



**HAL**  
open science

## Behind the Communities, a Focus on the Sparse Part of a Network

Mehdi Djellabi, Bertrand Jouve, Frédéric Amblard

► **To cite this version:**

Mehdi Djellabi, Bertrand Jouve, Frédéric Amblard. Behind the Communities, a Focus on the Sparse Part of a Network. 6th International Workshop on Complex Networks and their Applications , Nov 2017, Lyon, France. pp.216-228, 10.1007/978-3-319-72150-7\_18 . hal-02864407

**HAL Id: hal-02864407**

**<https://hal.science/hal-02864407v1>**

Submitted on 11 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:  
<http://oatao.univ-toulouse.fr/22279>

### Official URL

[https://doi.org/10.1007/978-3-319-72150-7\\_18](https://doi.org/10.1007/978-3-319-72150-7_18)

**To cite this version:** Djellabi, Mehdi and Jouve, Bertrand and Amblard, Frédéric *Behind the Communities, a Focus on the Sparse Part of a Network*. (2017) In: 6th International Workshop on Complex Networks and their Applications (), 29 November 2017 - 1 December 2017 (Lyon, France).

Any correspondence concerning this service should be sent to the repository administrator: [tech-oatao@listes-diff.inp-toulouse.fr](mailto:tech-oatao@listes-diff.inp-toulouse.fr)

# Behind the Communities, a Focus on the Sparse Part of a Network

Mehdi Djellabi<sup>1</sup>(✉), Bertrand Jouve<sup>2</sup>, and Frederic Amblard<sup>3</sup>

<sup>1</sup> Institut de Recherche en Informatique de Toulouse, Toulouse, France  
djellabi.mehdi@gmail.com

<sup>2</sup> FRAMESPA & Institut de Mathématiques de Toulouse, CNRS, Université de Toulouse, Toulouse, France

<sup>3</sup> Institut de Recherche en Informatique de Toulouse,  
Université Toulouse 1 Capitole, Toulouse, France

**Abstract.** We propose a method that allows to detect the subset of the sparse nodes in a complex network, providing supplementary informations about its structure and features. The aim is to produce a complementary approach to the classical ones dealing with dense communities, and in the end to develop mixed models of community classification which are articulated around the network's sparse skeleton. We will present in this article different metrics that measure sparsity in a network, and introduce a method that uses these metrics to extract the sparse part from it, which we tested on a toy network and on data coming from the real world.

## 1 Introduction

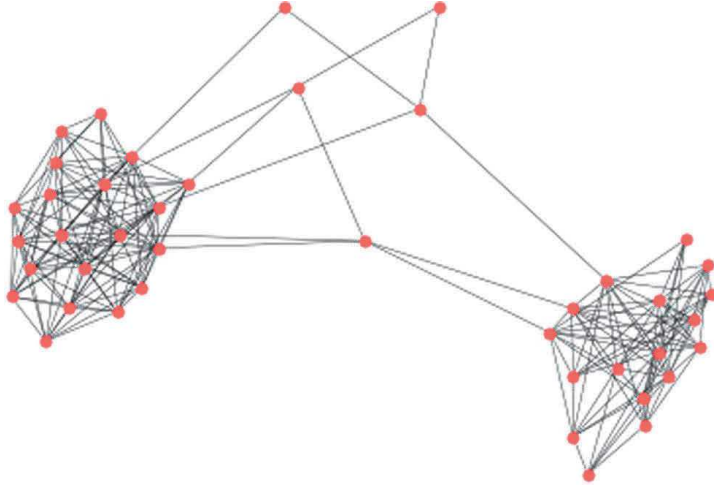
Networks are nowadays taking an important place in the everyday life, whether by means of the social media that are used by nearly everyone or our dependence on the world wide web, and more specifically in the scientific domain thanks to the evolution of the calculation means. Now we use networks as a representation for a lot of problematics, making their modeling easier. As examples we can cite the biological systems [15] through the study of proteins and the relation between their structure and function, ecological systems [10], and other examples in economics [14] or even history [7] and geography [2]. Therefore researches concerning the structure and the fundamental properties of networks have been achieved to allow their better comprehension, since the understanding of the network structure is generally the first step before a detailed study. Among multiple structural characteristics, some turned out to be common to a large class of networks, even if these networks model completely different systems, such as high a transitivity coupled to the small world effect, and the scale free property identified in most social networks [12]. An other specificity of the real world networks is the self-organization of their nodes in communities, and by community we mean a group of nodes which are much more connected between themselves than they are to the rest of the network [11]. Thus the community detection problem became very important these last years leading to the emergence of many approaches

[5]. However, few studies have focused on understanding the links between communities. One can only cite several works on hierarchical organization in complex networks [13]. The aim of our study is to go further in the modeling of arrangement of communities in a network by detecting the nodes that do not belong to any community, or at least whose belonging can be discussed. These nodes are the basis to reveal the structuration of the inter-part communities. Classical approaches are focusing on the partitioning of nodes into communities but Burt [3] for instance shows the importance of the sparsely connected individuals in a social context. We want to develop mixed models of community classification articulated around the network's sparse skeleton: a unique subset of nodes which share the same properties rather than a partition with several components as it is done in community detection methods, all with the purpose to apply it to dynamical networks and thus highlight both the dynamics (possibly different) of the dense and the sparse parts. The paper is structured in three parts: the first one presents methods of density evaluation and the extraction of the sparse parts, the second one describes the datasets and the models on which we tested our method, and the last part is dedicated to the presentation and the discussion of their results.

