Artif Intell Law (2018) 26:23–47
https://doi.org/10.1007/s10506-017-9204-y

ORIGINAL RESEARCH

Network approach to the French system of legal codes part II: the role of the weights in a network

Romain Boulet[1] · Pierre Mazzega[2] · Danièle Bourcier[3]

Published online: 12 August 2017
© Springer Science+Business Media B.V. 2017

Abstract Unlike usual real graphs which have a low number of edges, we study here a dense network constructed from legal citations. This study is achieved on the simple graph and on the multiple graph associated to this legal network, this allows exploring the behavior of the network structural properties and communities by considering the weighted graph and see which additional information are provided by the weights. We propose new measures to assess the role of the weights in the network structure and to appreciate the weights repartition. Then we compare the communities obtained on the simple graph and on the weighted graph. We also extend to weighted networks the amphitheater-like representation (exposed in a previous work) of this legal network. Finally we evaluate the robustness of our measures and methods thus taking into account potential errors which may occur by getting data or building the network. Our methodology may open new perspectives in the analysis of weighted networks.

Keywords Weighted network · Structural measures · Graph · Legal code · Codified legal system · Communities

✉ Romain Boulet
romain.boulet@univ-lyon3.fr

Pierre Mazzega
pierre.mazzegaciamp@get.omp.eu

Danièle Bourcier
daniele.bourcier@cersa.cnrs.fr

[1] Univ. Lyon, Jean Moulin, iaelyon, Magellan Research Center, 6 Cours Albert Thomas, 69008 Lyon, France

[2] UMR5563 Geosciences Environment Toulouse, CNRS, University of Toulouse, 14 Avenue E. Belin, 31400 Toulouse, France

[3] CERSA CNRS, Université de Paris 2, 10 Rue Thénard, 75005 Paris, France

1 Introduction

1.1 Motivations

Complexity in real networks often lies on a high number of vertices and these networks are large and sparse. We propose here to study another kind of network: a small and dense network; complexity does not rely on a large number of vertices but on a high number of links. The network we consider in this paper is the network of French Legal Codes (FLC network for short) representing the dependences (that is the citations) between the different French legal codes (Mazzega et al. 2009a, b). The nodes of this network are 52 French legal codes such as the environmental code or the code of education (see Table 11 of Appendix for a complete list). There is a link between two codes if one code explicitly refers to the other code by a citation, for instance there is a link between the environmental code and the mining code because article L211–10 of the environmental code refers to article L413–1 of the mining code: "Notwithstanding the provisions of Article L413–1 of the mining Code, the samples, documents and information of interest for research, production or the behaviour of underground water fall immediately into the public domain."[1] In this paper these links are considered as undirected but weighted: the weight of the link between X and Y is the number of citations between the two codes, that is the number of times code X cites or is cited by code Y.

Although this network is a weighted network, the structure of the underlying simple graph should not be neglected as it can be seen as the skeleton of the network, this analysis previously performed (Boulet et al. 2011), referred as Part I in this paper, is taken into account in this paper.

Performing a network analysis on the simple graph and on the weighted graph would allow knowing whether the weights of the links play an important role in the structure of the network and, if it is the case, where this role is emphasized. In this paper we first define new indices to analyze the role played by these weights on the skeleton (Sect. 2). Then in Sect. 3 we extend to weighted graphs the partition processes performed in Part I and analyze the behavior of the communities previously highlighted. Section 4 is devoted to the study of robustness by introducing small changes to the weights. A deeper analysis of the role played by the weights in our network is made in Sects. 5 and 6 concludes the paper.

From a legal point of view this kind of approach is quite new. This study is performed with the French Legal Codes but can also be done with other legal networks or with another set of codes such that the Italian Legal codes or any country whose codification process is well advanced. Moreover the approach introduced in this paper and the computation of the new measures we define can be applied to other kinds of networks such as social networks or biological networks.

[1] In French: « Nonobstant les dispositions de l'article L. 413–1 du code minier, les échantillons, documents et renseignements intéressant la recherche, la production ou le régime des eaux souterraines tombent immédiatement dans le domaine public ».

1.2 Previous results

Although, on the one hand, the problem of legal complexity has been and is often dealt (Schuck 1992) and on the other hand the development of graph-based tools is developing (Brandes and Erlebach 2005), we have seen since the last decade the analysis of networks (whether from a mathematical point of view or from a computer science point of view) interact with the legal sciences (Bourcier and Mazzega 2007, Fowler et al. 2007): a network being an object modeling interactions, it is useful for modeling citations between legal texts. In this approach we find the use of networks to study an aspect of the complexity of Law (Katz and Bommarito 2014), to visualize this complexity (Winkels and Boer 2014) and they are even now regarded as an object of study and not only as a model (Tarissan and Nollez Goldbach 2015). Workshops like *Network Analysis in Law* (NAIL) and surveys (Whalen 2016) are now devoted to this field of research. In this paper we study the weighted graph FLC and, in order to perform a comparison with the simple graph and emphasize the role played by the weights, the sketch of the analysis is the same as the one performed in Part I on the simple graph. We first have a look on structural indices the definitions of which have been changed to take the weights into account and we introduce new indices to assess the role played by the weights in this network. Then we look for a rich-club and continue the analysis with the research of communities using the same algorithms and methods, but extended to weighted networks, than the ones used in Part I.

1.3 Usual notions in graph theory

For the reader not acquainted with graph theory, we remind here the meanings of some terms used in this study. The usual notions of network analysis present in this paper can be found in reference books (such as Brandes and Erlebach 2005).

- A *graph* $G(V,E)$ is defined by a set of vertices V and a set of edges E, an edge linking two vertices.
- A *sub-graph* induced by a subset $W \subset V$ of the set of vertices of the graph G is the graph whose set of vertices is W and the set of edges is constituted by all the edges of G linking two elements of W.
- Two vertices are *neighbors* if they are linked by an edge.
- The *neighborhood* Γ_v of a vertex v, is the set of neighbors of v.
- The *degree* of a vertex is the number of its neighbors, that is $|\Gamma_v|$.
- The *density* of a graph is the ratio between the number of edges that actually exists in the graph and the total number of possible edges of the graph. Therefore, the density is given by $2m/[n(n-1)]$ where n is the number of vertices of the graph and m is the number of edges.
- A *path* between two vertices v_1 and v_k is a sequel of different vertices $(v_1, v_2, \ldots, v_k)$ where the vertices v_j and v_{j+1} (for $j < k$) are linked by an edge. The *length* of the path is $(k-1)$.

- A *shortest path* between two vertices is a path of minimal length between these vertices. Several shortest paths can exist between two given vertices (with the same length).
- The *average path length* is the mean of the length of the shortest paths between each pair of vertices (Watts and Strogatz 1998).
- The *characteristic path length* of a graph is the median of the means of the shortest paths from a vertex to the other vertices (Watts and Strogatz 1998).
- The *diameter* of a graph is the longest of the shortest paths between two any vertices of the graph.
- A *clique* of a graph G is a complete sub-graph of G that is a sub-graph in which all the vertices are pair-wise linked by an edge. A clique with k vertices is also called a k-clique. A clique is maximal if it is not contained into another clique.
- A *connected component* of a graph G is a sub-graph H of G where there is a path between any two vertices of H and no path between a vertex of H and a vertex of G not in H.
- An *Erdös-Rényi random graph* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, denoted by $G(n, p)$, is a graph with n vertices where each edge exists with a uniform probability p (Erdös and Rényi 1959).
- The adjacency matrix of a graph G with n vertices labeled from 1 to n is the matrix whose (i, j)th entry is 1 if there is an edge between the vertices i and j and 0 otherwise.
- A *rich-club* of a network is a strongly interconnected group of nodes with highest degree. A rich-club may not exist if nodes with highest degree are not connected.

