Mapping Organizations' Goals and Leanings in the Lobbyist Network in Banking and Finance

Borut Sluban[1(✉)], Jasmina Smailović[2], Petra Kralj Novak[2], Igor Mozetič[2], and Stefano Battiston[1]

[1] Department of Banking and Finance, University of Zurich, Andreasstrasse 15, Zürich, Switzerland
[2] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova 39, Ljubljana, Slovenia
borut.sluban@uzh.ch

Abstract. We address the question of how can publicly accessible information be used to make a map of the political actors and their leanings, that would benefit both policy makers and stakeholders in the European Commission's 'Better regulation agenda' and contribute to social stability. We explore this possibility by using data from the Transparency Register and the open public consultations of the European Commission in the area of Banking and Finance. We compare lobbying organizations active in this area according to three criteria: (i) their formal categorization in the Transparency Register, (ii) their self-declared goals and activities, and (iii) their leanings towards policy issues as derived from their responses to public consultations. We combine methods from information retrieval, text mining, and network analysis to obtain insights on the policy arena. We find that constructing a similarity network based on preference patterns adds a crucial dimension in the understanding of how lobby organizations engage in the policy making process.

1 Introduction

The process of policy reforms is often triggered by initiatives of the stakeholders that are going to be affected by that future policy, e.g., a specific sector of the industry. In democratic regimes so far policy makers typically consult a limited number of experts and the largest stakeholders directly involved before issuing a new policy proposal. However, this process may leave citizens and smaller stakeholders underrepresented in the process of shaping the policy. Therefore in many countries, governments are working on improving the communication with citizens and stakeholders to increase their involvement in the law-making process. For example, the European Commission (EC) has been making a significant effort to try and engage an increasing number of citizens in the EU law-making process by means of open public consultations[1]. Typically, the responses are lim-

---

[1] http://ec.europa.eu/assets/epsc/pages/60-years.

ited in number to few hundreds, mostly coming from the lobbying organizations that are active in the policy area concerned by the proposal.

There are several empirical studies of interest group mobilization in the EU examining the number and type of interest groups politically active in the EU. The Commission's lobbying register was inspected in [6] to assess the density and diversity of the interest group population per policy domain. The density of interest organizations per economic sector in the EU is explained in [4] on the basis of political and economic institutional factors. Based on an analysis of EC online consultations, [11] found that organized interests potentially can act as a transmission belt between the public and decision makers. Higher mobilization rates were found on those issues that fall within policy areas that are regarded as salient by the general public and those with consequences for budgetary spending. However, little research has been carried out so far on the structure of the networks in which lobbies operate [14]. In [15], data from the EU transparency register and the Orbis database was combined to construct a multiplex lobby network consisting of the affiliation, shareholding, interlocking and client relations between lobby organizations. No simple relation was found between the network centrality of organizations and their size, for instance in terms of funds deployed in lobbying. However each network layer was found to provide complementary information to characterize organizations' influence.

In this work, we study how lobby organizations respond to the EC's public consultations in the area of Banking and Finance, in order to make a map of the political actors and their leanings, that would benefit both policy makers and stakeholders within the EC's better regulation agenda, and would contribute to social stability. We combine methods from information retrieval, text mining, and network analysis to obtain insights from the working of the political arena. By comparing the formal categorization of organizations, their self-declared goals and activities, and their response patterns to public consultation we find that network analysis adds a new dimension in the understanding of how lobby organizations engage in the policy creation process.

2   Data and Preprocessing

In our study we focus on lobby organizations registered in the EC Transparency Register[2] and active in the area of Banking and Finance[3]. We analyze and compare three aspects of these organizations: (i) their formal categorization, (ii) their self-described goals and activities from the Transparency Register, and (iii) their responses to the open public consultations of the EC. Our study includes 363 organizations that responded to multiple choice questions in the public consultations, and that have sufficiently long description of goals and activities in English in the Transparency Register.

Transparency Register was set up by the European Parliament and the EC to increase the open access to information about "what interests are being pursued,

---

[2] http://ec.europa.eu/transparencyregister.
[3] https://ec.europa.eu/info/consultations-banking-and-finance_en.