## 2 From Sparsity Measurement to the Extraction of the Sparse Parts

Before we discuss the details of the calculation it is better to start by telling which are the characteristics that make a node belong to the sparse part (SP in the following), and it is worth noticing that it depends on our viewpoints and that other are of course possible. So we do not give a strict definition of SP but only a qualitative description of what we expect it to be based on our observations, it is therefore a similar scheme to those adopted in any community detection approach: the results we find are directly related to the descriptions that we make of the entities we look for. As for the criteria that should share the nodes of SP, we rely on what is traditionally used for density estimation in networks and we'll enrich it with a few characteristics that seem adapted. We stipulate that it is possible for certain types of networks (typically the real world networks) to split their nodes into two separate categories: on one hand we have the dense part which is the subset of nodes belonging to communities or more generally to clusters, and on the other hand the nodes which do not. Figure 1 is an illustration of a small network made of two clusters/communities and five nodes between them. This is one of the advantages that may provide the separation proposed here, instead of facing a situation in which it is hard to tell to which community belongs each of these five nodes, we classify them in the SP. After that it becomes easier to separate the two remaining communities by (for example) a modularity based algorithm. One criterion that is frequently used for estimating the node's density is the ratio between this node's degree and the size of the network (number of nodes). But this is far from describing precisely the notion of density, the best counter example we can provide is the

comparison between a high degree star node and a smaller degree node that lays inside a cluster. In fact a node may have a large number of neighbors but if these neighbors are not well connected to each other, then it is not likely that it belongs to the dense part. On the other hand we can have a lesser degree node strongly connected to a community making it an element of the dense part.



**Fig. 1.** Here we can easily identify two clusters and 5 nodes weakly linked to them, but we cannot clearly establish which of these 5 nodes is in which cluster.

Such considerations have led us to the elaboration of our own metrics, which report the quality of each node in terms of density so that we can rank them before we extract the SP. Three metrics are presented and detailed in the following subsections, each of them relies on a different considerations.

## 2.1 The Clustering Coefficient Based Metric

The upcoming approaches are simply driven by the impossibility to capture a non-local property such as the belonging (or not) of a node to a dense zone, and by zone we mean a more or less close neighborhood, only on the basis of local properties (specific to that node). So we developed approaches in which the scores attributed to nodes depend on the state of their neighborhoods, and thus we evaluate each node by having a more global vision, contrarily to what we would have obtained if we only measured (for example) their degree. To clarify this we propose a first method that is based on the clustering coefficient, and we'll explain the model that allows its extension to the neighborhood evaluated. We remind that the clustering coefficient associated to a node  $i$  is given by the ratio between the number of links that exist between this node's neighbors and the number of pairs they form.

$$C_i = \frac{2(|E(Ego_i)| - d_i)}{d_i(d_i - 1)} \quad (1)$$

Here we note  $Ego_i$  the subnetwork formed by the node  $i$  and its neighbors,  $k_i$  the degree of  $i$  and  $E(G)$  the set of edges in a network  $G$ . This coefficient is equal to 0 for a star node and 1 if  $i$ 's neighborhood is a clique (a subset of nodes in which every pair is linked by an edge). However as said in the previous paragraph, we do not only restrain to the node's coefficient to evaluate its density, we rather extend this measure to its neighbors and their neighbors etc. We also modulate the contribution of each group of nodes with an inversely proportional function to the distance to the evaluated node. So the coefficient averaged over the first neighbors has a greater contribution to the node's score than that of the second neighbors which at their turn contribute more than the group of nodes whose distance to our node is 3 and so on. Indeed for a node to be in a dense zone does not only mean to have a high clustering coefficient, its (more or less) close neighbors must also share the property. This provides us with a larger vision centered around the node whose density is evaluated, as large as we want it to be just by modifying the modulating function, but still making sure it decreases with the distance. Let  $C(i)$  be the clustering coefficient of node  $i$  to obtain the metric we just described, we do a breadth first search on the network starting from node  $i$ , which provides us a set

$$K = \{k_1, k_2, \dots, k_L\}$$

in which an element  $k_l$  is the subset of nodes reachable from  $i$  through a path of length  $l$ . Then we average the clustering coefficient over each of the subsets, let's call  $\rho_i$  the average over  $k_i$ :

$$\rho_i = \frac{1}{card(k_i)} \sum_{j \in k_i} C(j) \quad (2)$$

The last step is to sum these averages and give them a weight that corresponds to their importance in the final score, like:

$$Q(i) = \left( \sum_{i=1}^l \rho_i \cdot f(i) \right) \quad (3)$$

where  $f$  is the decreasing function that we determine empirically.