1.4 Basic notions in statistics

This paper will make use of some basic notions of univariate statistics which are recalled here.

Let $X_1, X_2, \ldots, X_n$ be a dataset of n observations. Descriptive statistics provides measures to describe a dataset. Two measures are frequently used to summarize a sample: central tendency and dispersion (or variability). A first pair of measurements consists of the mean and the standard deviation. The mean, denoted by $\bar{X}$, is given by $\frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ that is $\frac{1}{n}\sum_{i=1}^{n} X_i$. The standard deviation is the square root of the variance defined by $\frac{1}{n}\left((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2\right)$ that is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. This measures the dispersion around the mean. A second pair of measurements is given by the median and the interquartile range. The median is the value Q_2 for which half of the X_i is greater than or equal to Q_2 and half of the X_i is lower than or equal to Q_2. The interquartile range is the difference between the third quartile (the value Q_3 for which 75% is lower than or equal to Q_3) and the first quartile (the value Q_1 for which 25% is lower than or equal to Q_1).

Analyzing the link between two variables X and Y could be made through the computation of the covariance given by $cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$ but this measure is not easy to interpret and we prefer the correlation coefficient denoted by r and given by covariance between X and Y divided by the product of the standard deviation of X and the standard deviation of Y. The correlation coefficient lies between -1 (perfect linear relationship of two variables evolving in the opposite direction) and $+1$ (perfect linear relationship of two variables evolving in the same direction). In case of a correlation coefficient significantly non-zero one can perform a linear regression, that is computing the best line explaining variable Y in function of variable X and the slope of this line is called the regression coefficient.

2 Structural analysis of the weighted graph

To study the structure of a weighted network one can perform a structural analysis on the underlying simple graph, this is done for instance in (Grossman 2002) or (Latapy and Magnien 2006) and this is what we have done in Part I. In this section we mainly describe how the structure of the FLC skeleton affects the weight arrangements.

First of all we measure central tendencies and dispersion of the weights like mean and standard deviation or median and quartiles (Table 1). The first observation is that these weights are heterogeneous (they are neither uniform nor Gaussian) with significantly high values (the mean is considerably higher than the median) and spreading out over a large range of values (the dispersion values like standard deviation or interquartile range are high compared with their respective position index: mean and median).

The cumulative distribution is then plotted (Fig. 1) and we remark that it follows a truncated power law, that is, a power-law with a higher decrease for high values.

In order to evaluate the influence of the graph structure on the weight distribution we introduce a set of new measures (graphical or quantitative). In order to have a baseline and to be sure that the phenomenon described by this indices is not due to random, we perform the same measures on a graph the weights of which are assigned randomly, that is a graph whose underlying simple graph is the same than the considered graph (here the FLC network) and the set of weights $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ are randomly assigned to the edges. In practice we average the values obtained on a high number of such randomly generated graphs.

2.1 The degree attractiveness

The first measure we propose relates weight distribution according to degree of vertices. Of course a vertex with a high degree in a simple graph tends to have a high weighted degree; for this reason it is better to measure the average weight on edges incident to a vertex and to plot it in function of the simple degrees. An increasing function means that weights tend to reinforce the high degrees. A decreasing function means that vertices with a low simple degree are in fact vertices with a higher number of relations than expected by measuring the simple degree; in this case the weights give strength to these vertices. In order to sum up this graphic

Table 1 classical univariate indices of the weights of the FLC network

Minimum	Maximum	Mean	Standard deviation	1st quartile	Median	3rd quartile	Interquartile range
1	385	14.99	34.38	2	4	14	12

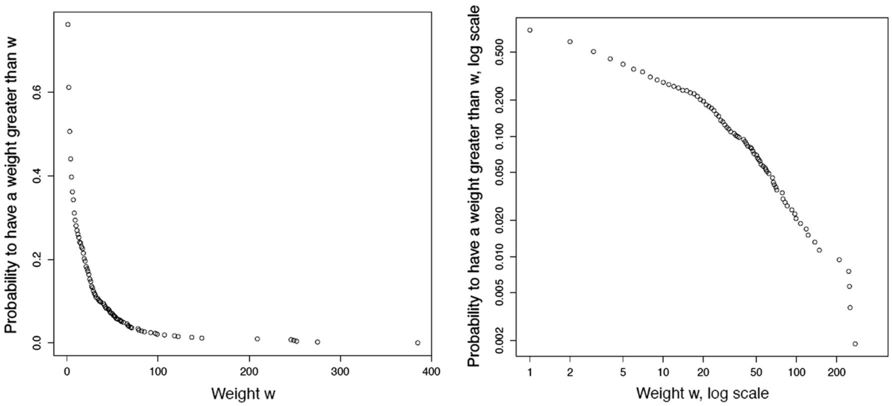

Fig. 1 Cumulative weight distribution (*left* standard scale, *right* log–log scale)

into a single index we compute the linear regression coefficient which we call the *degree attractiveness*. If the degree attractiveness is positive then the weights tends to reinforce the high degrees, if it is negative then they give strength to low-degree vertices. Another way to check the importance of this relation is to compute the correlation coefficient between average weight on edges incident to a vertex and the simple degree of this vertex; the value of the linear regression coefficient is relevant only if the correlation coefficient is high.

We can establish a formula for the degree attractiveness (hereafter denoted by a) involving the degrees and the weights. Let d_i be the degree of vertex i and let $w_{i,j}$ be the weight of the edge linking vertex i and vertex j (this weight is zero if there are no edges between i and j). We denote by n the number of vertices in the graph, by m the number of edges in the simple underlying graph, by $\bar{d}$ the mean degree (which is equal to $\frac{2m}{n}$), by $\bar{w}$ the mean weight (which is equal to $\frac{\sum w_{i,j}}{m}$). We recall that the regression coefficient between X and Y is given by $\frac{\mathrm{cov(X,Y)}}{\mathrm{var}(X)}$ where variance and covariance could be rewritten in $\mathrm{cov}(X,Y) = \frac{1}{n}\sum_i X_i Y_i - \left(\frac{1}{n}\sum_i X_i\right)\left(\frac{1}{n}\sum_i Y_i\right)$ and $\mathrm{var}(X) = \frac{1}{n}\sum_i X_i^2 - \left(\frac{1}{n}\sum_i X_i\right)^2$. In our case,

$$X_i = d_i$$

and

$$Y_i = \frac{1}{d_i}\sum_{j=1}^{n} w_{i,j}$$

so

$$a = \frac{\frac{1}{n}\sum_i\left(d_i\frac{1}{d_i}\sum_j w_{i,j}\right) - \left(\frac{1}{n}\sum_i d_i\right)\left(\frac{1}{n}\sum_i\frac{1}{d_i}\sum_j w_{i,j}\right)}{\frac{1}{n}\sum_i d_i^2 - \left(\frac{1}{n}\sum_i d_i\right)^2}$$

that is

$$a = \frac{2m\bar{w} - \frac{2m}{n}\sum_{i,j}\frac{w_{i,j}}{d_i}}{\sum_i d_i^2 - \frac{4m^2}{n}}$$

and we have:

$$a = \frac{\bar{w} - \frac{1}{n}\sum_{i,j}\frac{w_{i,j}}{d_i}}{\frac{1}{2m}\sum_i d_i^2 - \bar{d}}$$

In the general case, the degree attractiveness of a random graph with random edges is zero (or is not significantly different from zero). We randomly generated 100,000 graphs with n vertices (n chosen uniformly at random between 50 and 500) and m edges (m is such that the density of the graph is chosen at random between 0.01 and 0.5), each graph having edges weight (each weight is chosen uniformly at random between 1 and 40). The average degree attractiveness on these random graphs is 5.9×10^{-4}.