Table 1. Transparency Register categories and distribution of the 363 lobby organizations considered in this study

Main/Subcategory	Org.
I - Professional consultancies/law firms/self-employed consultants	**5**
Professional consultancies	5
II - In-house lobbyists and trade/business/professional associations	**279**
Companies & groups	101
Other organisations	12
Trade and business associations	140
Trade unions and professional associations	26
III - Non-governmental organisations	**60**
Non-governmental organisations, platforms and networks and similar	60
IV - Think tanks, research and academic institutions	**11**
Academic institutions	1
Think tanks and research institutions	10
VI - Organisations representing local, regional and municipal authorities, other public or mixed entities, etc.	**8**
Other public or mixed entities, created by law whose purpose is to act in the public interest	4
Other sub-national public authorities	1
Regional structures	1
Transnational associations and networks of public regional or other sub-national authorities	2
Total	**363**

by whom and with what budgets". We are particularly interested in two aspects of the data on the organizations from the register.[4] First, there is a formal categorization of the organizations (i.e., *mainCategory* and *subCategory*), and second, we analyze the organizations' *goals* and *activities* (as provided in the fields *activityEuLegislative*, *activityRelevantComm*, *activityHighLevelGroups*, *activityConsultCommittees*, *activityExpertGroups*, *activityInterGroups*, and *activityIndustryForums*).

The Transparency Register provides information about main- and subcategory in which an organization is registered. The majority (76.86%) of 363 organizations we are focused on are in the "II – In-house lobbyists and trade/business/-professional associations" main category, while the organizations are more spread over the subcategories. Distribution of categories are in Table 1.

---

[4] https://data.europa.eu/euodp/en/data/dataset/transparency-register.

In the process of registering an organization in the Transparency Register the organization's describe their goals and activities. We extract these descriptions and merge them into a single document for each organization. We remove URLs as we are focused on the content, and not on the links to other sources. We split the documents into sentences and take into account only those which are classified as English text.[5] Splitting into sentences is necessary since some documents contain texts written in more than one language. After language detection, we take into account only the documents which contain at least 50 characters.

Public Consultations[6] are used by the EC to involve citizens and stakeholders in the law-making process. Our study is focused on the area of Banking and Finance. Since October 2015, responses to consultations have been collected by the EC in a more structured manner. For the time period of one year we acquired data about 16 consultations.

We extracted data from the consultations questioners for organizations which provided at least one answer to a multiple choice question. This allows us to find exact matches of their responses in contrast to open ended questions where comparison of two answers is more ambiguous. This narrowed down our study to 12 consultations. Each response to a consultation was transformed into a binary vector denoting which of the answers to the multiple choice questions were provided. For each stakeholder participating in at least one consultation, a concatenated vector from all 12 consultations was created. We use such 1373-dimensional binary vectors as representations of the preferences of responding stakeholders.

3   Modeling Lobby Organizations

We examine lobbying organizations in the area of Banking and Finance by comparing their: (i) formal categorization, (ii) self-described goals and activities, and (iii) responses to public consultations.

To study the conduct of these organization we contrast predefined subcategories from the Transparency Register (see Table 1), to the clusters automatically discovered from the descriptions of goals and activities, and with the communities detected in a similarity network of preference patterns. In this section we describe the methods for obtaining clusters of organizations based on their own descriptions (Sect. 3.1), and communities of organizations extracted from the network structure imposed by the similarity between preference patterns expressed in the consultations (Sect. 3.2).

3.1   Clustering of Descriptions

K-means clustering [8] partitions the input data into k clusters, in a way that similar items are grouped into the same cluster. We apply the *KMeansClusteringFast* algorithm from the LATINO library for clustering English descriptions of

---

[5] Using a language detector from the LATINO text mining library (https://github.com/LatinoLib/LATINO).

[6] https://ec.europa.eu/info/consultations_en.

Table 2. Results of clustering ($k = 3$) applied to textual self-described goals and activities. Each cluster is identified by a short name, the number of organizations it covers, and the top ten centroid terms together with their weights

Cluster ID	Org.	Top centroid terms
Advisory/Representation	35	advisory(0.224); advisory committee(0.201); workers(0.171); committee(0.165); free(0.148); free movement(0.121); movement(0.105); trade(0.097); trade unions(0.096); member advisory(0.088);
Energy/Sustainable investment	207	investment(0.125); energy(0.118); group(0.106); financial(0.102); european(0.101); sustainable(0.094); market(0.092); directive(0.09); insurance(0.089); eu(0.089);
Banking/Insurance	121	banking(0.214); payments(0.17); association(0.119); members(0.1); european(0.097); represents(0.097); interests(0.09); financial(0.088); insurance(0.085); industry(0.081);

the organizations' goals and activities. Before applying the clustering algorithm, we prepare the data by performing standard text preprocessing [7] in terms of tokenization and stemming, removing stop words, constructing unigrams and bigrams, and constructing feature vectors by employing tf-idf weighting scheme and normalization. The resulting vectors are given to the clustering algorithm together with a predefined value of k.