## 2.2 The Modified Betweenness Metric

The second metric is based on a slight modification of the betweenness centrality, so that it reflects the sparsity in a network. Remind that the betweenness centrality assigns scores to the network's nodes according to their implication in geodesics, and by geodesic we mean the shortest path between a pair of nodes.

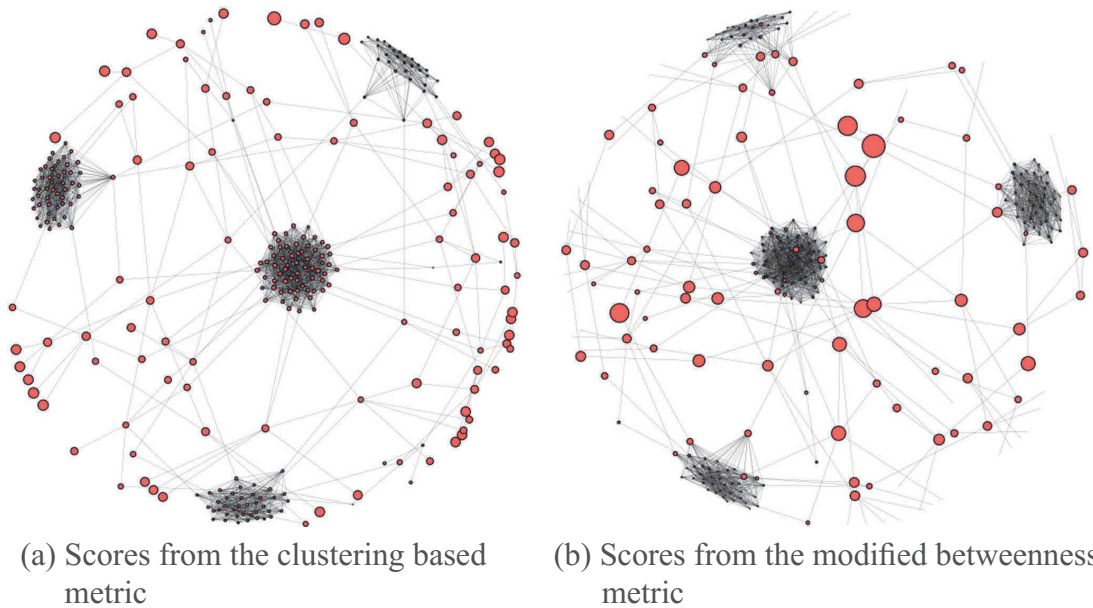
$$C_b(i) = \sum_{(u,v) | i \notin \{u,v\}} \frac{\sigma(u, v|i)}{\sigma(u, v)} \quad (4)$$

where  $\sigma(u, v|i)$  is the number of shortest paths between nodes  $u$  and  $v$  that cross the node  $i$ , and  $\sigma(u, v)$  the number of shortest paths that exist between

$u$  and  $v$ . This centrality is inadequate to measure density in a network, and we would like to improve it in order to obtain a measure that favors the SP nodes, that is those who typically lay between communities. The disadvantage of betweenness centrality is that when a node is inside a community, it is involved in a great number of geodesics which link the pairs of this same community, leading to the attribution of large scores to their nodes (sometimes dominant, according to the community size) which is not what we want. To overcome this difficulty one could use the following consideration: the geodesics of the same community nodes are short, whereas a SP node should be implied in longer ones. So, one solution consists in modifying the upper formula by discriminating geodesics, which is easy if we give to each one a weight proportional to its length:

$$C_b(i) = \sum_{(u,v)|i \notin \{u,v\}} f(l_{uv}) \frac{\sigma(u,v|i)}{\sigma(u,v)} \quad (5)$$

where this time  $f$  is an increasing function whose variable  $l_{uv}$  is the length of the shortest path separating  $u$  and  $v$ . The results obtained from testing these two metrics on an artificial “toy” network (that we detail later in this paper) are shown in the Fig. 4.



**Fig. 2.** In (a) the node sizes are inversely proportional to the score they get from the clustering metric, and we can see that nodes inside the clusters have the highest scores. In 4(b) the size of each node is proportional to the score they get from the betweenness metric. If we inverse the scores of the classification by the clustering metric, we notice a similarity in the results of (a) and (b), with the difference that the betweenness metric gives a null score to the leaves (1-degree nodes) since there is no geodesic that passes by any of them. We can overcome this issue since it is easy to identify the leaves in a network and then adopting a convention to consider them as nodes of the sparse part.