The results for the FLC network are depicted on Fig. 2 and the degree attractiveness is 0.69, it is positive (and significantly different from zero, p value $<0.1\%$) so the weights of the edges are "attracted" by edges incident to high-degree vertices. In our context that means that codes linked to a high number of other codes tend to have more citations (in average) than codes linked to a fewer number of other codes. To have a baseline, we computed this coefficient on 100,000 graphs with weights randomly distributed, that is the simple graph is the same as the FLC network and the set of weights are the same as the FLC network but they are randomly distributed on the edges; we found a degree attractiveness of 1.5×10^{-5} (with a correlation coefficient 0.032).

2.2 The clustering attractiveness

Now we want to relate the effect of clustering on this weight distribution: is the weight distribution attracted by clustered regions of the graph or has the clustering no effect on weight distribution? Remind that a way to measure clustering coefficient in a graph is to compute the ratio between the number of triangles and the number of connected triples. We compute the average weight of edges belonging to triangles (Eq. 1) and the mean weight of edges belonging to connected

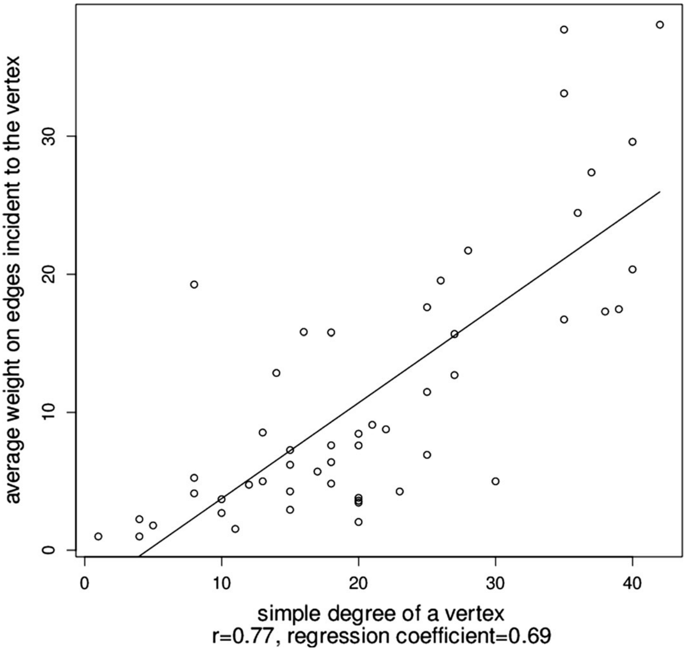

Fig. 2 Degree attractiveness: plot of average weight on edges incident to a vertex in function of the simple degree of this vertex

Table 2 Weight distribution on triangles and connected triples, between brackets the same value for a graph with the same skeleton as FLC but with randomly assigned weights obtained by averaging the values on 1000 generated such graphs

Mean weight of edges belonging to a triangle	21.72 (14.95)
Mean weight of edges belonging to a connected triple different from a triangle	13.09 (15.00)
Ratio	1.66 ($\approx$1)

triples that are not a triangle (Eq. 2). We then define the *clustering attractiveness* to be the ratio between these two measures. In a graph with randomly allocated weights on the edges, this ratio is about 1. So if the clustering attractiveness is close to 1 then clustered regions have no effect on the weight dispersion; on the contrary a ratio greater than 1 (resp. lower than 1) shows that clustered regions attracts (resp. repels) the weights.

For the weighted FLC network, this ratio is 1.66 (see Table 2). That means that the average weight is increased by 66% when there is a closure of a connected triple; a triangle attracts more weights than a connected triple different from a triangle.

$$\frac{1}{|T|}\sum_{\{a,b,c\}\in T}\frac{1}{3}\left(w_{a,b}+w_{a,c}+w_{b,c}\right) \tag{1}$$

where T is the set of triangles. Equation 1: average weight of edges belonging to triangles.

$$\frac{1}{|\hat{T}|}\sum_{(a,b,c)\in\hat{T}}\frac{1}{2}\left(w_{a,b}+w_{a,c}\right) \tag{2}$$

where $\hat{T}=\{\{a,b,c\}: a\sim b, a\sim c, b\nsim c,\}$. Equation 2: mean weight of edges belonging to connected triples that are not a triangle.

2.3 The clique attractiveness

Given that a triangle is a clique of size 3, we can extend the idea of clustering attractiveness by computing the mean weight of edges belonging to (maximal) cliques (see Tables 3 and 4), by analogy with the previous measures we call this set of computations the *clique attractiveness*. In order to make comparisons, the same computations have been done on a set of 1000 graphs whose weights have been randomly assigned to the edges, with these experimental results, it turns out that the mean weight of the edges belonging to a clique is between 14.45 and 15.08 whatever the size of the considered clique; this hardly comes as a surprise since the mean weight (of all the edges in the graph) is 14.98.

For the FLC network, we remark that the mean weight of the edges is an increasing function of the size of the clique. That means that the weights are attracted by the cliques, that is, more a region of a graph is clustered, more the mean weight will be important.

Table 3 Mean weights of edges belonging to a clique, between brackets the same value for a graph with the same skeleton as FLC but with randomly assigned weights obtained by averaging the values on 1000 generated such graphs

Size of the clique	Number of cliques	Mean weight of the edges	Size of the clique	Number of cliques	Mean weight of the edges
2	231	14.99 (14.99)	8	15,943	38.23 (14.95)
3	2680	21.72 (14.96)	9	7979	40.19 (14.94)
4	8144	26.63 (14.95)	10	2745	41.97 (14.94)
5	16,369	30.42 (14.95)	11	610	43.61 (14.94)
6	22,811	33.47 (14.95)	12	77	45.19 (14.93)
7	22,559	36.02 (14.95)	13	4	46.76 (14.92)

Table 4 Mean weights of edges belonging to a maximal clique, between brackets the same value for a graph with the same skeleton as FLC but with randomly assigned weights obtained by averaging the values on 1000 generated such graphs

Size of the maximal clique	Number of maximal cliques	Mean weight of the edges	Size of the maximal clique	Number of maximal cliques	Mean weight of the edges
2	1	1.00 (14.45)	8	32	32.22 (15.01)
3	0		9	60	29.28 (14.98)
4	4	22.04 (14.96)	10	55	35.10 (14.99)
5	15	23.67 (15.07)	11	45	37.19 (14.96)
6	13	24.94 (15.08)	12	25	42.74 (14.96)
7	20	25.35 (15.03)	13	4	46.76 (14.92)

2.4 The shortcut attractiveness

The last index we introduce is related to the small-world effect (Watts 2003) and in particular to the short length of paths linking any two vertices. A first idea would be to compute the average path length on geodesics[2] but this would not give relevant results mainly because each edge is a geodesic between the vertices it links and averaging among all edges has no interest. A better approach is to focus attention on edges playing an important role on this small world effect, that is, edges belonging a lot of times to geodesics. This is measured by the edge betweenness (Girvan and Newman 2002) whose principle is similar to that of vertex betweenness (Freeman 1979).

