in the methods choice. Their alignment with the practitioners' research question plays a crucial role for the final choice of a particular method. In the framework of the case study presented in this paper, the choice of the Edge Betweenness method results from the analysis of socio-economic characteristics and the exploration of the distribution of Ululers' profiles. This way, we have identified 3 robust families of platform's communities and their distinctive features, i.e. organization, number of participants, collaboration intensity, thematic specialization, and performance in the fundraising campaigns. Depending on the context and available data, different socio-economic indicators may be mobilized to obtain the communities' classification and a range of further business-oriented questions may be addressed: e.g. precise distribution of string-based, star-based or clique-based forms in communities' families, life circle and evolution dynamics of the communities and many others.

Acknowledgements The authors of this study would like to thank Arnaud Burgot, the CEO of Ulule, very warmly for providing us with the data from the platform and for giving us his confidence in carrying out this work.

References

1. Agrawal, A., Catalini, C., Goldfarb, A.: Entrepreneurial finance and the flatworld hypothesis: Evidence from crowdfunding entrepreneurs in the arts (2010)
2. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 10, 10008 (2008)
3. Chakraborty, T., Dalmia, A., Mukherjee, A., Ganguly, N.: Metrics for community analysis: A survey. ACM Comput. Surv. 50(4), 1–37 (2017). DOI 10.1145/3091106
4. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E 70(6) (2004). DOI 10.1103/physreve.70.066111
5. Dao, V.L., Bothorel, C., Lenca, P.: Community structures evaluation in complex networks: A descriptive approach. In: E. Shmueli, B. Barzel, R. Puzis (eds.) 3rd International Winter School and Conference on Network Science, NetSci-X 2017, pp. 11–19. Springer International Publishing, Cham (2017)
6. Dao, V.L., Bothorel, C., Lenca, P.: An empirical characterization of community structures in complex networks using a bivariate map of quality metrics. CoRR abs/1806.01386 (2018)
7. Dao, V.L., Bothorel, C., Lenca, P.: Estimating the similarity of community detection methods based on cluster size distribution. In: L.M. Aiello, C. Cherifi, H. Cherifi, R. Lambiotte, P. Lió, L.M. Rocha (eds.) Complex Networks and Their Applications VII, pp. 183–194. Springer International Publishing, Cham (2019)
8. Dao, V.L., Bothorel, C., Lenca, P.: Community structure: A comparative evaluation of community detection methods. Network Science p. 1–41 (2020). DOI 10.1017/nws.2019.59
9. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Physics Reports 659, 1–44 (2016). DOI 10.1016/j.physrep.2016.09.002
10. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99(12), 7821–7826 (2002). DOI 10.1073/pnas.122653799

11. Inbar, Y., Barzilay, O.: Community impact on crowdfunding performance (2014). DOI 10.2139/ssrn. 2524910
12. Jebabli, Malek, Cherifi, Hocine, Cherifi, Chantal, & Hamouda, Atef. (2018). Community detection algorithm evaluation with ground-truth data. Physica a: Statistical mechanics and its applications, 492, 651–706.
13. Karrer, B., Newman, M.E.J.: Stochastic blockmodels and community structure in networks. Physical Review E 83(1) (2011). DOI 10.1103/physreve.83.016107
14. Kuppuswamy, V., Bayus, B.L.: Crowdfunding creative ideas: The dynamics of project backers. In: The Economics of Crowdfunding, pp. 151–182. Springer (2018)
15. Lyubareva, I., Brisson, L., Bothorel, C., Billot, R.: Une plateforme de crowdfunding et son réseau social : l'exemple ulule. Revue Franccaise de Gestion (2020)
16. Meo, P.D., Ferrara, E., Fiumara, G., Provetti, A.: Mixing local and global information for community detection in large networks. Journal of Computer and System Sciences 80(1), 72–87 (2014). DOI 10.1016/j.jcss.2013.03.012
17. Newman, M.E.J.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103(23), 8577–8582 (2006). DOI 10.1073/pnas.0601602103
18. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69(2) (2004). DOI 10.1103/physreve.69.026113
19. Pons, P., Latapy, M.: Computing communities in large networks using random walks (long version). ArXiv Physics e-prints (2005)
20. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 76(3) (2007). DOI 10.1103/physreve.76.036106
21. Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. Physical Review E 74(1) (2006). DOI 10.1103/physreve.74.016110
22. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105(4), 1118–1123 (2008). DOI 10.1073/pnas.0706851105
23. Schaub, M.T., Delvenne, J.C., Rosvall, M., Lambiotte, R.: The many facets of community detection in complex networks. Applied Network Science 2(1) (2017). DOI 10.1007/s41109-017-0023-6
24. Smith, N.R., Zivich, P.N., Frerichs, L.M., Moody, J., Aiello, A.E.: A Guide for Choosing Community Detection Algorithms in Social Network Studies: The Question Alignment Approach, American Journal of Preventive Medicine, Volume 59, Issue 4, 2020, Pages 597-605, https://doi.org/10.1016/j.amepre.2020.04.015
25. Xie, J., Szymanski, B.K.: Community detection using a neighborhood strength driven label propagation algorithm. In: Proceedings of the 2011 IEEE Network Science Workshop, NSW '11, pp. 188–195. IEEE Computer Society, Washington, DC, USA (2011). DOI 10.1109/NSW.2011.6004645
26. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep. 2016;6:30750. https://doi.org/10.1038/srep30750
27. Zheng, H., Li, D., Wu, J., Xu, Y.: The role of multidimensional social capital in crowdfunding: A comparative study in china and us. Information & Management 51(4), 488–496 (2014)