In order to find the best clustering setting, we test k in the range between 2 and 30 and also ten different seeds for initial clustering for each tested k. For estimating the quality of clustering results we calculate the Silhouette coefficient [12]. Based on the evaluation results we choose $k = 3$ and a clustering seed which maximizes the Silhouette coefficient for the selected k value.

The clustering results for $k = 3$ are shown in Table 2 and Fig. 1. Table 2 shows, for each cluster, how we named it, number of associated organizations, and top ten centroid terms together with their weights. On the other hand, the fifty most important centroid terms for each of the clusters are visualized as tag clouds in Fig. 1. The size of the tag clouds corresponds approximately to the number of organizations in the clusters. As can be seen from the results, lobby organizations could be, according to their goals and activities, roughly divided into three groups engaged in *Advisory/Representation*, *Energy/Sustainable investment*, and *Banking/Insurance*.

3.2 Community Detection in the Co-voting Similarity Network

We assessed the difference in how organizations responded to the consultations by calculating the cosine similarity between their binary vectors of answers that are representing their respective preferences. Cosine similarity is a measure of similarity between vectors $\mathbf{a}$ and $\mathbf{b}$. It is calculated as the normalized dot product between vectors $\mathbf{a}$ and $\mathbf{b}$: $\text{sim}(\mathbf{a}, \mathbf{b}) = \cos(\angle(\mathbf{a}, \mathbf{b})) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \cdot |\mathbf{b}|}$. It is a suitable

Fig. 1. Tag clouds of the three clusters, resulting from the self-described goals and activities of the 363 lobby organizations.

measure for comparing responses to consultations, as it is a normalized count of common answers provided by two organizations.

From the calculated similarities between all pairs of organizations we construct a weighted undirected network.[7] Nodes of the network represent the 363 organizations and 19,328 weighted edges reflect the similarity between their responses to the public consultation on Banking and Finance. Two nodes are connected with an edge if the similarity measure is greater than zero. On this weighted network we apply the Louvain method for community detection [5,9] with the following parameters: $Randomize = Off$, $Useedgeweights = On$, $Resolution = 1.0$. This results in five communities and a modularity value of 0.509.

Table 3 presents the characteristics of individual communities. By analyzing the members of each community and their participation in the consultations, we observe that the detected communities represent organizations engaging on a certain topic or participating in particular consultations. In Fig. 2 we show for each community the distribution of its members' responses to the individual consultations, and the consultations for each community are ordered according to the coverage of responses to a consultation by the community members. Therefore, we labelled the communities according to their predominant field of engagement or with the corresponding name of a consultation, see Figs. 2 and 3, and Table 3.

We can observe that communities 1, 3, and 4 evolve around organizations mainly responding to one consultation either on Corporate tax transparency, Building a capital markets union, or Non-financial reporting guidelines, respectively. In the first community also all organizations responding to Extractive forestry are included. Organizations in communities 2 and 5 tend to respond to multiple consultations. Community 2 is formed by the similar preferences about

---

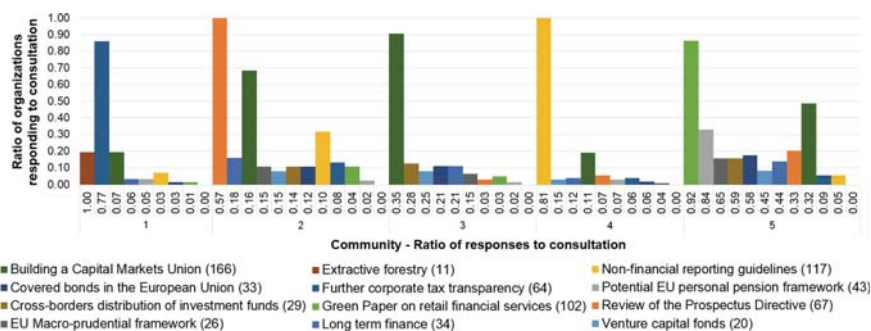[7] The network is constructed and analyzed in Gephi (https://gephi.org/) [3].

