Multiscale Detection of Stable Communities Using Wavelets on Networks

Nicolas Tremblay*, Pierre Borgnat

Laboratoire de Physique, ENS Lyon, CNRS UMR 5672, Université de Lyon, France

Abstract Graph wavelets, a recently defined signal processing tool, are used to tackle a classical complex system problem: community mining in networks. By construction, graph wavelets encode local and scale-dependent information, providing "ego-centered" views (one for each node and each scale) of the network. A careful definition of the wavelet kernel enables us to use these wavelets for multi-scale community mining: groups of nodes that have similar "ego-centered" views of the network at a given scale will tend to be in the same community at that scale. Moreover, we propose a method based on partition stability that estimates the different "natural" scales of a network. All in all, given a network, the method detects relevant scales if there are any, and outputs the associated partitions. We successfully test this method on a graph benchmark having hierarchical communities.

Keywords Complex Networks, Community Mining, Multiscale, Detection

1 Introduction

In many complex systems, data are naturally represented as networks (or weighted graphs): social networks, sensor networks, Internet networks, neuronal networks, transportation networks, biological networks,... A striking property of many networks, and a common way of simplifying the network's analysis, is their modular structure, i.e., there exists groups of nodes, called communities [3], that are more connected with themselves than with the rest of the network. As nodes in a same community tend to share common properties, community mining provides both a sketch of the structure of a network, and some insight on nodes' properties. One issue in community mining is defining the scale at which one wants to analyze the network. Many algorithms (see the review [3]) are based on the optimisation of appropriate evaluation functions such as the popular modularity [14] and generally discard the question of the scale or propose only *ad-hoc* discussions. Modularity for instance is known to favour an intrinsic scale of description [4, 10].

The present work is twofold. First, we develop a scale-dependent procedure which identifies community structures, i.e., that classifies nodes according to their community, at different scales. After one decides on a scale of interest (or a collection of scales), the objective is to mine for the communities at this(ese) scale(s). Using multiple scales will provide a fully multiscale community description of the network. Some authors have proposed multiscale community mining either based on random walk processes [17, 11], or on definitions of parametric modularities [15, 1]. Our proposition for community mining is to rely on the recent construction of graph wavelets based on spectral graph theory [5]. By nature, the wavelet associated to a node a and a scale s is local. It is centered around this node and spreads on its neighbourhood so that the larger is s,

*Corresponding author. Email: firstname.name@ens-lyon.fr

the larger is the spanned neighbourhood. In some sense, wavelets give an "egocentered" view of how a node "sees" the network at that scale. Taking advantage of this local information encoded in wavelets, we develop an approach that clusters together nodes whose local environments are similar, i.e., whose associated wavelets are correlated.

The second part of this work focuses on how to assess the relevance of a given scale. Indeed, the method will always output a partition per scale, but does not decide whether these partitions are relevant or not. Based on partition instability measures and on a statistical test where we compare the graph of interest to Erdös-Renyi random graphs, we propose a way to detect the relevant scales of the network.

Section 2 recalls useful background material on spectral graph theory and graph wavelets. In section 3, the multiscale community mining algorithm is presented. In section 4, we discuss partition stability and automatic detection of relevant scales. We conclude in section 5.

2 Spectral Graph Theory and Wavelets

2.1 The Graph Fourier Transform

Let $\mathcal{G} = (V, E, A)$ be a undirected weighted graph with V the set of nodes, E the set of edges, and A the weighted adjacency matrix such that $A_{ij} = A_{ji} \geq 0$ is the weight of the edge between nodes i and j. Note N the total number of nodes. Let us define the graph's Laplacian matrix $L = D - A$ where D is a diagonal matrix with $D_{ii} = d_i = \sum_{j \neq i} A_{ij}$ the strength of node i. The normalized Laplacian matrix reads $\mathscr{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. $\mathscr{L}$ is real symmetric, therefore diagonalisable: its spectrum is composed of $(\lambda_l)_{l=1\dots N}$ its set of eigenvalues that we decide to sort: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N \leq 2$; and of χ the matrix of its normalized eigenvectors: $\chi = (\chi_1 | \chi_2 | \dots | \chi_N)$. Considering only connected graphs, the multiplicity of eigenvalue $\lambda_1 = 0$ is 1. For more properties of this spectrum, see [2] (theory) or [12, 13] (empirical). By analogy to the continuous Laplacian operator whose eigenfunctions are the continuous Fourier modes and eigenvalues their squared frequencies, χ is considered as the matrix of the graph's Fourier modes, and $\left(\sqrt{\lambda_l}\right)_{l=1\dots N}$ its set of associated "frequencies". For instance, the graph Fourier transform $\hat{f}$ of a signal f defined on the nodes of the graph reads: $\hat{f} = \chi^\top f$.

