Magnetic eigenmaps for community detection in directed networks

Michaël Fanuel,* Carlos M. Alaíz,† and Johan A. K. Suykens‡
KU Leuven, Department of Electrical Engineering (ESAT)
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics,
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
(Dated: July 15, 2016)

Communities in directed networks have often been characterized as regions with a high density of links, or as sets of nodes with certain patterns of connection. Our approach for community detection combines the optimization of a quality function and a spectral clustering of a deformation of the combinatorial Laplacian, the so-called magnetic Laplacian. The eigenfunctions of the magnetic Laplacian, that we call magnetic eigenmaps, incorporate structural information. Hence, using the magnetic eigenmaps, dense communities including directed cycles can be revealed as well as "role" communities in networks with a running flow, usually discovered thanks to mixture models. Furthermore, in the spirit of the Markov stability method, an approach for studying communities at different energy levels in the network is put forward, based on a quantum mechanical system at finite temperature.

I. INTRODUCTION

The investigation of network structure has been performed with the help of a wealth of techniques [1] with various advantages, a famous example being modularity optimization [2, 3] in undirected networks. One of the main disadvantages that some of these methods have to face is the resolution limit. In particular, extensions of modularity for changing the resolution limit have been developed [4], whereas other related methods are inspired from statistical physics. Indeed, a general framework was developed in [5], and later, a specific Potts model was explained to have no resolution limit [6]. On the other hand, among the community detection methods, the discrete Laplacians in undirected networks [7, 8] have also been used in order to unravel the graph structures. A common feature of these definitions of communities is that they rely on the density of links in the network.

Recently, the community structure of complex networks has been studied with the help of several dynamical processes [9, 10]. For instance, flow communities are detected in the information theory based framework "Infomap" of [10]. In this dynamical paradigm, the "Markov stability" method uses a dynamical process governed by a random walk or a continuous time Kolmogorov equation to unravel the network geometry at different scales [11, 12]. An asset of this method is that it naturally contains previous methods such as modularity optimization and Fiedler partitioning. As a matter of fact, time evolution allows to span different scales, since at different times the eigenmodes of the dynamics have another relative importance. These eigenmodes incorporate structural information.

The importance of diffusion processes for investigating network structure has been highlighted in recent works [9, 12, 13]. Furthermore, these approaches have emphasized the relevance of a different type of community structure: the flow-based communities, which are intuitively defined as the structures retaining the diffusion for a certain period of time. Nonetheless, while the diffusion processes on networks are increasingly understood in the case of undirected networks, there has been less focus in the past on diffusion on directed networks. Many of the networks of interest in biology, internet or social sciences are directed and have attracted attention in the physics literature [14–18]. In dynamical frameworks, where the existence of a stationary distribution is crucial, directed networks are explored thanks to a random walk with teleportation as for instance in the Markov stability framework [12], the Infomap method [10], or in the definition of the LinkRank method [16]. A natural question is: can we spare the use of a random walk with teleportation?

Although alternatives approaches to Markovian processes have also been studied in [19], we propose here a method based on quantum mechanics to uncover communities in directed networks. Our approach relies on a deformation of the combinatorial Laplacian suited to directed graphs and does not fit into the theory of Markov processes. More explicitly, the magnetic Laplacian [20–22] is a generalization of the combinatorial Laplacian to a line bundle, and can be understood as describing the dynamics of a free quantum mechanical particle on a graph under the influence of magnetic fluxes passing through the cycles in the network. It is widely known in the physics community that the presence of a magnetic flux can be detected in quantum mechanics thanks to the Aharonov-Bohm effect [23]. Whereas the combinatorial Laplacian was designed long ago as a discrete differential operator, it has been used only more recently for community detection in networks [7, 8, 24]. Similarly, the magnetic Laplacian is a well-known object in mathematical and condensed matter physics [20, 21], however, to the best of our knowledge, it has never been used for

*Electronic address: Michael.Fanuel@esat.kuleuven.be
†Electronic address: CMAlaiz@esat.kuleuven.be
‡Electronic address: Johan.Suykens@esat.kuleuven.be

community detection in directed networks, apart from the purely theoretical work [25]. As a matter of fact, the major asset of our quantum mechanical approach over dynamical frameworks relying on random walks or Kolmogorov equations is that the generator of the dynamics can be here a complex valued Hermitian operator. Actually, the question of the existence of a stationary distribution is irrelevant in our case.