Plotting the edge weights in function of edge betweenness (especially edges with a high betweenness) allows visualizing relations between high edge betweenness and edge weights. If high weights lie on edges with a high betweenness then this means on one hand that shortcut edges attract the weights but on the other hand the shortcut aspect of these edges is weaken (a shortcut with weight $\mathbf{w}$ can be seen as $\mathbf{w}$ shortcuts of weight 1 and if $\mathbf{w}$ is large the importance of one shortcut is reduced because there are $\mathbf{w-1}$ other ones). On the opposite, if edges with a high betweenness only have weak weights then the central aspect and the importance of these edges is reinforced.

For the FLC network, Fig. 3 shows that we are in the case where edges with a high betweenness have low weights. This means these central edges have small weights which emphasizes the importance of the corresponding citations: they are weak ties but important to create shortcuts in the network.

[2] A geodesic is a shortest path between two vertices.

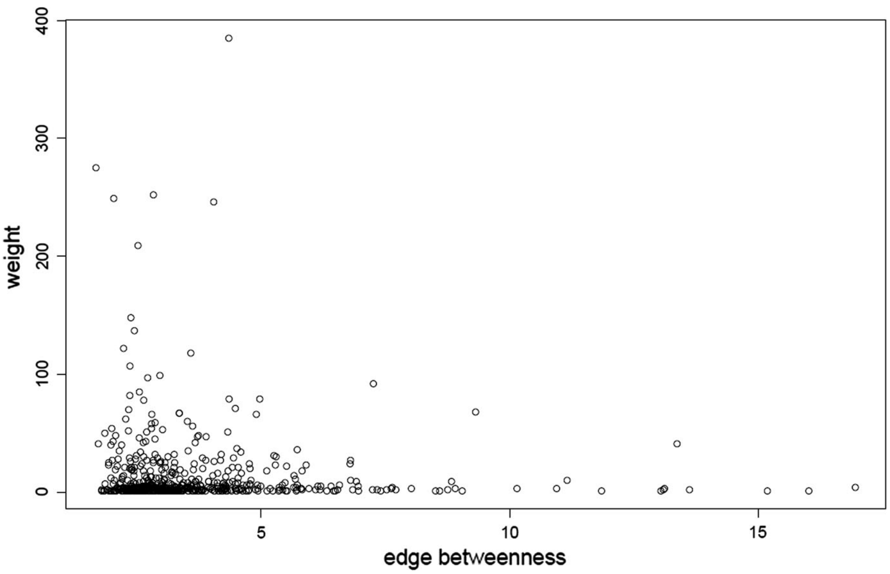

Fig. 3 Shortcut attractiveness: plot of edge weights in function of edge betweenness. We do not take into account the cut-edge linking IMN (a vertex with degree 1) and PEN owing to its huge betweenness due to its belonging to all the geodesics linking this vertex to another one (so its betweenness is 50; the weight allocated to this edge is 1)

3 Partitioning the weighted network

3.1 The rich-club

Nine codes among the ten codes of the rich club found in Part I belong to the ten codes with highest total weighted degree (Table 5). We can say that vertices with high degree in the skeleton graph still have a high degree in the weighted network; a code with a high simple degree is a code having a high capacity to make links with several other codes so it is not surprising that they also have a high weighted degree. Moreover we showed in Sect. 2.1 that the network has positive degree attractiveness involving that a vertex with a high simple degree is very likely to have a high weighted degree.

We may remark that the ENV[3] code is the eleventh code with highest total degree; due to this ranking and his belonging to the rich-club in the "skeleton" network (remind that the ENV code has a high centrality measure), we consider that the ENV code is still an important code. Another difference induced by the weights is the presence of the MOF code: MOF was the twelfth code with highest degree. The density of the underlying simple sub-graph induced by the eleven codes (PEN, SSC, SAP, TRA, RUR, COM, GCT, PPE, CIV, MOF, ENV) equals 96% and is still high; this group of eleven codes constitutes the rich-club of the FLC network.

[3] The correspondence between the acronyms and the (translated) names of the codes are given in Appendix Table 11.

Table 5 The 10 codes with highest weighted degree

Total weighted degree	
PEN	1600
SSC	1321
SAP	1184
TRA	1159
RUR	1013
COM	880
GCT	814
PPE	681
CIV	657
MOF	608

In order to study how the weights enrich the rich-club structure we search a core from another point of view: we fix a threshold t and delete all the arcs having a weight lower than t. For a large t, only edges with a high weight remain and it is interesting studying the remaining components. Figure 4 represents the density and the size of the largest connected component, two plates (corresponding to two cores) emerge: one with two vertices (the PEN and SAP codes) and one with six codes (the PEN, SAP, PPE, RUR, SSC and TRA codes). These last six codes can be considered as a tight core of the rich-club.

As a result, on a network with positive degree attractiveness, the rich club can be established into two steps: first extracting the *club* (a highly interconnected sub-graph) of vertices with highest simple degree (the *rich* vertices) and then exploring the cores in that club by using the weighted degrees as exposed here above.

3.2 Communities

In Part I, in order to face with the problem of high density, we first deleted vertices of the rich-club (which can be seen as a community). Then we used three algorithms to detect communities in the remaining non-weighted network and we gather the results and extract stable (or robust) communities that are set of vertices which are always put together whatever the algorithm used. As a result we obtained two main stable communities, one related to “codes for territories and resources”[4] and one related to “codes for social issues”.[5]

The walktrap algorithm (Pons and Latapy 2005) and the fast-greedy algorithm (Clauset et al. 2004; Newman 2004) used in Part I can be generalized to weighted graphs by maximizing the modularity, the definition of which is extended to weighted graphs (Eq. 3).

[4] Containing the codes of CHA (Housing), DOE (State-owned property), DPF (Public rivers), EUP (Expropriation in public interest), FOR (Forestry), GPP (Property legal person), MIN (Mining), PAT (Estate), PMA (Seaports), URB (Urbanism), VOR (Road system).

[5] Containing the codes of ART (Handicraft), ASF (Social service), ASS (Insurance), CNS (Consumer), EDU (Education), JUA (Administrative court), JUF (Financial court), MUT (Mutual society), REC (Research), ROU (Traffic), SPO (Sport), TOU (Tourism).

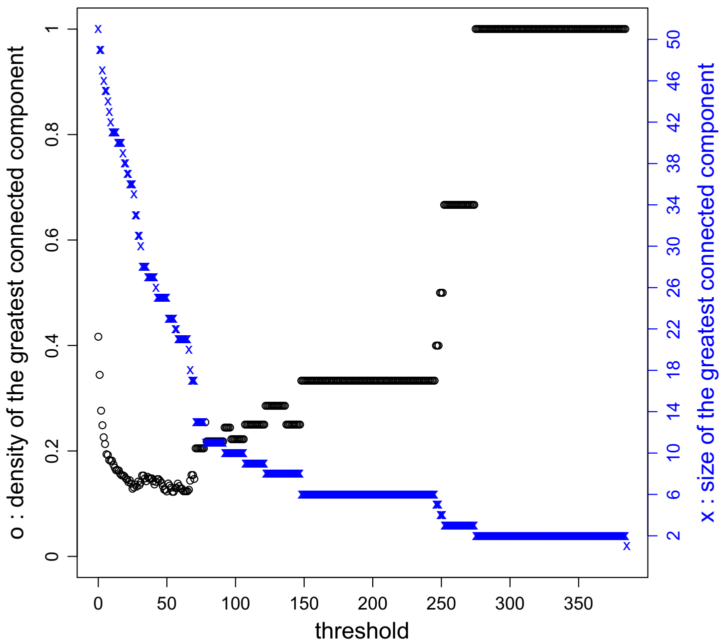

Fig. 4 Density and size of the largest connected component of the graph obtained by deleting edges in increasing order of their weights

$$\sum_{i=1}^{k}\left[\frac{m_i}{m}-\left(\frac{\sum_{u\in V_i} d(u)}{2m}\right)^2\right] \tag{3}$$

Equation 3: Modularity for a weighted graph, m_i is the number of weights in V_i, m is the total sum of weights and $d(u)$ denotes the weighted degree of the vertex u.