2.2 The Graph Wavelets

Graph wavelets were defined in [5] using the graph Fourier modes. Let us note $\psi_{s,a}$ the wavelet centered around node a. Its construction is based on band-pass filters defined in the graph Fourier domain, generated by stretching a band-pass filter kernel $g(\cdot)$ by a scale parameter $s > 0$. The stretched filter has matrix representation $\hat{G}_s = diag(g(s\lambda_1), \dots, g(s\lambda_N))$. Hence, the wavelet basis at scale s reads as: $\Psi_s = (\psi_{s,1} | \psi_{s,2} | \dots | \psi_{s,N}) = \chi \hat{G}_s \chi^\top$.
Here, our main use is of the localized wavelets $\psi_{s,a}$ themselves. The intuition behind this definition of wavelets on graphs is that, at small scales (small scale parameter s), the filter $g(s\,\cdot)$ is stretched out, thus letting through high frequency modes essential to good localization: corresponding wavelets extend only to their close neighbourhood in the graph. At large scales (large s) the filter function is compressed around low frequency modes and this creates wavelets encoding a coarser description of the local environment. The details of the band-pass kernel we use as well as the boundaries of the scale parameter s are discussed in section 3.1 of [18].

3 Multiscale Community Mining

3.1 Elements of the method

Clustering techniques in data-mining generally rely on three choices [8]: of feature vectors for each object under consideration; of a distance to quantify whether vectors are close or not; and of

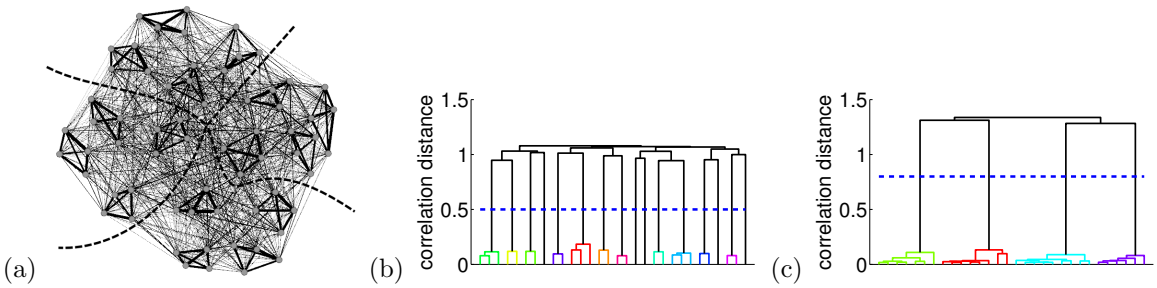

Figure 1: (a) Sketch of the benchmark graph as defined in section 4.2: each node displayed is in fact a community of 10 nodes. The thickness of each link is proportional to the total number of links between the two corresponding communities. The method outputs one dendrogram per scale. We decide to cut each dendrogram at the maximal gap between two of its nodes. We illustrate this on a dendrogram at small scale (b) and a dendrogram at large scale (c). The horizontal dashed line represents the cut.

a clustering algorithm to separate the objects in several clusters. Let us describe our multiscale community mining through these three key points.

1. Scale-dependent feature vectors. The aim is to group together nodes whose topological environments are similar. As the local information is encoded in the wavelets, the feature vector associated to each scale s and each node a is the wavelet $\psi_{s,a}$ after normalization in energy.

2. Correlation distance. We choose the correlation distance (equal to 1 - the correlation coefficient) to quantify two feature vectors' closeness. For instance, consider two nodes a and b, their distance reads: $d_s^g(a,b) = 1 - (\tilde{\psi_{s,a}})^T \tilde{\psi_{s,b}}$, where $\tilde{\psi_{s,a}}$ is the normalized wavelet. Experimentally, it yields better results than, e.g., the Euclidean distance.