## 2.3 A Metric Based on the Force-Directed Layout Algorithm

### 2.3.1 The Principle of Nodes Positioning by Optimizing the Energy Function

The problem of the “nice” positioning of nodes when it comes to draw a network has been proposed in the 80’s [4] and aims to add the visualization process to the study and exploration of large networks. The idea is to model a network as a physical system in which nodes correspond to steel rings and edges to springs following the original structure of the network. Algorithms calculate the total energy of the system and consider the best layout as the one that minimizes this energy. There are several models of this type but we only cite two popular ones, starting by Fruchterman and Reingold’s [6] layout model. It considers a spring between each pair of nodes, but with a free length that is proportional to the length of the geodesic between pairs (the longer the distance the greater the free length). For a network of  $N$  nodes represented by Fruchterman and Reingold’s model we obtain a total energy that can be written as:

$$\sum_{1 \leq i < j \leq N} k_{ij} (|x_i - x_j| - l_{ij})^2 \quad (6)$$

Here we call  $x_i$  the position of node  $i$ ,  $l_{ij}$  the length of the shortest path between  $i$  and  $j$ , and  $k_{ij}$  the spring constant which depends on parameters that we do not discuss here. So in this model the smaller the distance between nodes the more they are susceptible to be spatially close according to the layout. An other model is proposed in [9] considers that the springs are only between the pairs which are actually linked by an edge, and adds a repulsive term to the energy functional between the pairs which are not. The energy to minimize in this model is written:

$$\sum_{i=1}^N (\sum_{j < i} A_{ij} |x_i - x_j|^2 + D_{x_i}) \quad (7)$$

with  $A_{ij}$  the  $(i, j)$ ’th term in the network’s adjacency matrix, and  $D_{x_i}$  the repulsive term related to each node  $i$ . In this model we have two terms, the first brings closer the linked nodes and the other one drives away from each other the pairs which are not. Therefore the nodes we aim to detect have a small spatial density. In both models presented we need to minimize a function of  $N$  variables, which is in fact impossible to achieve analytically, except in extremely rare cases. We then need to resort to an optimization algorithm. These functions contain a great number of local minima and the result returned by the algorithm are generally not the one that corresponds to the global minimum, but rather a local one and so an approximation of the exact solution. However the layout provided by these approximations is good enough to be suited for our metric.

### 2.3.2 The Metric

We can exploit these layouts to measure density, simply by connecting the network density to its spatial density (the number of nodes per unit surface) provided by a layout. This method has the advantage to be much quicker than the



two we previously introduced. We evaluate each node by calculating the number of nodes (neighbors or not) which lay inside a circle of diameter  $r$  centered around the evaluated node, and divide this number by the surface of the disk to obtain a normalized surface density. We can chose (empirically) the value of the parameter  $r$ , or get rid of this constraint by following a similar scheme to that in Sect. 2.1: sum the score of different diameters until a certain  $r_{max}$  at which the density becomes the same for all nodes, and in the same time modulate these contributions by a decreasing function.

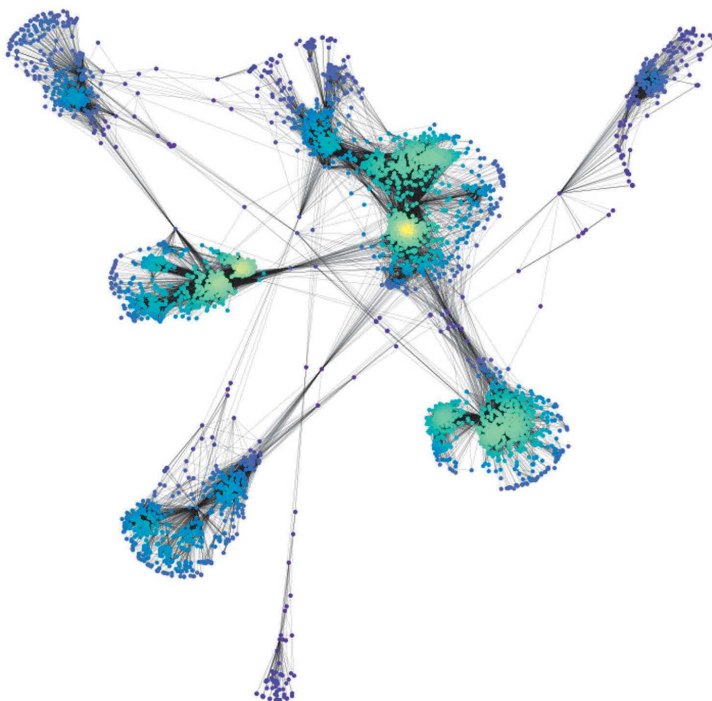
The motivation behind this method is that the communities tend to form spatial clusters in the geometric space, which results in the creation of a strong repulsive potential to push away all nodes that do not belong to it (for the case of the model in [9]). Now let  $\rho_i(r)$  be the number of nodes contained inside a circle, centered around the position of node  $i$  and of diameter  $r$  and normalized by the surface of the disk. We attribute to node  $i$  the following score:

$$Q(i) = \int_0^{r_{max}} \rho_i(r) \cdot f(r) \cdot dr \quad (8)$$

where  $f$  is a decreasing function and  $r_{max}$  is defined whatever the node  $i$  by

$$\max(\rho_i(r_{max})) - \min(\rho_i(r_{max})) \leq \epsilon$$

The results of this metric on a 4092 nodes real world network [8].