Fig. 2. Distribution of the community members' responses to consultations. Consultations are ordered by decreasing coverage of the community members, for the five detected communities. Absolute number of respondents per consultation is given in parentheses next to each consultation.

Table 3. Characteristics of the detected communities. Communities are labelled by the consultation in which the member organizations were mostly engaged. The estimated yearly expenses (from the Transparency Registry) and the average expense per community member are in €

Community label	Org	Expenses [€]	Avg./Org. [€]
1 – Further corporate tax transparency	57	37, 971, 414	666, 165
2 – Prospectus Directive	38	21, 136, 500	556, 223
3 – Building a capital markets union	64	20, 538, 791	320, 918
4 – Non-financial reporting guidelines	95	57, 853, 462	608, 983
5 – Retail fin. services/Personal pension framework	109	66, 563, 238	610, 671

the Prospectus directives and Building a capital market union. Community 5 is the largest and is mostly concerned about Retail financial services, the Personal pension framework, and about Building a capital market union, however it also includes the majority of organizations responding to Covered bonds, Cross-border distribution of investment funds, and the Macro-prudential framework.

The similarity network with two different parameters for the node size is in Fig. 3. In the left part (a) of the figure, the node size corresponds to the number of consultations in which the organization participated. As in Fig. 2, we see that in communities 2 and 5 there is a larger number of organization responding to several consultations. In the right part (b), the node size is related to the organization's estimated lobbying expenses. Compared to other communities organizations of community 3 spend on average the least on lobbying (see also Table 3). Different colors of nodes correspond to different communities. The legend in the lower left part of the figure shows the number and the percentage of nodes (organizations) which belong to each community. The network layout is produced by OpenOrd [10], a force-directed layout algorithm suitable for large undirected graphs. The algorithm emphasizes division of node clusters, and
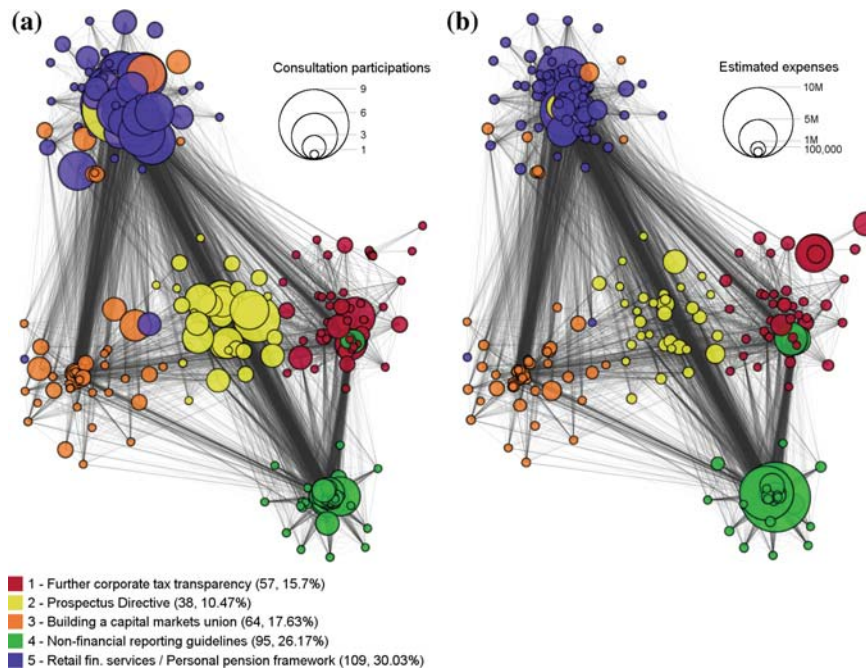
Fig. 3. Similarity network of lobby organizations' co-voting on EU public consultations on Banking and Finance. Node size: (a) number of consultations to which an organization responded, (b) the organization's estimated lobbying expenses. Different colors of nodes correspond to the five detected communities.

includes components such as edge-cutting, multilevel graph coarsening, average-link clustering, and parallel execution. We employed the OpenOrd algorithm with an edge-cutting value of 0.95 and 850 iterations.

4   Comparisons and Discussion

The lobby organizations studied in this paper are grouped according to different criteria. First, they belong to different categories and subcategories (see Table 1). Second, clustering of textual self-described goals and activities yields three clusters (see Fig. 1). And third, co-voting similarity of their responses to policy consultations yields five communities (see Table 3). We perform empirical, visual and qualitative comparison of the three types of lobby organization groupings, to discover the similarities and differences in the conduct of their engagement in the area of Banking and Finance.