3. Clustering algorithm. We use a hierarchical "average-linkage" clustering algorithm [8, 9]. This hierarchical algorithm gives a dendrogram as its output.

Where should we cut the dendrogram? As we do not know beforehand how many clusters there are in the network, we have to define a criterion to cut the dendrogram. Inspired by the gap statistics method [6], we decide to cut the dendrogram at the maximal gap between two of its nodes (see Fig. 1 for an illustration).

3.2 Protocol for multiscale community mining

The proposed method is summarized as follows.

1. Choose the scale s at which one wants to study the community structure.

2. Compute the normalized wavelets at this scale, and use them as feature vectors.

3. Cluster the feature vectors using the correlation distance and hierarchical average-linkage clustering, and output the induced dendrogram.

4. Cut the dendrogram at its maximal gap, thereby obtaining a partition of the network.

The method outputs the community structure at scale s. To obtain a proper multiscale description, repeat the procedure with several scales.

4 Detection of stable partitions

4.1 Two instability measures

Let us consider $S = \{s_1 = s_{min}, s_2, \ldots, s_M = s_{max}\}$ a set of scales. We choose them logarithmically spaced because the density of eigenvalues on the interval $[0, 2]$ is not uniform: they are much

more grouped around 1 than 0 for complex graphs with communities [12]. Therefore, a small difference at small scale (a small scale takes into account the largest eigenvalues) has a much bigger impact on the clustering than the same small difference at large scale. As for the number of scales M we decide to scan, we choose, by analogy to the classical 1-D discrete wavelets case: $M = K \log_2(N)$ where K is typically inferior to 10.

In order to detect the most relevant scales of S, we follow the following points. We first pre-select potentially stable scales (points 1 and 2), before applying a classical resampling method on each of these scales (point 3).

1. Consider the dendrogram at scale s, and δ_1 and δ_2 its first and second largest gaps. We define $\beta(s) = \frac{\delta_2(s)}{\delta_1(s)}$ an instability indicator. Indeed, if the first largest gap is clearly larger than the second (if $\beta(s)$ is close to 0) then the choice of cutting the dendrogram at δ_1 is univocal: the partition is probably stable. On the other hand, if $\beta(s)$ is close to 1, i.e. if $\delta_1 \simeq \delta_2$, then the choice of cutting the dendrogram at δ_1 is less relevant: the partition is probably unstable.

2. We look for the scales at which the corresponding partition is the more stable, i.e. the local minima of β. Some of the partitions associated to these minima may be identical. For a given partition, we choose to keep only the local minimum that has the smallest instability. Moreover, we discard all partitions that have more than $\frac{N}{2}$ communities (this makes sense because we consider only connected graphs). We thereby obtain a set of potentially interesting scales S^*.

3. Finally, we apply a resampling method [11] on these few scales: we create many (typically 100) resampled graphs by randomly adding $\pm 10\%$ to the weight of each link. Then, for each pre-selected scale $s \in S^*$, we compute the associated partition of all the resampled graphs, as well as the similarity (using the Adjusted Rand Index [7]) between all pairs of partitions. We then estimate the mean $m(s)$ and the standard deviation $\sigma(s)$ of the similarity and define a second instability indicator: $\gamma(s) = 1 - (1 - \sigma(s)) * m(s)$ ($\gamma \in [0, 1]$). For instance, at a given scale s, if all partitions of the resampled graphs are exactly the same, the similarity of all pairs of partitions is 1, and $m(s) = 1$, $\sigma(s) = 0$, i.e. $\gamma = 0$: this particular scale is very stable. The K scales that give the smallest values of $\{\gamma(s), s \in S^*\}$ are the K most relevant scales, and the associated partitions the K most relevant partitions of the graph.

This protocole outputs the K "best" scales, but it does not inform us how good the associated partitions are. In fact, the protocole will find the "best" K partitions of an Erdös-Renyi (ER) graph, even though ER graphs have no community structure at any scale. The same issue exists with methods based on modularity. Indeed, modularity maximisation has a solution even for ER graphs: but what is the threshold value of modularity after which one decides that a given partition is interesting? In the framework of this article we rephrase this in: what is the threshold value γ_t under which one may say the partition is sufficiently stable? We tackle this problem in the following with statistical tools.