In Section II, the properties of the magnetic Laplacian will be reviewed with a particular emphasis on its connection to the network topology. Then, the so-called flux communities will be introduced in Section III, and a method for uncovering them will be discussed in Section IV. Subsequently, a multiscale method for studying flux communities will be proposed in Section V, while the results will be illustrated on artificial and real-life networks in Section VI. An analogue of the spectral clustering method in the complex domain will be introduced in Section VII as a tool to uncover role communities in directed networks with a running flow. Finally, the paper will end with some conclusions in Section VIII.

A summary of the different methods that we propose is included in Table I.

Notation In the sequel, a directed network will be considered, with its set of nodes V and links E. An undirected link between i and j will be denoted by $\{i,j\}$, while a directed link will be written $[i,j]$.

II. MAGNETIC LAPLACIAN AND LINE BUNDLES

A conventional manner to study the structure of directed networks is to symmetrize its weight matrix in order to make the network undirected so that a spectral method, based on the combinatorial Laplacian

$$\left(\hat{\mathcal{L}}^{\mathrm{C}}\psi\right)(i) = \sum_j w_{\mathrm{s}}(i,j)\left(\psi(i) - \psi(j)\right),$$

can be used to partition the network. In general, the function of the nodes ψ is taken to be real-valued. Another approach consists in defining an analogue of the PageRank random walk on the network, where the walker follows exclusively the edge directions. For technical reasons, a teleportation parameter has to be added so that a stationary distribution can be defined, if the network is not strongly connected and aperiodic. In this article, we consider an intermediate possibility which uses the symmetrization of the weights by keeping relevant information about the edge directions in an edge flow 1-form. Indeed, decomposing the weight matrix, we define the symmetrized weight $w_{\mathrm{s}}(i,j) = \left(w(i,j) + w(j,i)\right)/2$ and the skew-symmetric non-dimensional function of the oriented links $a(i,j)$, satisfying $a(i,j) = 1$ if $i \to j$, $a(i,j) = -1$ if $j \to i$, and $a(i,j) = 0$ if $\{i,j\}$ is reciprocal. Separating the asymmetric part of the weight matrix was already proposed in [26], however no similar deformation of the combinatorial Laplacian was proposed earlier in the literature.

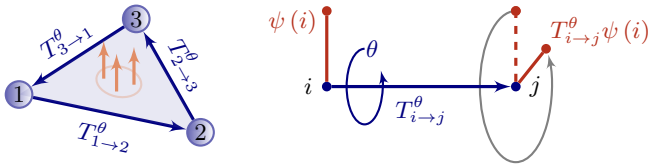


FIG. 1: Parallel transport along a directed cycle (here a triangle is chosen as an example), and an illustration of the effect of a transport along a directed link. It is intuively clear that $T^{\theta}_{i\to j}$ assigns a rotation to a directed link.

A. Magnetic Laplacian

As a consequence of these elementary remarks, we propose to describe the Hamiltonian of our quantum mechanical system as the Hamiltonian of a free charged particle on an undirected network in presence of a space varying magnetic field, given by the so-called "magnetic Laplacian" [20–22],

$$\left(\hat{\mathcal{L}}_{a,\mathrm{i}\theta}\psi\right)(i) = \sum_j w_{\mathrm{s}}(i,j)\left(\psi(i) - T^{\theta}_{j\to i}\psi(j)\right), \qquad (1)$$

with $T^{\theta}_{j\to i} = \exp\left(\mathrm{i}\theta a(j,i)\right)$, depending on a real deformation parameter θ interpreted as the electric charge of a particle. Obviously, the combinatorial Laplacian of the symmetrized weight matrix is recovered if either $\theta = 0$ or $a = 0$, i.e. $\hat{\mathcal{L}}^{\mathrm{C}} = \hat{\mathcal{L}}_{a,0} = \hat{\mathcal{L}}_{0,\theta}$. The dynamics is invariant under $\theta \to \theta + 2\pi$, and therefore the parameter θ is interpreted as being an angle, as illustrated in Figure 1. We may call this version of electrodynamics "compact electrodynamics".