**Fig. 3.** Here we chose a color map that assigns a bright yellowish color to the highly rated (dense) nodes and the dark purple color to the nodes with small scores (sparse), the layout algorithm is the one in [6]

## 2.4 Interruption Criterion and SP Extraction

So far we did not talk yet about the way to use the metrics we introduced to determine the sparse parts, this is what we do in the following subsection. We emphasize that we could just use a thresholding method on the metric values, if there are points at which a threshold is adapted, like an important gap in their ordered values. Otherwise, in general case we give our interruption criterion for the extraction of sparse parts: SP are the maximal subnetwork whose nodes are the most unfavorably evaluated ones (by the metric) and with a density that is clearly smaller than the density of the whole network, in comparison with its configurational model. Remained that a network’s configurational model is the network in which each node keeps its degree but is randomly rewired to the other nodes of the network.

### 2.4.1 SP Extraction

We visit the network’s nodes in the order provided by the metric and at each step  $i$  we add the  $i$ th node to the subnetwork  $U_{i-1}$  constituted by the  $i - 1$  previous ones, measure the density of the subnetwork  $U_i$  and repeat until  $i = N$  ( $U_N = G$  the studied network). We do the same for the configurational model of the network (rewire, classify and visit its nodes in the provided order). Finally we draw both the corresponding outputs. From this outputs we can be in one of the two following situations: the first and preferable one is when we have a deep minimum (smaller than one if we divide it by the density of the whole network) in the density function at a certain point  $i$  which means that we clearly have a subnetwork whose density is way smaller than that of the whole network. If no such minimum exists (second case) which happens frequently, we can compare the density function obtained from the study of the original network to that obtained from its randomized version, and look for the point at which the difference (random minus original) between the two is maximal. The resulting subnetwork is considered to be the SP.

## 3 Datasets

### 3.1 The TreeCom Model

To ensure that our model works correctly on simple cases, this network was meant to present a clear separation between the dense parts and the sparse part. Its construction is simple and contains a sparse skeleton, some of its extremities attached to dense clusters of different sizes. First we construct the “skeleton” as a tree, which is a sparse network obtained by calculating the spanning tree over a Barabasi-Albert model [1]. The reason we chose this specific tree is to have the maximum heterogeneity in the degree distribution of the nodes, to show that large degree nodes in the tree are still classified as nodes of the SP by our model. The next step consists in a random choice of nodes to which the clusters are going to be attached, with a constraint on the number of edges that link them to the tree nodes. This model has already been used to show the results of two metrics in Fig. 2(a) and (b).

### 3.2 The Real World Network

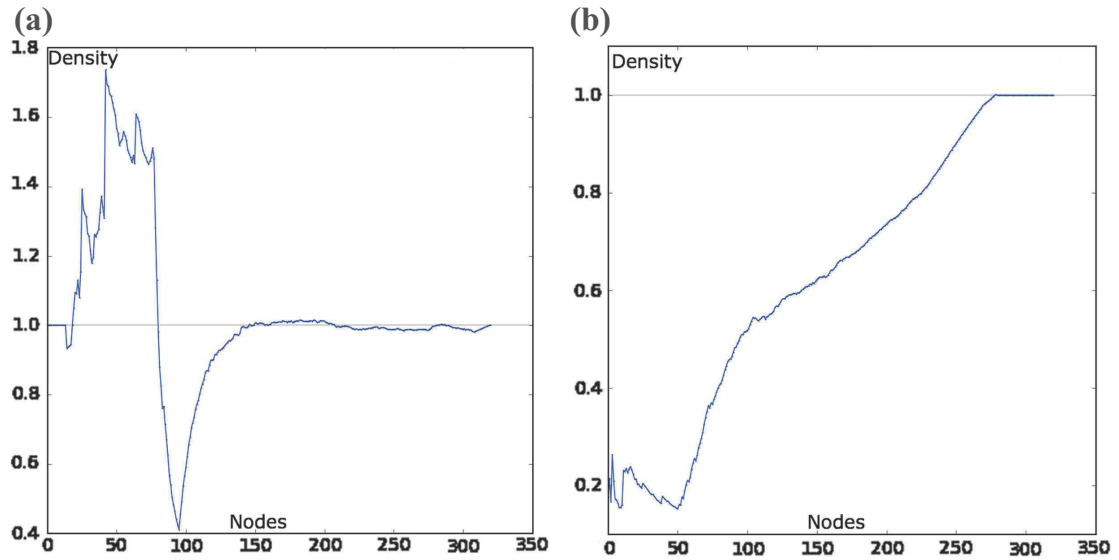
This one is a small sample of the Facebook network, contains 4100 nodes but has the interesting features of real world networks, like a high average clustering coefficient, a community structure, and a rather small average diameter. It consists in “circles” (or “friends lists”) from Facebook. Facebook data was collected from survey participants using the application. The dataset includes node features (profiles), circles, and ego networks [8], anonymized for ethical reasons.