We apply the B^3 measure [2] to asses the relations between the three different groupings. B^3 is considered the most appropriate measure for extrinsic evaluation of clustering [1]. The B^3 measure decomposes the evaluation into calculating the precision and recall associated with each node in two groupings. The

comparison of two groupings is measured as the average value over all nodes, i.e., in our case all 363 organizations.

Let N be the set of all nodes in two groupings, say grouping 1 and 2. For each node $n \in N$, we denote with $L(n)$ the set of nodes with the same group label as n, i.e., members of the same group (category, cluster, or community) in grouping 1. With $C(n)$, we denote the set of all nodes which are members of the same group as n in grouping 2. The B^3 precision of a node n, $P(n)$, is computed as the fraction of nodes which have the same label as n in both groupings, from all the nodes which are in the same group as n in grouping 2. Similarly, the B^3 recall of a node n, $R(n)$, is computed as the fraction of nodes with the same label in both groupings, from all the nodes with the same label as n in grouping 1. The precision and recall are further combined into the F_1 score,[8] which is a harmonic mean of the precision and recall:

$$P(n) = \frac{|L(n) \cap C(n)|}{|C(n)|} \qquad R(n) = \frac{|L(n) \cap C(n)|}{|L(n)|} \qquad F_1(n) = 2\,\frac{P(n)\,R(n)}{P(n) + R(n)}$$

The F_1 score balances precision and recall. The precision, recall, and F_1 score of a grouping is a micro average of the scores of all the nodes. The B^3 precision, recall and F_1 scores between the categories, clusters, and communities of organizations are presented in Fig. 4.
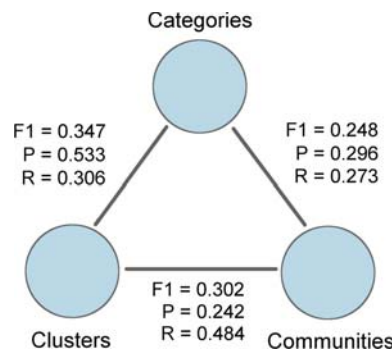

Categories

F1 = 0.347
P = 0.533
R = 0.306

F1 = 0.248
P = 0.296
R = 0.273

F1 = 0.302
P = 0.242
R = 0.484

Clusters   **Communities**

Fig. 4. The B^3 precision, recall, and F_1 score between the categories, the three clusters, and the five detected communities.

All measures have relatively low values, indicating that there is no greater matching, inclusion nor containment among the groups in different partitions.

---

[8] The F_1 score is a special case of Van Rijsbergen's effectiveness measure [13], where precision and recall can be combined with different weights.
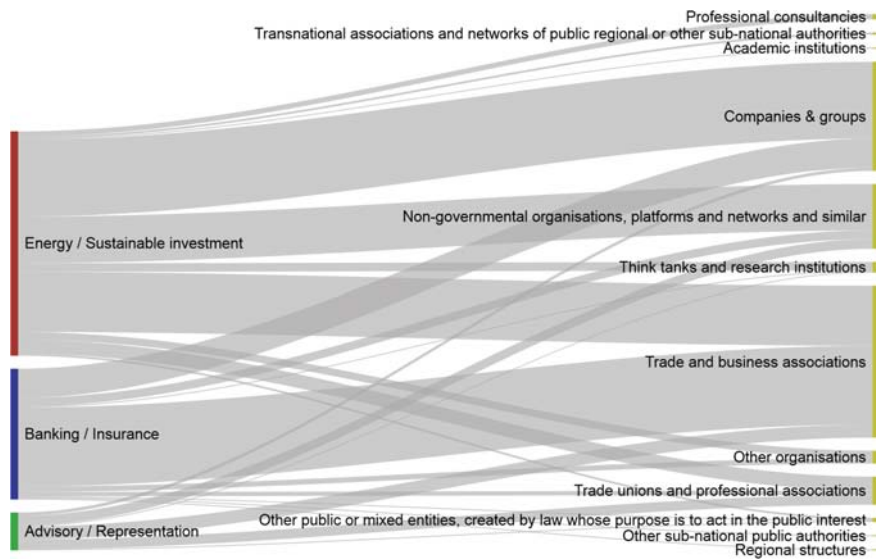
Fig. 5. Correspondence between the three clusters of goals and activities (left), and the subcategories in which the lobby organizations are registered (right).