The self-adjoint operator $\hat{H} = \hat{\mathcal{L}}_{a,\mathrm{i}\theta}$ is actually positive semi-definite and can be understood as a deformation of the combinatorial Laplacian on the undirected graph. More details are given in Appendix A 1. Furthermore, it is a special case of vector bundle Laplacian [27, 28]. Indeed, the factor $T^{\theta}_{j\to i} = \exp\left(\mathrm{i}\theta a(j,i)\right)$ is interpreted as a unitary parallel transporter. Although the magnetic Laplacian was already present in the physics literature long ago, it can be also understood in the "connection Laplacian" framework of Singer and Wu [29, 30], but for the case of a complex unitary representation of U(1) instead of real representations of SO(d), as explained in Appendix A 2. Incidentally, Cucuringu has proposed recently a ranking algorithm based on a U(1) connection Laplacian [31].

B. Discrete Hodge theory and gauge transformations

The customary gauge transformation $a' = a + \mathrm{d}h$ of a magnetic Laplacian, for a discrete gradient $\mathrm{d}h(i,j) =$

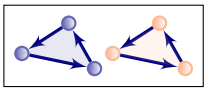
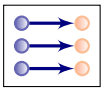
	Flux		Flow
	Energy λ_ℓ	Finite temperature $1/\beta$	
Illustration			
Correlation	$x_{\ell,\theta}(i,j) = \mathrm{Re}\big(\chi_{\theta,\ell}(i)\, T^{\theta}_{i\to j}\chi^{*}_{\theta,\ell}(j)\big)$	$x_{\beta,\theta}(i,j) = \mathrm{Re}\Big(T^{\theta}_{i\to j}\,(\rho_\beta)_{i,j}\Big)$	$\xi_{\theta,0}(i,j) = \mathrm{Re}\big(\chi_{\theta,0}(i)\,\chi^{*}_{\theta,0}(j)\big)$
Adj. matrix	$X_{\ell,\theta}(i,j) = x_{\ell,\theta}(i,j) + \lvert x_{\ell,\theta}(i,j)\rvert$	$X_{\beta,\theta}(i,j) = x_{\ell,\beta}(i,j) + \lvert x_{\ell,\beta}(i,j)\rvert$	$\Xi_{\theta,0}(i,j) = \xi_{\theta,0}(i,j) + \lvert\xi_{\theta,0}(i,j)\rvert$

TABLE I: Summary of the community detection methods using the magnetic eigenmaps $\chi_{\theta,\ell}$.

$h(j) - h(i)$, gives a unitary equivalent operator

$$\hat{\mathcal{L}}_{a',\mathrm{i}\theta} = e^{-\mathrm{i}\theta h} \circ \hat{\mathcal{L}}_{a,\mathrm{i}\theta} \circ e^{\mathrm{i}\theta h}, \tag{2}$$

which obviously shares the same spectrum. Graph gauge theory has been already discussed for instance in [32], where no explicit application to directed networks was presented.

The magnetic Laplacian is closely related to the topology of the network and we will take advantage of this feature in order to uncover various types of community structures. Actually, the topology of graphs and simplicial complexes has been shown to incorporate relevant information about data. Here, because electromagnetism is associated to the gauge group U(1), the relevant topological structures uncovered are the cycles. In this spirit, we can perform the Hodge decomposition [33] of the edge flow

$$a = a_{\mathrm{H}} + \mathrm{d}h_a + \mathrm{d}^{\star}\omega_a = a_{\mathrm{M}} + \mathrm{d}h_a,$$

where a_{H} is the harmonic component related to the existence of magnetic flux through undirected k-cycles ($k > 3$) in the network and $\mathrm{d}^{\star}\omega_a$ is a co-exact form related to a magnetic flux through the undirected triangles of the graph. This Hodge decomposition allows to consider components of the edge flow related to cycles in the network. Incidentally, these topologically non-trivial components actually play an important role in the non-perturbative dynamics of gauge theories [34].

An asset of our description is that the magnetic Laplacian will emphasize the importance of groups of nodes organized as a directed cycle, which is actually a structure causing problems when the directed network is explored by means of a random walk.