## 4 Results and Discussion

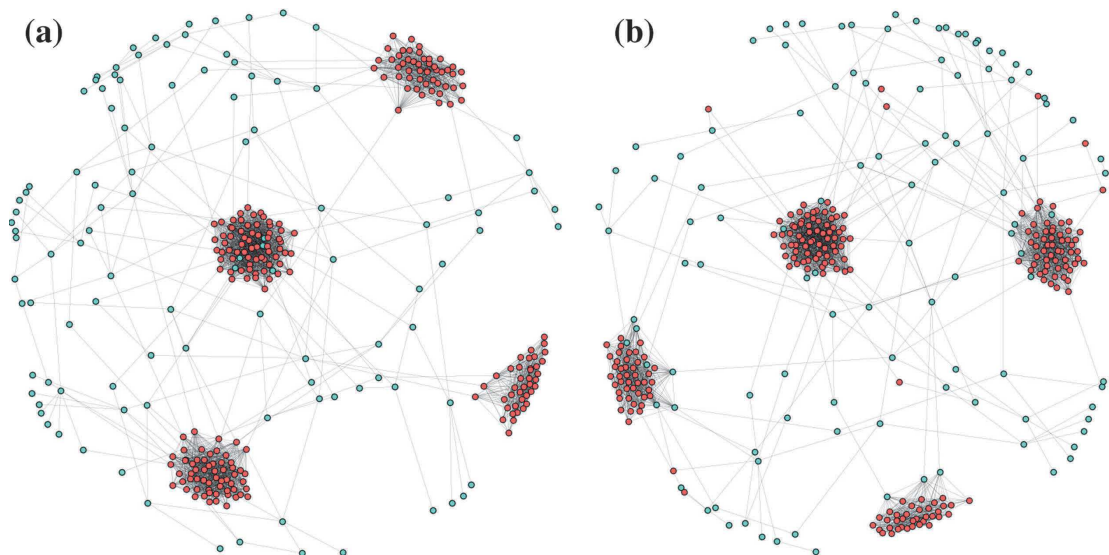
In the examples that are shown below, the two first concern a TreeCom model, which parameters are a tree of a 100 nodes, 4 clusters of averagely 50 nodes each, and a maximum of 4 links between each cluster and the tree nodes. The results of the classification have already been shown in Fig. 2(a) and (b) so we directly give the graphics corresponding to the interruption criteria following these 2 first metrics (clustering and modified betweenness). We can see in Fig. 4(a) that the minimum is very close to 100 nodes (the few tree nodes attached to the clusters are considered denser) which corresponds to the number of nodes in SP. It is not the case in Fig. 4(b) (which is the result of the same analysis applied to the same network but using a different metric), and the reason is just that the betweenness based metric attributes a null score to the 1-degree nodes. On the contrary the clustering based metric classifies them as SP nodes. Indeed the minimum around 50 nodes in Fig. 4(b) corresponds to the subnetwork constituted by all the nodes in the tree which are not leaves (1-degree nodes). The number of these leaves has been measured and happens to be approximately 50, the whole sums up to one hundred and if we adopt the convention to consider leaves in the SP we find a very similar result to the one obtained from the use of the clustering based metric. This is in the following figures.

The Facebook network has been studied by the application of the three metrics presented in Sect. 2 but only the results obtained from the use of the layout based metric (Cf Fig. 3) are presented here, the ensuing interruption criterion is shown in Fig. 6. We evaluate the density of each subnetwork (in the interruption criterion) by the measure of the spatial density of the nodes, since the metric which classifies them uses it. We can see in Fig. 6(a) that the blue curve has no minimum, but it seems to increase at different rates while the orange curve grows more smoothly after a quick increase in the beginning. In such a situation we can threshold the blue curve at the most convenient point which we chose as the maximal distance to the randomized case. That corresponds in Fig. 6(b) to the second maximum of the curve, without taking into account the first one which is a consequence of the perturbed behavior in the randomized case (orange curve) due the fact that leaves in the original network remain leaves in its configurational model, the rest of the nodes are homogeneously mixed (Fig. 7).