The spectral partitioning (von Luxburg 2007) can be generalized to weighted graphs by using a generalization of the Laplacian matrix to weighted graphs (Chung 1997) defined by $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$ where $\mathbf{W}$ is the weighted adjacency matrix and $\mathbf{D}$ is the diagonal matrix of weighted degrees.

The communities obtained with these three algorithms on the weighted graph (where vertices of the rich club are deleted) are exposed in Table 6 where we remind the stable communities obtained on the simple graph. Then, from these communities we extract the stable groups (Table 7) for the weighted graph.

For the FLC network, we see that we have four main robust clusters when we consider the weighted graph instead of two with the simple graph; the weights thus provide a more detailed understanding of the network. The robust community related to territories and resources is roughly maintained whereas the codes related to social issues are divided into two groups.

Table 6 Communities found with three algorithms performed on the weighted network FLC without the rich-club (one line per community)

Walktrap	*CHA** *DOE** ELE *EUP** *FOR** *GPP** *MIN** *PAT** **TOU** *URB** *VOR**
	ASF$^{\circ}$ **EDU**$^{\circ}$ **JUF**$^{\circ}$ SDA **JUA**$^{\circ}$ **REC**$^{\circ}$ MPU **SPO**$^{\circ}$
	AVI DEF *DPF** DOU OGJ *PMA** PCI PIT
	ASS$^{\circ}$ **CNS**$^{\circ}$ **MUT**$^{\circ}$ PCO **ROU**$^{\circ}$
	FAS JUM SNA
	DMM TMA
	ART$^{\circ}$
Fast-greedy	*CHA** *DOE** *EUP** *FOR** *GPP** ICI *MIN** *PAT** **TOU**$^{\circ}$ *URB**
	AVI DEF DMM ELE FAS JUM SNA TMA
	ART$^{\circ}$ **ASS**$^{\circ}$ **CNS**$^{\circ}$ **MUT**$^{\circ}$ PCO **ROU**$^{\circ}$ *VOR**
	DPF* DOU MPU OGJ *PMA** PCI PIT
	ASF$^{\circ}$ **EDU**$^{\circ}$ SDA **JUF**$^{\circ}$ **JUA**$^{\circ}$ **REC**$^{\circ}$ **SPO**$^{\circ}$
Spectral	**ART**$^{\circ}$ AVI DEF DMM *DOE** ELE FAS ICI JUM MPU SNA TMA
	CHA* *EUP** *FOR** *GPP** *MIN** *PAT** **TOU**$^{\circ}$ *URB** *VOR**
	ASF$^{\circ}$ **EDU**$^{\circ}$ SDA **JUF**$^{\circ}$ **JUA**$^{\circ}$ **REC**$^{\circ}$ **SPO**$^{\circ}$
	DPF* DOU OGJ *PMA** PCI PIT
	ASS$^{\circ}$ **CNS**$^{\circ}$ **MUT**$^{\circ}$ PCO **ROU**$^{\circ}$

Codes are mentioned by their acronym, we also recall the 2 stable communities found in Part I: the codes for social issues are in bold and are followed by the degree symbol, the ones related to territories and resources are in italic and are followed by an asterisk. Codes in roman black were not in one of these two communities

Table 7 Stable communities which ensue from the three partitionings of the weighted network (one line per stable community)

CHA* *EUP** *FOR** *GPP** *MIN** *PAT** TOU$^{\circ}$ *URB**
ASF$^{\circ}$ **EDU**$^{\circ}$ SDA **JUF**$^{\circ}$ **JUA**$^{\circ}$ **REC**$^{\circ}$ **SPO**$^{\circ}$
DPF* DOU OGJ *PMA** PCI PIT
ASS$^{\circ}$ **CNS**$^{\circ}$ **MUT**$^{\circ}$ PCO **ROU**$^{\circ}$
FAS JUM SNA
DMM TMA
AVI DEF

Codes are mentioned by their acronym, we also recall the 2 stable communities found in Part I: the codes for social issues are in bold and are followed by the degree symbol, the ones related to territories and resources are in italic and are followed by an asterisk

3.3 Amphitheater-like representation

The last step of our analysis is to represent the vertices of the FLC network on an amphitheater. For this purpose we need to endow the network with a dissimilarity taking into account the structure of the network (in particular the weights of the links) and, in order to maintain the same approach as that performed in Part I, this dissimilarity may be a generalization of the Czekanovski-Dice dissimilarity given in Part I. According to (Jouve et al. 2002) the Czekanovski-Dice dissimilarity can be rewritten

$$\delta^2(i,j) = 1 - 2\frac{\langle a_i, a_j \rangle}{\|a_i\|^2 + \|a_j\|^2}$$

where a_i is the **i**th column of the adjacency matrix **A** of a simple graph and <a, b> denotes the dot product between columns a and b and $\|a\|^2 = \langle a, a \rangle$.

We extend naturally this dissimilarity to weighted graphs:

$$\delta^2(i,j) = 1 - 2\frac{\langle w_i, w_j \rangle}{\|w_i\|^2 + \|w_j\|^2}$$

where w_i is the **i**th column of the weighted adjacency matrix **W** of a graph.

Applying the method described in Part I we obtain the amphitheater-like representation drawn in Fig. 5.

4 Robustness of the results

By nature it is unlikely that there are errors in the construction of the unweighted graph associated to the French legal system (hypercode): it is enough to identify and check the existence of an occurrence of a citation between codes X and Y to draw the edge between the two vertices representing these codes. The construction of the weighted graph is more prone to errors because these quotations and references are expressed in natural language and therefore give rise to a diversity of forms relevant

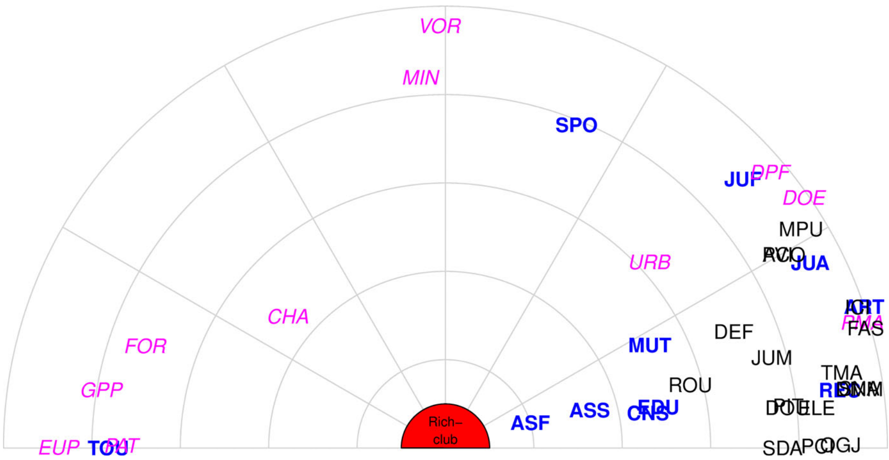

Fig. 5 The amphitheater-like representation of the weighted FLC network with the stable communities found in Part I (unweighted network): the codes for social issues are in *blue* and **bold**, the ones related to territories and resources are in *pink* and *italic*. (Color figure online)

to the construction of the weighted graph but difficult to identify exhaustively. We do not estimate the reliability of the automatic identification of citations with criteria such as recall and precision or sensitivity and specificity but we present in this section an analysis of the robustness of our results. The allocation of a citation to a misidentified couple of codes, if it can happen, should be rare; the most likely source of error in constructing the weighted graph is the omission of citations resulting from the diversity of expression allowed by the natural language.