4.2 A statistical test

A first observation. We consider a first graph that has a well defined multi-scale community structure that we know *a priori*. We choose here a realisation of a Sales-Pardo (SP) graph [16] with the following hierarchical community structure: there are 64 communities of 10 nodes, nested in 16 communities of 40 nodes, themselves nested in 4 communities of 160 nodes. This graph is controlled by two parameters: ρ controls how well separated the levels are, and the mean degree $\bar{k}$ controls how dense the graph is. The bigger is ρ and the smaller is $\bar{k}$, the harder it is to recover the community structures. We choose here $\rho = 1$ and $\bar{k} = 16$, which is considered as a fairly challenging set of parameters for community detection [16, 11]. This graph is represented in Fig. 1a. We compare this graph to an Erdös-Renyi (ER) random graph with same mean degree, i.e. with a probability of existence of link $p = \frac{16}{640-1} \simeq 0.025$. By construction, this graph is supposed to have no community structure at any scale. Fig. 2 displays the results.

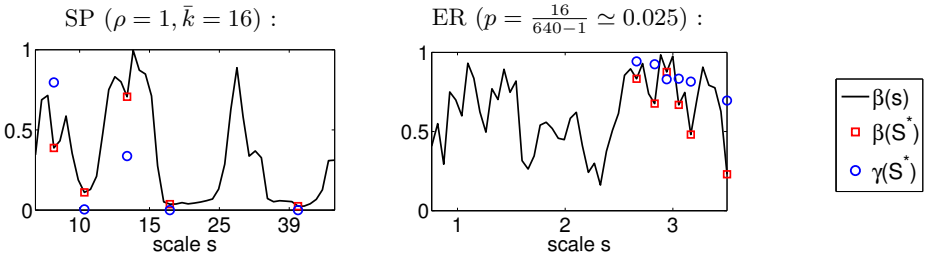

Figure 2: Comparison between a SP graph (left) and an ER graph (right) with same number of nodes and same average degree. The instability $\beta(s)$ is the black continuous line, the red squares represent $\{\beta(s), s \in S^*\}$ and the blue circles $\{\gamma(s), s \in S^*\}$.

The results are fundamentally different and two arguments tell us that this ER graph has no community structure. First of all, the instability $\beta(s)$ of the ER graph has many local minima and none of them seems more important than the other. Whereas the SP graph shows a much more regular $\beta(s)$. More important, we see that the three smallest values of $\gamma(S^*)$ for the SP graph are null, or very close to zero: the associated partitions are very stable (and they actually correspond to the three "true" levels of description). Whereas $\gamma(S^*)$ remains very high for the ER graph at all scales. In order to give statistical significance to this first observation, we propose a statistical test to automatically decide which scales of a given graph are statistically significant.

The test. Consider a graph $\mathcal{G}_0 = (V_0, E_0, A_0)$ and its instability function $\gamma_0(S_0^*)$. The goal is to test each scale $s \in S_0^*$ and decide whether it is interesting or not. To this end, we propose the following test:

1. Formulate the Null Hypothesis H0: $\mathcal{G}_0$ has no community structure at any scale.

2. Generate a large number R of ER graphs with same mean degree.

3. Compute their $\gamma(S^*)$ and obtain the general statistics $\mathcal{S}_\gamma$ of the values of γ.

4. For each scale $s \in S_0^*$, if $\gamma_0(s)$ is lower than the first α-quantile of $\mathcal{S}_\gamma$, then we reject H0 with a confidence of $1 - \frac{1}{\alpha}$ (if $R >> \alpha$): $\mathcal{G}_0$ has a community structure at this particular scale. Typically, we use $\alpha = 100$ and $R = 1000$.

For instance, if we consider the SP graph of the last section, and apply this statistical test, the first 100-quantile of $\mathcal{S}_\gamma$ for $\bar{k} = 16$ is $\gamma_t = 0.32$. In this context, out of the 5 scales of S^* and their associated instability values $\gamma(S^*)$ (shown on the left plot of Fig. 2 with blue circles), 4 are under γ_t: 3 have zero instability and they correspond exactly to the three levels of description, one at $\gamma = 0.25$ is close to the threshold, and corresponds to a partition of 42 communities where the smallest communities of 10 nodes are partially aggregated together. This partition is understandable but we still consider it as a false positive. In information retrieval, *precision* and *recall* are two widely-used measures to characterize such tests. In our context, precision is the fraction of retrieved scales that are relevant, while recall is the fraction of relevant scales that are retrieved. In this particular example, we have a precision of 0.75 and a recall of 1.