III. AHARONOV-BOHM EFFECT AND FLUX COMMUNITIES

A. Directed networks and Aharonov-Bohm phases

Considering the Schrödinger equation on a network with the magnetic Laplacian as Hamiltonian, we can write the conservation of probabilities in terms of the divergence of a probability current. Indeed, the time evolution of the probability distribution $p_t(i) = |\psi(i,t)|^2$ on the network is governed by

$$\frac{\partial}{\partial t} p_t(i) = \sum_j w_{\mathrm{s}}(i,j)\, J_t(i,j), \tag{3}$$

with the real valued and gauge invariant probability current

$$J_t(i,j) = 2\,\mathrm{Im}\big(\psi(i,t) T^{\theta}_{i\to j}\psi^{*}(j,t)\big), \tag{4}$$

where $T^{\theta}_{i\to j}\psi(i,t)$ is the value of ψ at node i and time t, transported at node j.

Although we do not have here a diffusion-like dynamics, (3) governs the time evolution of a probability. Nevertheless, we are interested in the phase correlations of the nodes and therefore we are going to study the real part in (4) instead of the imaginary part. Incidentally, since we are going to consider first the case where the state ψ is an eigenstate of the magnetic Laplacian, we will not consider the associated probability current (4). Let us consider the orthonormal basis of functions on the nodes $\langle\delta_i|\delta_j\rangle = \delta_{i,j}$, localized on a specific node ($\delta_{i,j}$ is the Krönecker delta). The matrix elements of the magnetic Laplacian are

$$\langle\delta_i|\hat{\mathcal{L}}_{a,\mathrm{i}\theta}|\delta_j\rangle = \sum_{\ell=0}^{N-1} \lambda_\ell \chi_{\theta,\ell}(i)\,\chi^{*}_{\theta,\ell}(j), \tag{5}$$

where $\chi_{\theta,\ell}$ is the eigenfunction associated to the eigenvalue λ_ℓ, satisfying $\lambda_0 \le \lambda_1 \le \lambda_2 \le \dots$. These eigenfunctions are called in this paper the *magnetic eigenmaps*.

Actually, it is well-known that the eigenfunctions $\chi_{\theta,\ell}$ are stationary states, i.e., their probability density does not evolve in time. In fact, the rhs of (3) vanishes. Nevertheless, in general, the real part of the same matrix elements can be interesting. Let us first assume that the eigenvalue λ_ℓ is non degenerate. Hence, we want to find the partition which maximizes the following correlation due to the magnetic field:

$$x^{(a)}_{\ell,\theta}(i,j) = \mathrm{Re}\big(\chi_{\theta,\ell}(i)\, T^{\theta}_{i\to j}\chi^{*}_{\theta,\ell}(j)\big), \tag{6}$$

which can be negative. Therefore, we introduce the following matrix with positive matrix elements:

$$X^{(a)}_{\theta,\ell}(i,j) = \left|\chi_{\theta,\ell}(i)\,\chi^*_{\theta,\ell}(j)\right| + x^{(a)}_{\theta,\ell}(i,j) \tag{7}$$

for all $i,j \in V$ such that $\{i,j\} \in E$, and $X^{(a)}_{\ell,\theta}(i,j) = 0$ otherwise. The matrix elements $X^{(a)}_{\ell,\theta}(i,j)$ may be understood as link weights, and they are gauge invariant, i.e. the correlation is the same either if $a = a_{\mathrm{M}} + \mathrm{d}h$ or a_{M} are used to compute it. Hence, we have

$$X^{(a)}_{\theta,\ell}(i,j) = X^{(a_{\mathrm{M}})}_{\theta,\ell}(i,j).$$

For the sake of simplicity, we will now omit the superscript indicating the dependence on a and merely write $X_{\theta,\ell} = X^{(a)}_{\theta,\ell}$. Let us explain the definition of the correlation of (6) with the example of Figure 1. First of all, finding the eigenvector of the magnetic Laplacian with the lowest eigenvalue can be formulated as the following optimization problems:

$$\begin{aligned}&\min_{\chi}\ \langle\chi|\hat{\mathcal{L}}_{a,\mathrm{i}\theta}\chi\rangle \text{ s.t. } \langle\chi|\chi\rangle = 1\\ \equiv &\min_{\chi} \sum_{\{i,j\}\in E} w_{\mathrm{s}}(i,j)\left|\left(\mathrm{D}_{a,\mathrm{i}\theta}\chi\right)(i,j)\right|^2 \text{ s.t. } \langle\chi|\chi\rangle = 1,\end{aligned} \tag{8}$$

where the covariant derivative is defined as

$$\left(\mathrm{D}_{a,\theta}\chi\right)(i,j) = e^{\mathrm{i}\theta a(i,j)/2}\chi(j) - e^{-\mathrm{i}\theta a(i,j)/2}\chi(i).$$