Omission errors result mostly in the underestimation of the weights associated to some edges. We also do not know which edges are underweighted in the graph. In order to test the robustness of the results presented in the previous sections, we perturb the weighted graph, perform its network analysis and search for the previously identified structures (communities) that remain stable. This procedure was performed for a set of perturbations, each generated perturbation being characterized by two parameters: **p** is the percentage of edges in the graph whose weights are disturbed; **N** is the maximum amplitude of the weight perturbation. Indeed we generate the value of the perturbation of the weight of an edge as a random variable with uniform distribution with support on the real segment **[−N/**4;**N].**

Owing to the execution time of some calculations, especially to compare communities, we focus on four sets of parameters which may reflect the human errors of data entries. These parameters are:

−**p** = 5% and **N** = 10.
−**p** = 5% and **N** = 20.
−**p** = 10% and **N** = 10.
−**p** = 10% and **N** = 20.

In the following subsections we first analyze the structural indices of the perturbated graphs and then we compare the communities found while partitioning these perturbated graphs. Performing only one permutation for each set of parameter is not relevant: we perform several perturbations for each set of parameters.

4.1 Structural indices of perturbated graphs

In this section, we compute on the randomly perturbated graphs the indices introduced in Sect. 2 in order to assess their robustness. We remark that there are no changes on the degree attractiveness: the average correlation coefficient and the average regression coefficient computed on 10,000 randomly perturbated graphs remains the same except for the last set of parameters ($p = 0.10$ and $\mathrm{N} = 20$) where the average correlation coefficient is 0.76 instead of 0.77. The triangle attractiveness (Table 8), is not clearly altered (there is a slight decrease, less that 2.5%, occurring with strong perturbations $p = 10\%$ and $\mathrm{N} = 20$). Similarly, the clique attractiveness is not clearly changed: the absolute maximal difference between the measure on the unperturbated graph and the average attractiveness in the perturbated graph is 0.8 (the maximum relative difference is <3.5%). The same is done for the shortcut attractiveness: it occurs that, for an edge, the maximum difference between its average weight in the perturbated graph and its weight in the unperturbated graph is 1.14.

Table 8 Weight distribution on triangles and connected triples of perturbated weighted graphs (for each set of parameters, 1000 randomly perturbated graphs are generated)

	$p = 0$ N = 0	$p = 0.05$ N = 10	$p = 0.05$ N = 20	$p = 0.10$ N = 10	$p = 0.10$ N = 20
Mean weight of edges belonging to a triangle	21.72	21.92	22.11	22.11	22.51
Mean weight of edges belonging to a connected triple different from a triangle	13.09	13.29	13.48	13.48	13.87
Ratio	1.66	1.65	1.64	1.64	1.62

4.2 Partitioning perturbated graphs

We first check if, when perturbating the graph, the rich-club remains the same; in other words, are the first eleven codes with highest degree the same? To answer this question, we perform 10,000 perturbations for each set of parameters and we compute, for each set of parameters, the percentage of perturbations having the same rich-club as the unperturbated graph. We obtain that the rich club is clearly robust to perturbations because, for each set of parameters, more that 99.9% of the perturbations have the same rich club as the one found in Sect. 3.1.

Now the most tedious point is to check whether the partitionings obtained are robust. For each partitioning algorithm (walktrap, fast-greedy, spectral) we perform several perturbations for each set of parameters and compare the communities obtained with the communities of the unperturbated graph. In order to be able to measure the differences between two partitionings, we define a distance between two partitionings.

The communities of the unperturbated graph are numbered from 1 to k which induces a labelling of the vertices according to their belonging in a community (a vertex belonging to community i has label i). Now we number the communities (and therefore label the vertices) of the perturbated graph in order to minimize the number of vertices the label of which in the unperturbated graph is different than its label in the perturbated graph. Then we shall use this number of vertices with different labels to define a 'distance' between two partitionings. Let P and P′ be the vectors where the ith component gives the community of the ith vertex. For instance

$$\mathrm{P} = (3,4,4,2,4,1,2,2,1,5,5,3,2,3,1,2,1,1,1,3,3,2,5,1,4,5,1,5,4,5,5,3,4,2,3,1,2,1,4)$$

$$\mathrm{P}' = (\mathbf{5},4,4,\mathbf{3},4,1,2,2,1,\mathbf{3},\mathbf{3},\mathbf{5},2,\mathbf{5},1,2,1,1,1,\mathbf{5},\mathbf{5},2,5,1,4,5,1,\mathbf{3},4,5,\mathbf{3},\mathbf{5},4,2,\mathbf{5},1,2,1,4)$$

Let $d_H(P,P')$ be the Hamming distance between P and P′, that is the number of positions at which the corresponding labels are different. In our example $d_H(P,P') = 12$. But we can relabel P′ by exchanging the roles of 3 and 5 (this does not affect the partitioning, we just change the names of the communities). Let σ be the permutation which transposes 3 and 5 and let P'_σ be the resulting labeling. We have:

$$P'_{\sigma} = (3,4,4,\mathbf{5},4,1,2,2,1,5,5,3,2,3,1,2,1,1,1,3,3,2,\mathbf{3},1,4,\mathbf{3},1,5,4,\mathbf{3},5,3,4,2,3,1,2,1,4)$$

and the Hamming distance between P and $P'_{\sigma}P'\ _{\sigma}$ is $d_H(P,P'_{\sigma}) = 4$.

As a result we define the distance[6] between two partitioning P and P′ by:

$$d(P,P') = \min_{\sigma \in S_k} d_H\left(P,P'_{\sigma}\right)$$

where k is the number of communities in P′ and S_k is the set of all permutations in {1,…, k}.

The results of these distance computations are shown in Table 9. The high measures obtained for the spectral partitioning can be explained by the fact that, in the unperturbated cases, there are two partitioning with a highest modularity: the first one (described in Table 6) has modularity 0.4247 and a second one has modularity 0.4241 and the distance between these two partitionings is 9. We remind that a step in spectral clustering consists in performing a k-means, but since the result of the k-means algorithm depends on its initialisation, several k-means are performed that is, for each simulation, 5000 k-means (with random initializations) are performed in order to choose the one with the greatest modularity. As a result the resulting partitioning of a perturbated graph may be close to one of these two partitionings.

Moreover, notice that if we exchange two vertices belonging to two different communities of size at least two then the distance between the partitionings is at least 2. Similarly, if a new community is created by taking three vertices from other communities then the distance between the partitionings is at least 3. Let us notice that even in the unperturbated graph, assignment of vertices into a community varies depending on the algorithm chosen for partitioning; this is the case for the electoral code (ELE) for instance. This phenomenon was also present at the study of unweighted graph and may be caused by the high number of links of the hypercode network.

Therefore to have a better idea of the robustness to perturbations of these partitionings we focus on stable communities, that is, group of vertices that are always gathered together whatever the partitioning algorithm used. These stable communities for the weighted unperturbated graph are listed in Table 7 and we focus on the first four stable communities which are the main ones. To measure the variability of these stable communities when we perturb the graph, we compute the average number of vertices which are absent from one of the four stable communities when we cluster the perturbated graph; results are shown on Table 10 and show a very low variability of these stable communities.