Performance of the test. We follow this procedure on 200 realisations of a SP random graph with $\rho = 1$ and $\bar{k} = 16$ and obtain the Precision and Recall diagram of Fig.3.b. In average, we detect the three "true" partitions (Recall is almost always equal to 1), but we also have around three false positives (Precision around 0.5). These false positives correspond to partitions that are very similar to the three "true" ones (sometimes only a difference of one node permutation) but not perfectly equal. The precision is thereby conservative and remains low. On the same figure, we also show the results for $\rho = 0.8$ (a) and $\rho = 1.4$ (c). As expected, as ρ decreases and the different hierarchical levels become more and more separated, the test becomes more reliable.

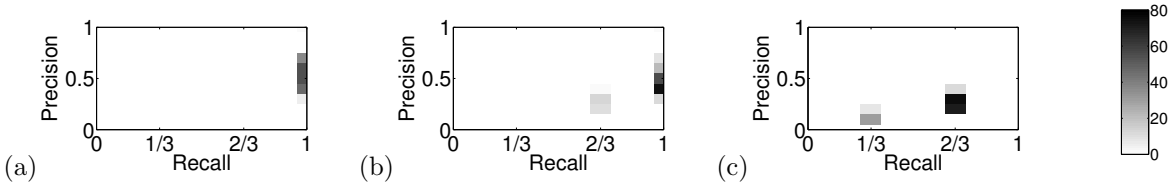

Figure 3: Precision and Recall diagramm for $\bar{k} = 16$ and (a) $\rho = 0.8$, (b) $\rho = 1$, and (c) $\rho = 1.4$ for 200 realisations of a Sales-Pardo random graph.

5 Conclusion

This work is a methodological contribution to multi scale community mining as well as the first work -up to our knowledge- that takes advantage of the local information encoded in graph wavelets to tackle classical complex systems issues. This method not only gives a multi-scale view of the network, but it also automatically gives an intrinsic value (via notions of partition instability) to each scale. Due to lack of space, we do not show here applications of this method to real graphs but preliminary results are fruitful and promising, and will be the object of future publication.

References

[1] A. Arenas, A. Fernandez, and S. Gomez. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10(5):053039, 2008.

[2] F.R.K. Chung. *Spectral graph theory*. Number 92. Amer Mathematical Society, 1997.

[3] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[4] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *PNAS*, 104(1):36, 2007.

[5] D.K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

[6] Trevor Hastie, Robert Tibshirani, Jerome Friedman, et al. The elements of statistical learning: data mining, inference, and prediction, 2001.

[7] L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[8] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[9] B. King. Step-wise clustering procedures. *Journal of the American Statistical Association*, 62(317):86–101, 1967.

[10] J.M. Kumpula, J. Saramäki, K. Kaski, and J. Kertesz. Limited resolution in complex network community detection with Potts model approach. *Eur. Phys. J. B*, 56(1):41–45, 2007.

[11] R. Lambiotte. Multi-scale modularity in complex networks. In *Proc. 8th Int. Symp. WiOpt*, pages 546–553, 2010.

[12] Marija Mitrovic and Bosiljka Tadic. Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. *Phys. Rev. E*, 80:026123, Aug 2009.

[13] Y. Nakatsukasa, N. Saito, and E. Woei. Mysteries around the graph Laplacian eigenvalue 4. *Arxiv preprint arXiv:1112.4526*, 2011.

[14] M.E.J. Newman. Modularity and community structure in networks. *PNAS*, 103(23):8577, 2006.

[15] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006.

[16] M. Sales-Pardo, R. Guimera, A.A. Moreira, and L.A.N. Amaral. Extracting the hierarchical organization of complex systems. *PNAS*, 104(39):15224–15229, 2007.

[17] M.T. Schaub, J.C. Delvenne, S.N. Yaliraki, and M. Barahona. Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. *PloS one*, 7(2):e32210, 2012.

[18] N. Tremblay and P. Borgnat. Multiscale community mining in networks using spectral graph wavelets. *arXiv:1212.0689*, 2012.