Finally we compare the results of the different metrics on our datasets, but for several reasons (given our datasets) it is not relevant enough to compare the structure of the sparse parts obtained from each of these metrics, for the

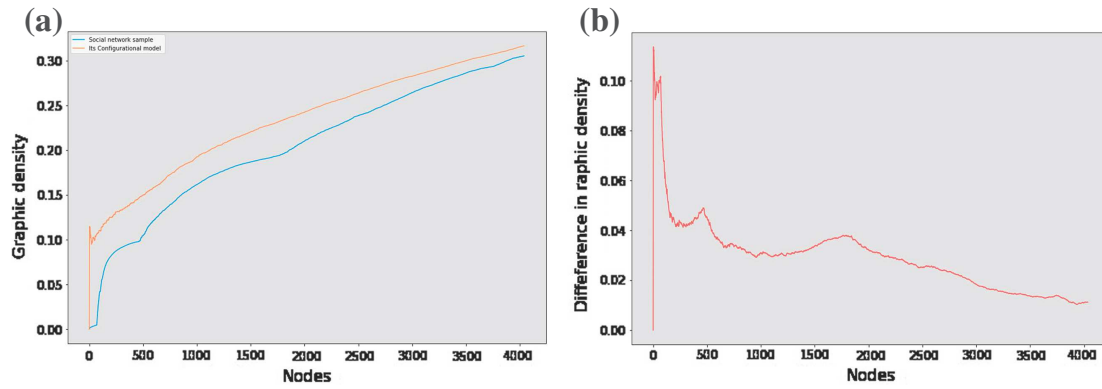


**Fig. 4.** (a) The evolution of density of the subnetwork versus its size according to the clustering based metric. (b) The evolution of density of the subnetwork versus its size according to the modified betweenness metric

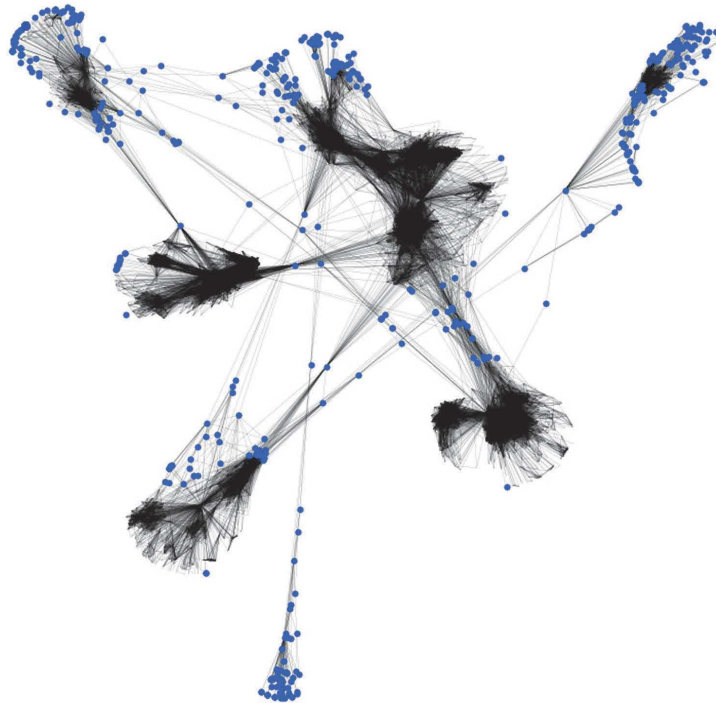


**Fig. 5.** The blue nodes are those selected by our method as the SP. The results of the clustering metric is the one in (a) and that of the modified betweenness metric is in (b). We adopted the convention to consider the leaves among the SP in the (b) case.

toy network the structure of the sparse part is trivial, and represented by a tree which is easily recovered using each metric (cf Fig. 5), and on the other hand we have a sample of a real world network whose calculated sparse parts are composed of multiple small non connected components. We rather compare the structure of the dense parts in each network, obtained from depriving the original network of its sparse parts obtained using each of our metrics, this does not give us informations about the structure of the sparse parts but it highlights the density gain in each case. Since the resulting dense subgraphs



**Fig. 6.** (a) The evolution of the spatial density of the subnetwork versus its size in both the original and the configurational model. (b) The difference between the graphical density of the Facebook network and its configurational model versus the size of the subnetworks



**Fig. 7.** The blue nodes are those who have been selected by the (most realistic) maximum which is around 500 nodes, the rest of them are not drawn but their edges are, to show the general positioning of the SP nodes in the network

are not connected, the density is averaged over all the connected components of significant size (more than 3 nodes). These results are shown in Table 1. The density of the clusters in the studied treeCom network are all equal to 0.5 which results in an average density of the same value in its dense subnetwork (once we take away its sparse part). In the other hand we cannot compare the real world dense subnetwork to any ground truth reality, but the fact that the results in Table 1 are of similar magnitude is encouraging.