Notice that the covariance property of this gradient under the transformation $a' = a + \mathrm{d}h$ reads

$$\left(\mathrm{D}_{a',\theta}\chi\right)(i,j) = e^{-\mathrm{i}\theta(h(i)+h(j))/2}\left(\mathrm{D}_{a,\theta}e^{\mathrm{i}\theta h}\chi\right)(i,j).$$

In analogy with the case of the combinatorial Laplacian, the solution to this minimization problem will satisfy $\left(\mathrm{D}_{a,\mathrm{i}\theta}\chi\right)(i,j) \approx 0$. Let us explain why the weight $x^{(a)}_{0,\theta}(i,j)$ is large if $[i,j]$ is part of a directed cycle with an example. More precisely, in the case of the directed triangle of Figure 1, if the lowest energy eigenfunction $\chi_{\theta,0}$ should satisfy $\left(\mathrm{D}_{a,\mathrm{i}\theta}\chi_{\theta,0}\right)(i,j) = 0$ for the three links, then

$$\begin{cases}\chi_{\theta,0}(3) = e^{\mathrm{i}\theta}\chi_{\theta,0}(2),\\ \chi_{\theta,0}(2) = e^{\mathrm{i}\theta}\chi_{\theta,0}(1),\\ \chi_{\theta,0}(1) = e^{\mathrm{i}\theta}\chi_{\theta,0}(3).\end{cases} \tag{9}$$

Hence, $\exp(\mathrm{i}3\theta) = 1$, so θ should be selected as $\theta = 2\pi/3$. Indeed, in that case we have

$$\begin{aligned}\chi_{\theta,0}(1)T^{\theta}_{1\to2}\chi^*_{\theta,0}(2) &= \chi_{\theta,0}(1)e^{\mathrm{i}\theta}\chi^*_{\theta,0}(2)\\ &= \left(\chi_{\theta,0}(1)e^{\mathrm{i}\theta}\right)\left(e^{\mathrm{i}\theta}\chi_{\theta,0}(1)\right)^*\\ &= \left|\chi_{\theta,0}(1)\right|^2.\end{aligned}$$

Since this is valid for the directed links $1\to2$, $2\to3$ and $3\to1$, the correlations (6) between subsequent nodes in the triangle are maximal, i.e.

$$x^{(a)}_{0,\theta}(1,2) = x^{(a)}_{0,\theta}(2,3) = x^{(a)}_{0,\theta}(3,1) = \left|\chi_{\theta,0}(1)\right|^2.$$

Hence, this triangle will be considered as a flux community.

In the case of an eigenspace of dimension larger than one, i.e. when λ_ℓ is degenerate, the relevant matrix element is $\mathbb{P}^{(\ell)}_{i,j}T^{\theta}_{i\to j}$, where $\hat{\mathbb{P}}^{(\ell)}$ is the projector on this eigenspace. Hence, we define in this case

$$X_{\theta,\ell}(i,j) = \left|\mathbb{P}^{(\ell)}_{i,j}\right| + \mathrm{Re}\left(\mathbb{P}^{(\ell)}_{i,j}T^{\theta}_{i\to j}\right),$$

for all $i,j \in V$ such that $\{i,j\} \in E$ and $X_{\ell,\theta}(i,j) = 0$ otherwise. In practice, eigenspaces are rarely exactly degenerate. However, some eigenvalues may be approximately equal if a threshold is defined. The same issue arises in the case of spectral clustering of the combinatorial Laplacian. In order to circumvent this difficulty, it is possible to avoid the computation of the eigenvalues and use a method based on a stability criterion with respect to a change in a parameter as in [11]. A similar idea can be used here, as explained in Section V.