[6] It's not difficult to check that this indeed defines a distance from a mathematical point of view, that is an application verifying the axioms of symmetry, separation and triangular inequality.

Table 9 Mean distance between the partitioning of the unperturbated graph and the perturbated graphs

Parameters of the perturbations	Fast-greedy (100 simulations)	Walktrap (100 simulations)	Spectral (100 simulations)
$p = 0.05$ and $N = 10$	2.32	2.69	8.14
$p = 0.05$ and $N = 20$	4.14	4.84	8.96
$p = 0.10$ and $N = 10$	4.09	4.72	9.58
$p = 0.10$ and $N = 20$	6.97	7.39	10.96

Table 10 Mean number of vertices absent from one of the four stable communities (first four lines of Table 6) after perturbing the graph

Parameters of the perturbations	Fast-greedy (100 simulations)	Walktrap (100 simulations)	Spectral (100 simulations)
$p = 0.05$ and $N = 10$	0.07 0.12 0.08 0.29	0.04 0.33 0.24 0.28	0.01 0.23 0.09 1.19
$p = 0.05$ and $N = 20$	0.32 0.48 0.44 0.80	0.27 0.64 0.58 0.53	0.15 0.48 0.38 1.10
$p = 0.10$ and $N = 10$	0.16 0.42 0.38 0.54	0.15 0.42 0.42 0.35	0.01 0.40 0.33 1.00
$p = 0.10$ and $N = 20$	0.52 0.84 0.68 1.00	0.30 0.94 1.22 0.86	0.27 0.87 0.61 1.36

5 Comparison with previous results: the role of the weights

We studied in Part I the simple graph (*i.e.* unweighted) associated to the network of citations between French legal codes which can be seen as a skeleton of the structure of this network. By weighting the relations we add supplementary information about the structural organization of this network. In order to answer the questions how these weights are distributed and where are the strong ties concentrated we introduced four new measures: the degree attractiveness, the clustering attractiveness, the clique attractiveness and the shortcut attractiveness. We also extract communities in this weighted network. In this section, for measures introduced in Sect. 2, we shall give an example and an interpretation of the underlying phenomenon.

5.1 Degree attractiveness

The significantly positive value (0.69) of degree attractiveness computed in Sect. 2.1 involves that the weights are attracted by high degrees meaning that proportionally the codes with the most links to other codes are those for which these links will have the strongest weight. That is, they are the ones who share the most citations.

For example, the penal code that is related to 42 other codes (this is the code most related to others) has an average weight on its links of 38.09 (this is the highest average). This means that, on average, each of its links has a weight of 38.09: when the penal code is linked to another code, there are, on average, 38.09 citations between these two codes. The seaports code that is linked to 10 other codes (it is only the 44th code in the ranking of codes with the most links to other codes) has an average weight on its links of 2.7 (it is ranked 45th). This means that on average each of its links has a weight of 2.7: when the seaport code is linked to another code, there are, on average, 2.7 citations between these two codes. These two examples illustrate that weights are higher on edges incident to high degree vertices in the simple graph.

This can be seen as a confirmation of the snowball effect (or the richer-get-richer effect) that was identified via a power-law degree distribution of the simple graph (remind that such a degree distribution occurs with a preferential attachment process). In the simple graph, codes with the most links to other codes are codes that tend to be ″popular″ that is solicited (cited) or soliciting (citing); this is confirmed by the weights.

5.2 Clustering attractiveness

A clustering attractiveness greater that 1 (1.66) computed in Sect. 2.2 implies that the weights will be more present on dense zones in links, for example more present on triangles: the weights reinforce the interconnection between vertices.

For example, the "mutual society", "insurance" and "monetary and financial" codes are all linked, they form a triangle whose sum of the weights is 137, giving an average of 137/3 = 45.67. The "mutual society, "insurance" and "road" codes do not form a triangle because "road" is not linked to "mutuality". The sum of the weights is 70 giving an average of 70/2 = 35.

As soon as one begins to have a triangle this means that the interaction between the three codes begins to be important, this important interaction is then nourished by important citations.

A consequence of the degree attractiveness and clustering attractiveness is the presence of high weights in the rich-club as the rich-club contains high-degree vertices that are highly interconnected, therefore accumulating the degree attractiveness and the clustering attractiveness. One calculation shows that the average weight on the links inside the rich-club is 65.70, which is more than four times the average weight on the whole network (which is equal to 14.99); the rich-club containing 8% of the links however contains more than a third of the weights. Thus, the rich-club phenomenon revealed during the analysis of simple graph is also confirmed in the weighted graph: there is very little difference between the most central vertices (for degree centrality) in the simple graph than in the weighted graph. In addition to this confirmation, the weights on the edges are used to refine the structure in the rich-club and highlight a core composed of two codes (PEN and SAP) around which there is another core of six codes. Thus there is no contradiction with the results previously established on the simple graph but a refinement of the structure.

5.3 Shortcut attractiveness

As seen in Sect. 2.4, the shortcut edges do not have a high weight: their central aspect is strengthened. For example, the link between penal code and handicraft code is the 3rd link of strongest betweenness and it has a weight of 1: only one citation.

Links with high betweenness create shortcuts in a network, so they could be considered as “unexpected” links creating a relationship between two domains that would communicate little. As these are unexpected links it is consistent that we do not find a high weight on these links (there are only a few unexpected citations).

These three measures (degree, clustering and shortcut attractiveness) confirm and amplify the structure studied on the skeleton; the results obtained in the study of the weighted graph do not come into contradiction with the interpretations made on the simple graph.

5.4 Communities

The structure found in the first paper is not contradicted by taking the weights into consideration: we have a central and influential group (the rich-club) and two major communities: one related to social issues and the other related to territories and resources.

The differences are slight; concerning the behavior of these groups detected by partitioning the simple graph in Part I, we obtained the following results:

- The territories and resources group is more stable than the social group when we take the weights into consideration; the social group is split into two parts.
- The code of Tourism which was belonging to the social group in the simple graph ends up in the territories and resources community.
- The code of Asylum and the code of Post Communications, which weren’t belonging to a stable community in the simple graph, are now assigned to a social stable group.

Thus there are slight changes in the belonging to communities of these codes but there is however no major changes. The prior analyzes and the indices we introduced can give an explanation for this phenomenon. Indeed, we saw that the clustering attractiveness and clique attractiveness of the network was strong, that is highly interconnected vertices in the simple graph are highly interconnected with high weights in the weighted graph; therefore if we remind that a community aims at gathering highly interconnected vertices, it is not surprising that there is no major changes in the communities found. Then, we also saw that the network of french legal codes has a particularly high density (compared to usual networks studied in the literature) and we could think that the slight changes observed when divided the graphs (the simple one and the weighted one) into communities are due to the high density of the network: higher the density is, more difficult the partitioning is because when we perform a partitioning on the simple graph, a vertex belongs to one and only one community[7] but it is likely to belong to more than one community and, in this case, the weights could bring the missing information and adjust the membership of some vertices to communities.

The change concerning the code of Tourism could be explained by the social aspect of tourism (which links it to consumer code, insurance code and trade code), but also to its patrimonial aspect (linking it to the urbanism code, the forestry code or the seaports code). The weights being more important towards the patrimonial codes (in particular the code of urbanism) it has changed community.

6 Conclusion

Finally, although the legislative corpus complexity is not reduced to its single aspect of many cross-references between legal texts, we have initiated the process of defining measures of the complexity of the law based on its structure or its contents. In particular one aspect of the complex structure of law can be understood by considering networks induced by the references in text corpus. These references can involve multiple links between two texts which can be modeled by a weighted graph and we introduced in this paper new measures to assess the role of the weights in the network structure, thus allowing a better understanding of the underlying complex system.