**Table 1.** Density gain in a network without its sparse parts

	Original density	Clustering metric	Betweenness metric	Layout metric
Facebook sample	0.011	0.223	0.168	0.193
TreeCom model	0.039	0.499	0.468	0.431

## 5 Conclusion and Perspectives

We exposed a new method that aims for a better understanding of the networks structure, and although it is the beginning of a study that has to be extended, it has a few interesting results. Moreover there are a lot of points that can be deepened like the empirical definition of the SP, the choice of the layout algorithm (in this paper we used the simplest but not necessarily the best suited) and the parameters for each metric, which we took from visual considerations here. We can also think of the questions that can arise from the transposition of this method to the dynamical case, like the existence or not of a sparse core that evolves in time, at which case how would be its evolution? One more issue is the calculation time estimated to  $O(N^3 \cdot \log(N))$ . A few applications in the static case are possible though, and are the central subject of the upcoming studies such as the contribution our analysis can bring to the study of communities or the identification of the important spreaders in epidemics models. Finally the study of a network's sparse part provides informations about the structure of the nodes laying between its communities which is useful for the modeling of real world networks.

## References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47 (2002). <https://doi.org/10.1103/RevModPhys.74.47>
2. Barthélemy, M.: Spatial networks. *Phys. Rep.* **499**(1–3), 1–101. <http://linkinghub.elsevier.com/retrieve/pii/S037015731000308X> (2011). <https://doi.org/10.1016/j.physrep.2010.11.002>
3. Burt, R.: *Structural Holes: The Social Structure of Competition*. Harvard University Press. <https://books.google.fr/books?id=FAhiz9FWDzMC> (2009)
4. Eades, P., Foulds, L., Giffin, J.: An efficient heuristic for identifying a maximum weight planar subgraph. In: Billington, E.J., Oates-Williams, S., Street, A.P. (eds.) *Combinatorial Mathematics IX: Proceedings of the Ninth Australian Conference on Combinatorial Mathematics Held at the University of Queensland, Brisbane, Australia, August 24–28, 1981*, pp. 239–251. Springer, Berlin Heidelberg (1982). <https://doi.org/10.1007/BFb0061982>
5. Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**(3–5), 75–174. <http://linkinghub.elsevier.com/retrieve/pii/S0370157309002841> (2010). <https://doi.org/10.1016/j.physrep.2009.11.002>
6. Fruchterman, T.M., Reingold, E.M.: Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**(11), 1129–1164 (1991). <https://doi.org/10.1002/spe.4380211102>

7. Le Couédic, M., Leturcq, S., Rodier, X., Hautefeuille, F., Fieux, E., Jouve, B.: Du cadastre ancien au graphe. Les dynamiques spatiales dans les sources fiscales médiévales et modernes. *ArchéoSciences* **36**, 71–84. <http://archeosciences.revues.org/3758> (2012). <https://doi.org/10.4000/archeosciences.3758>
8. Leskovec, J., Mcauley, J.J.: Learning to discover social circles in ego networks. *Adv. Neural Inf. Process. Syst.* 539–547. <http://papers.nips.cc/paper/4532-learning-to-discover-social-circles-in-ego-networks> (2012)
9. Martin, S., Brown, W.M., Klavans, R., Boyack, K.W.: OpenOrd: an open-source toolbox for large graph layout, p. 786806 (2011). <https://doi.org/10.1117/12.871402>
10. Miele, V., Matias, C.: Revealing the hidden structure of dynamic ecological networks. *R. Soc. Open Sci.* **4**(6), 170251 (2017). <https://doi.org/10.1098/rsos.170251>
11. Newman, M.E.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**(23), 8577–8582. <http://www.pnas.org/content/103/23/8577.short> (2006)
12. Qiao, J., Huang, H.Q., Li, G.Y., Fan, Y.: Bridging the gap between different social networks. *Phys. A Stat. Mech. Appl.* **410**, 535–549. <http://linkinghub.elsevier.com/retrieve/pii/S0378437114004488> (2014). <https://doi.org/10.1016/j.physa.2014.05.067>
13. Ravasz, E., Barabási, A.L.: Hierarchical organization in complex networks. *Phys. Rev. E* **67**(2) (2003). <https://doi.org/10.1103/PhysRevE.67.026112>
14. Souma, W., Fujiwara, Y., Aoyama, H.: Complex networks and economics. *Phys. A Stat. Mech. Appl.* **324**(1–2), 396–401. <http://linkinghub.elsevier.com/retrieve/pii/S0378437102018587> (2003). [https://doi.org/10.1016/S0378-4371\(02\)01858-7](https://doi.org/10.1016/S0378-4371(02)01858-7)
15. Vuillon, L., Lesieur, C.: From local to global changes in proteins: a network view. *Curr. Opin. Struct. Biol.* **31**, 1–8. <http://linkinghub.elsevier.com/retrieve/pii/S0959440X1500024X> (2015). <https://doi.org/10.1016/j.sbi.2015.02.015>