Let us explain now why choosing special values of θ may be interesting. We shall focus on the first eigenvector with the smallest eigenvalue, the minimizer of (8). Considering again the example of the directed triangle illustrated in Figure 1 and combining the three equations of 9, we find the condition of flux quantisation

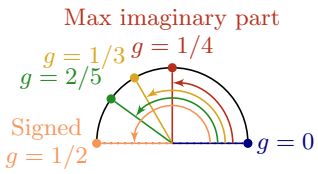
$$T^{\theta}_{1\to2}T^{\theta}_{2\to3}T^{\theta}_{3\to1} = 1,$$

i.e. $\exp(\mathrm{i}\theta\Phi(1,2,3)) = 1$ with $\Phi(1,2,3) = a(1,2) + a(2,3) + a(3,1)$. There is a similar relation for directed n-cycles, which is called "consistency" in the case of connection graphs in [30]. Therefore, if the links are in only one direction, i.e. $a(i,j) = \pm1$, and not reciprocal, then θ is taken such that $\theta\Phi(1,2,3,\ldots,n) = 0 \mod 2\pi$. This means that we can choose the parameter θ to take quantised values,

$$\theta = \frac{2\pi k}{n}, \quad \frac{k}{n} \notin \mathbb{Z},$$

in order to detect n-cycles with constant phase differences. A trivial consequence is that the value $X_{\theta,\ell}(i,j)$, on an edge which is not part of a flux community, will be suppressed. Hence, the unitary factor in (6) implementing the parallel transport on the line bundle is fundamentally important. In quantum mechanics, for instance in the case of Abrikosov vortices in Type II superconductors, it is well-known that the magnetic flux is quantised. We have here an analogous condition on the product of the "electric" charge and the magnetic flux. Moreover, in order to detect communities of the size given by a multiple of n, i.e. communities of a given magnetic flux, we prescribe to choose the quantised charge $\theta = 2\pi/n$, as shown in Figure 2. In the sequel, we will define the coupling constant

$$g = \frac{\theta}{2\pi}.$$



g	Structures
0	dense clusters
1/4	2, 4, 3-cycles
1/3	3, 2-cycles
2/5	3-cycles

FIG. 2: Different particular values of the electric charge $\theta = 2\pi g$. The lower half circle correspond to a network with opposite link directions.

B. Normalized magnetic Laplacian

The magnetic Laplacian was originally constructed in the case of a quantum particle on a lattice where, of course, the degree of the nodes is constant. Real-life networks have often an inhomogeneous degree distribution. It may be interesting to use instead of the magnetic Laplacian of (1) a degree normalized version

$$\hat{\mathcal{L}}^{\mathrm{N}}_{a,\mathrm{i}\theta} = \deg_{\mathrm{s}}^{-1/2} \circ \hat{\mathcal{L}}_{a,\mathrm{i}\theta} \circ \deg_{\mathrm{s}}^{-1/2}, \tag{10}$$

with the degree $\deg_{\mathrm{s}}(i) = \sum_j w_{\mathrm{s}}(i,j)$ associated to the symmetrized weight. The normalization can be understood as changing the definition of the covariant derivative by $\mathrm{D}_{a,\theta} \to \mathrm{D}^{\mathrm{N}}_{a,\theta} = \mathrm{D}_{a,\theta} \circ \deg_{\mathrm{s}}^{-1/2}$, which can be alternatively understood as a change of measure [35] or reweighting the inner product $\langle\psi|\psi'\rangle_{0,N} \triangleq \sum_{i\in V} \deg_{\mathrm{s}}(i)\, \psi^*(i)\, \psi'(i)$, so that nodes with a large degree have a larger weight. Loosely speaking, the upshot is that nodes with a large degree are effectively considered as a set of several nodes with a smaller degree. Notice that this construction is similar to the normalized version of the combinatorial Laplacian.

IV. CLUSTERING THE AUTO-CORRELATION OF THE MAGNETIC EIGENMAPS

Our proposal can be summarized as follows: from the directed network, a weighted undirected graph is constructed, so that the weights of the links emphasize certain structures describing flux communities. This novel weighted network can then be studied with a community detection method. In this paper, we propose the use of undirected modularity, however other methods could as well be used, with possibly different outcomes.