The highest values of precision (or recall reversely) can be observed between the clusters and the communities, and between clusters and categories, showing that clusters on average tend to contain around half of the communities or categories.

These results can be very intuitively presented with Sankey diagrams[9] of the three pairs of organization groupings, as shown in Figs. 5, 6, and 7. These proportional flow diagrams show how organizations in the partitions of one grouping are distributed into partitions of the other grouping. Thickness of connections corresponds to the number of organizations which are present in both observed partitions.

Figure 5 indicates that organizations declare similar goals and activities across different categories from the Transparency Register. Figure 6 shows that three communities extracted from the network of similar preference patterns (*2-Prospectus Directive*, *3-Building a capital market union*, and *5-Retail fin. services/Personal pension framework*), almost do not include organizations declaring their interest in Advisory and/or Representation, as they are mostly

---

[9] Sankey diagrams (https://developers.google.com/chart/interactive/docs/gallery/sankey) are based on work by Google (https://developers.google.com/terms/site-policies).
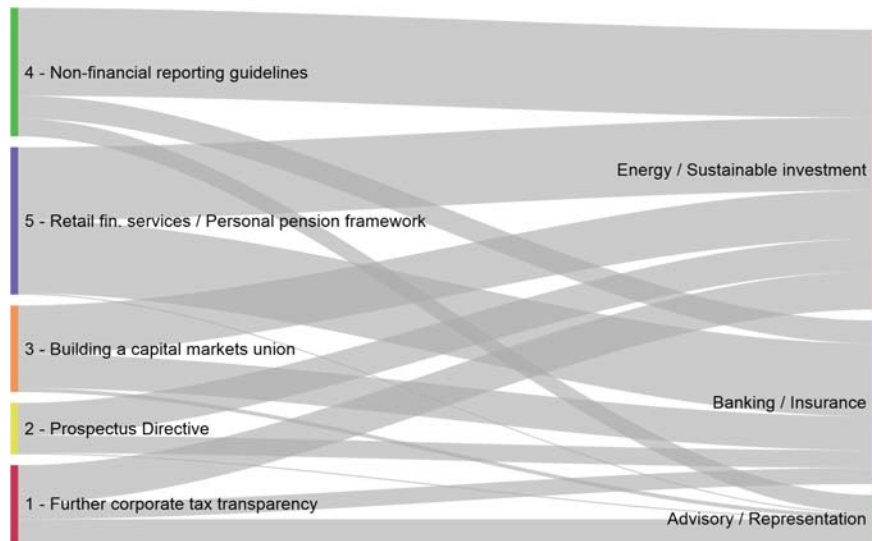
Fig. 6. Correspondence between the communities detected in the network of similar preference patterns (left), and the clusters of goals and activities (right)
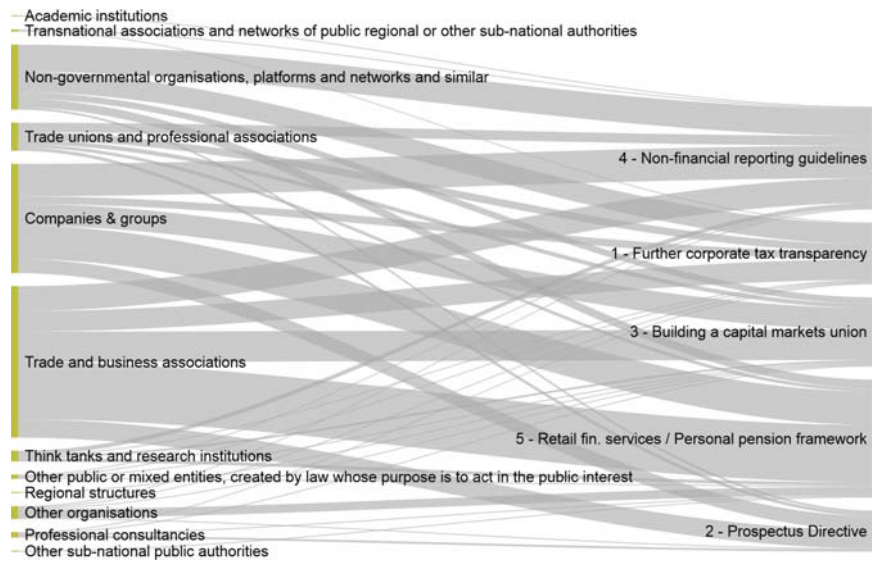