For the network of French legal codes, the weight distribution among the edges is far from random and we have outlined, through our new measures, structural patterns: the weights are attracted by vertices with high simple degree, by triangles, and more generally clustered regions. We also pointed out that edges with a high betweenness measure (the weak ties) have low weights, thus increasing the strength of these weak ties.

This emphasizes that the codification process is an intellectual process leading to the creation of a structure that we have highlighted. Codification, a process of reorganization by regrouping and rewriting laws, necessarily creates links of references between texts of different domains (and therefore of different codes): citations between codes are inevitable and numerous. Indeed, the general codification considerations assert that "*It is far preferable, as a general rule, to ensure a fair distribution of the texts between the codes and, if necessary, to resort to the simpler technique of citation without quote to a title, chapter or to articles of another code*."[8] and "*codification should not lead to a permanent disruption of the classification of the right and therefore of the codes*".[9] This paper aims at analyzing the distribution of these citations between dependent codes. The fact that the weights, therefore the citations, are far from being arranged randomly shows a goal of consistency inherent and essential to the codification process as evidenced in

[7] Note that to face this problem we used stable (or robust) communities.

[8] « Il est de loin préférable, en règle générale, de veiller à une juste répartition des textes entre les codes et, au besoin, de recourir à la technique plus simple du renvoi sans citation à un titre, à un chapitre ou à des articles d'un autre code. » in french, paragraph 1.42 of the legistical guide.

[9] « la codification ne doit pas conduire à un bouleversement permanent de la classification du droit et donc des codes » in french.

(François 2013) "*The codification of laws and regulations is an essential means of improving the accessibility and intelligibility of the law. It provides a streamlined presentation, both orderly and coherent to all legal provisions concerning a sector.*".[10]

Concerning the two stable communities found while studying the simple graph, the behavior of the community of codes for territories and resources is less affected by the addition of the weights on the edges than the community of codes for social issues (which is split into two stable communities). Indeed the biggest stable community of the weighted graph is mainly composed of codes belonging to the community of codes for territories and resources found in Sect. 3.2, we also remark that the tourism code (TOU) belongs to this stable community although it was in the social community when partitioning the simple graph.

Further works will consist in the computation of these new measures on other legal networks. Moreover, modeling can be enriched step by step: after having introduced the weights, another step would be to take into account the type of an edge because a network may contain links with different nature.

Acknowledgements We are very grateful to Mme Elisabeth Catta, *Rapporteure* for the Higher Commission for Codification, for her interest in our work and for her helpful comments. This study was funded by the *Réseau Thématique de Recherche Avancée (RTRA) Sciences et Techniques de l'Aéronautique et de l'Espace* (www.fondation-stae.net) in Toulouse, France (MAELIA project - http://maelia1.wordpress.com/). Statistical properties of networks have been computed with *R* and the library *igraph* (http://www.r-project.org/).

Appendix A

See Table 11.

[10] « La codification des textes législatifs et règlementaires constitue un moyen essentiel d'améliorer l'accessibilité et l'intelligibilité du droit. Elle permet une présentation rationalisée, à la fois ordonnée et cohérente, de l'ensemble des dispositions juridiques concernant un secteur. » in french.

Table 11 Labels and short name of the 52 French Legal Codes retained in this study

Label	Short name	Label	Short name
ART	Handicraft	LGA	Honor Legion
ASF	Social service	MIN	Mining
ASS	Insurance	MOF	Monetary & financial
AVI	Civil aviation	MPU	Public contract
CHA	Housing	MUT	Mutual society
CIV	Civil	OGJ	Admin. of justice
CNS	Consumer	PAT	Estate
COM	Trade	PCI	Civil procedure
DEF	Defense	PCO	Post communication
DMM	Mercantile marine	PEN	Criminal
DOE	State-owned property	PIT	Intellectual property
DOU	Customs	PMA	Seaports
DPF	Public rivers	PPE	Criminal procedure
EDU	Education	REC	Research
ELE	Elections	ROU	Traffic
ENV	Environment	RUR	Rural
EUP	Expropriation in public interest	SAP	Public Health
FAS	Family	SDA	Asylum
FOR	Forestry	SNA	National service
GCT	Local authorities	SPO	Sport
GPP	Property legal person	SSC	Social Security
ICI	Film industry	TMA	Marine employment
INM	Monetary media	TOU	Tourism
JUA	Administrative court	TRA	Employment
JUF	Financial court	URB	Urbanism
JUM	Military court	VOR	Road system

References

Boulet R, Mazzega P, Bourcier D (2011) A network approach to the French system of legal codes—part I: analysis of a dense network. Artif Intell Law 19(4):333–355

Bourcier D, Mazzega P (2007) Codification, law article and graphs. In: Jurix AR, Lodder, Mommers L (eds) Legal knowledge and information systems. IOS Press, Amsterdam, pp 29–38

Brandes U, Erlebach T (Eds) (2005), Network Analysis: Methodological Foundations. Lecture Notes in Computer Science, 3418

Chung FRK (1997) Spectral Graph Theory, American Mathematical Society

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70:066111

Erdös P, Rényi A (1959) On random graphs. Publ Math 6:290–297

Fowler JH, Johnson TR, Spriggs JF II, Jeaon S, Wahlbeck PJ (2007) Network analysis and the law: measuring the legal importance of precedents at the U.S. Supreme Court. Political Anal 15(3):324–346

François, D (2013) Consolidation et codification, simplifier et faciliter l'accès au droit. Forum européen des journaux officiels

Freeman LC (1979) Centrality in social networks: conceptual clarification. Soc Netw 1:215–239
Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99:7821–7826
Grossman JW (2002) The evolution of the mathematical research collaboration graph. Congr Numer 158:202–212
Jouve B, Kuntz P, Velin F (2002) Extraction de structures macroscopiques dans des grands graphes par une apporche spectrale. Extraction des Connaissances et Apprentissage 1:173–184
Katz DM, Bommarito M (2014) Measuring the Complexity of the Law: the United States Code. J Artif Intell Law 22(4):337–374
Latapy M, Magnien C (2006) Measuring Fundamental Properties of Real-World Complex Networks", http://arxiv.org/abs/cs.NI/0609115
Mazzega P, Bourcier D, Boulet R (2009a) The Network of French Legal Codes. Proc ICAIL. doi:10.1145/1568234.1568271
Mazzega P, Bourcier D, Boulet R (2009b) Code Communities in the French Legal System. In: Proceedings of GEMME 2009
Newman MEJ (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69:066133
Pons P, Latapy M (2005) Computing communities in large networks using random walks. J Graph Algorithms Appl 10:191–218
Schuck PH (1992) Legal complexity: some causes, consequences, and cures. Duke L J 42(1):1–52
Tarissan F, Nollez Goldbach R (2015) Temporal properties of legal decision networks: a case study from the International Criminal Court, In: 28th International conference on legal knowledge and information systems (JURIX'2015), Braga, Portugal
von Luxburg U (2007) A Tutorial on Spectral Clustering. Stat Comput 17:395–416
Watts DJ (2003) Small worlds: the dynamics of networks between order and randomness. Princeton University Press, Princeton
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
Whalen R (2016) Legal networks: the promises and challenges of legal network analysis. Mich. St. L Rev, Michigan, p 539
Winkels RGF, Boer A (2014) Finding and visualizing dutch legislative context networks. In: Winkels R, Lettieri N, Faro S (eds) Network analysis in law. Diritto Scienza Tecnologia, Rome, pp 157–182