A. Directed networks and $g \neq 0$

Fixing a partition of the network C constituted of communities $c \in C$, we propose to cluster the network by maximizing

$$P_{\theta,\ell}(C) = \sum_{c\in C} \sum_{i,j\in c} \left(X_{\theta,\ell}(i,j) - p_{\theta,\ell}(i,j) \right),$$

with the configuration null model $p_{\theta,\ell}(i,j) = k_i k_j / 2m$, with $k_i = \sum_j X_{\theta,\ell}(i,j)$ and $2m = \sum_i k_i$. As a consequence, it is possible to use a generalized Louvain method [36] to find the optimal partition based on the following customary formulation in terms of matrices:

$$P_{\theta,\ell}(C) = \mathrm{Tr}\big(\mathbf{H}^T \left[\mathbf{X}_{\theta,\ell} - \mathbf{p}_{\theta,\ell}\right] \mathbf{H}\big), \tag{11}$$

where $\mathbf{H} \in \mathbb{R}^{N\times|C|}$ is an indicator matrix of the communities in the partition, whose element $H_{i,j} = 1$ if node i belongs to community j and $H_{i,j} = 0$ otherwise. Finding the partition maximizing the quality function (11) is possible thanks to a greedy optimization algorithm[1].

B. Undirected networks and $g = 0$

In the case of an undirected graph, i.e. $a = 0$, this spectral clustering method yields the condition $\varphi_{\theta,\ell}(i) \approx \varphi_{\theta,\ell}(j) \mod 2\pi$ for any i and j in the same cluster, where $\varphi_{\theta,\ell} = \mathrm{phase}(\chi_{\theta,\ell})$. Therefore, because the eigenvectors can be made real, this procedure is equivalent to the grouping of nodes in two communities according to $\mathrm{sign}(\chi_\ell)$, which is the so-called Fiedler partition. Moreover, in the case $a = 0$, the matrix elements in (7) reduces simply to

$$X^{(0)}_{0,\ell=0}(i,j) \propto \delta\left(\{i,j\} \in E\right),$$

for any node i and j, so that $X^{(0)}_{0,\ell=0}$ is a binarized form of the symmetrized weight matrix w_{s}. Therefore, optimizing the modularity of the similarity matrix elements (7) using (11) is equivalent to the modularity optimization of the undirected graph defined as the binarized skeleton of the directed network.

C. Signed Laplacian and $g = 1/2$

For an electric charge $\theta = \pi$, the magnetic Laplacian is actually equal to a so-called signed Laplacian [37], associated to a specific signed network. Indeed, the parallel transporter is then real $T^{\pi}_{j\to i} = \exp(\mathrm{i}\pi a(j,i)) = T^{\pi}_{i\to j}$, so that $T^{\pi}_{j\to i} = 1$ if the link $[i,j]$ is reciprocal, and $T^{\pi}_{j\to i} = -1$ if the link $[i,j]$ is only in one direction. In fact, the peculiar value $\theta = \pi$ can be treated as a signed network were all reciprocal links are positive whereas all other links are negative (see Figure 3). Therefore, our community detection method will uncover dominantly the regions of reciprocal edges, i.e. 2-cycles.

[1] We used the code available at `http://netwiki.amath.unc.edu/GenLouvain/GenLouvain`), called `genLouvain`, used in [4]

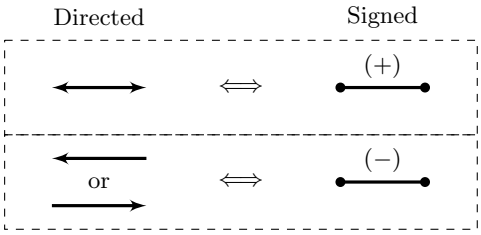
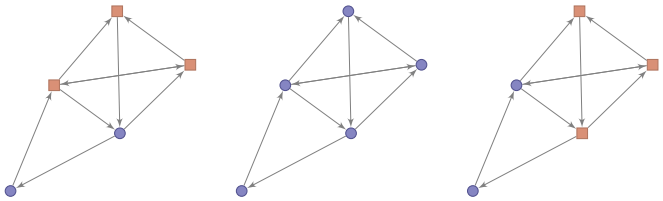


FIG. 3: The magnetic Laplacian with $\theta = \pi$ is equivalent to a signed Laplacian of the undirected signed network obtained as illustrated above.