Fig. 7. Correspondence between the subcategories in which the organizations are registered (left), and the communities detected in the network of similar preference patterns (right)

members of communities *1-Further corporate tax transparency* and *4-Non-financial reporting guidelines*. Figure 7 shows how are organizations from the categories scattered into communities, which explains also the lowest B^3 scores in

Fig. 4. While for most communities the majority of members are from the largest two categories (*Companies & Groups*, and *Trade and business associations*), the *Non-governmental organizations* and the *Trade Unions and professional associations* are substantially represented in communities *1-Further corporate tax transparency* and *4-Non-financial reporting guidelines*.

5   Conclusion

We presented how can publicly accessible information be used to assess the positions and leanings of major stakeholders in the policy creation process. We focused on policy reforms in the area of *Banking and Finance*, and used data from the EC's Transparency Register, and the EC's public consultations. By combining methods from information retrieval, text mining, and network analysis we studied different aspects of the lobby organizations which engage in policy shaping.

Our analysis showed that the categories representing the type of an organization do not align with the clusters based on their declared goals and activities, but can be divided in up to three main clusters in terms of their field of interest. Furthermore, from the organizations' consultation responses we constructed a similarity network representing inter-organization policy preferences. The community structure of this network reveals information about organizations' activity that cannot be obtained from the organizations' self-description of their goals and activities. This implies that the network analysis adds an important aspect that is complementary to text analysis in the understanding of how lobby organizations engage in the policy making process.

Our finding suggest that if we want to build a map of the policy making arena we should categorize lobby organizations based on their responses to policy issues via the consultations, rather than based on their general self-declared goals and activities, or based on their formal organization type categorization. This work represents only the first step of a novel approach towards building maps of the policy arena. Future work will analyse how the design of the consultations could be improved in order to better identify the positions of the stakeholders with respect to the policy issues. The insights from this type of analysis and its future development can support the current EU policy agenda on increasing the transparency of the policy making process by enabling stakeholders and citizens to better understand which interests the various organizations represent and how they are influencing the policy debates.

References

1. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. Inf. Retrieval **12**(4), 461–486 (2009)
2. Bagga, A., Baldwin, B.: Entity-based cross-document coreferencing using the vector space model. In: Proceedings of 17th International Conference on Computing Linguistics (COLING), pp. 79–85. ACL (1998)
3. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media (2009)
4. Berkhout, J., Carroll, B.J., Braun, C., Chalmers, A.W., Destrooper, T., Lowery, D., Otjes, S., Rasmussen, A.: Interest organizations across economic sectors: explaining interest group density in the European union. J. Eur. Public Policy **22**(4), 462–480 (2015)
5. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. **2008**(10), P10008 (2008)
6. Coen, D., Katsaitis, A.: Chameleon pluralism in the EU: an empirical study of the European commission interest group density and diversity across policy domains. J. Eur. Public Policy **20**(8), 1104–1119 (2013)
7. Feldman, R., Sanger, J.: Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, New York, NY, USA (2006)
8. Hartigan, J.A.: Clustering Algorithms. Wiley (1975)
9. Lambiotte, R., Delvenne, J.C., Barahona, M.: Laplacian dynamics and multiscale modular structure in networks (2009). https://arxiv.org/abs/0812.1770
10. Martin, S., Brown, W.M., Klavans, R., Boyack, K.W.: OpenOrd: an open-source toolbox for large graph layout. In: Proceedings of SPIE 7868, Visualization and Data Analysis (2011)
11. Rasmussen, A., Carroll, B.J., Lowery, D.: Representatives of the public? public opinion and interest group activity. Eur. J. Polit. Res. **53**(2), 250–268 (2014)
12. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
13. Van Rijsbergen, C.: Information Retrieval. Butterworth, London, UK (1979)
14. Wolf, M., Haar, K., Hoedeman, O.: The fire power of the financial lobby: a survey of the size of the financial lobby at the EU level. Corporate Europe Observatory, The Austrian Federal Chamber of Labour and The Austrian Trade Union Federation (2014). http://corporateeurope.org/sites/default/files/attachments/financial_lobby_report.pdf
15. Zeng, A., Battiston, S.: The multiplex network of EU lobby organizations. PloS one **11**(10), e0158062 (2016)