(a) $\ell = 0$. (b) $\ell = 1$. (c) $\ell = 2$.

FIG. 4: Partitions of an artificial network with a reciprocal link for $g = 1/4$, with the normalized magnetic Laplacian.

D. The particular case $g = 1/4$

Considering for simplicity a non-degenerate eigenvalue λ_ℓ, we study the particular case $g = 1/4$, so that the parallel transporter is pure imaginary $T^{\pi/2}_{j\to i} = \mathrm{i}a(j,i)$. Indeed, this case corresponds to the maximal deformation of the combinatorial Laplacian in the complex domain. The understanding of the importance of the edge directions is made easier because $a(i,j) = \pm 1$ indicates the directions of $[i,j] \in E$. We pretend that most of the effect of the edge directions is encoded in this difference of Aharonov-Bohm phases, that we have to compare with the sign of $a(i,j)$ indicating the directionality. Actually, the matrix elements of (7) satisfy

$$\frac{X_{\pi/2,\ell}(i,j)}{|\chi_{\theta,\ell}(i)||\chi_{\theta,\ell}(j)|} = 1 + a(i,j)\sin\left(\Delta\varphi_{\pi/2,\ell}(i,j)\right), \quad (12)$$

with the phase difference $\Delta\varphi_{\pi/2,\ell}(i,j) = \varphi_{\pi/2,\ell}(j) - \varphi_{\pi/2,\ell}(i)$. Let us discuss the interest of this particular expression. If there is a directed edge from i to j, i.e. $a(i,j) = 1$, then the value of (12) will be large if the phase difference is small, i.e. $0 \leq \Delta\varphi_{\pi/2,\ell}(i,j) < \pi$. A large value of $X_{\pi/2,\ell}(i,j)$ is expected if the nodes i and j belong to the same flux community. Otherwise, if the phase difference is too large, the value of (12) will be smaller and it will be more likely that the nodes i and j belong to different flux communities. Hence, the value of the charge $\theta = \pi/2$ is plausibly going to be the most robust choice for uncovering flux communities of different sizes. In the sequel, this value will always be chosen in the absence of reciprocal links. In the presence of reciprocal links, i.e. 2-cycles, the choice $\theta = \pi/2$ seems to give a large relative weight to the links satisfying $a(i,j) = 0$. If, for instance, we want to give more importance to directed 3-cycles, the choice $\theta = \pi/2$ may not be appropriate (actually, in real-life networks directed triangles, and more generally motifs, are thought to be very important for our understanding of social or biological networks).

E. Emphasis of reciprocal links and 3-cycles

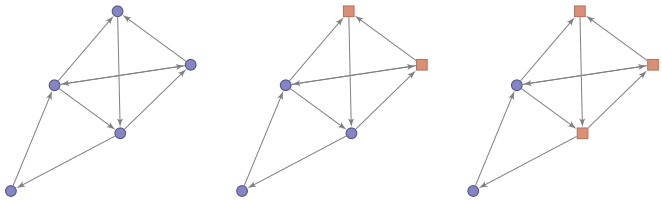
Recently, community detection techniques emphasizing the important of specific local structures have been proposed. For instance, it has been suggested to improve the methods relying on the density of links by incorporating information about triangles in the network [38, 39]. Furthermore, directed triangles in directed graphs were shown to be relevant for community detection in [40, 41].

The edge flow 1-form was defined so that $a(i,j) = \pm 1$ on a directed link, and $a(i,j) = 0$ on a reciprocal link. Let us denote the electric charge $\theta = 2\pi g$. It is clear that only the product $2\pi g a(i,j)$ is important, and we have shown that, for efficiency, the value of the charge has to be a ratio of integers. Choosing $g > 1/2$ is equivalent to a flip of the directions of the directed links, and therefore, we restrict ourselves to $g < 1/2$. Actually, the structures such as directed n-cycles will be emphasized by a choice $g = k/n$. In practice, we assume that directed n-cycles with $n \leq 5$ constitute the most significant structures in the context of community detection. Incidentally, directed triangles seem to incorporate a lot of information as explained in the recent papers [40, 41].

The choice that we recommend for networks without reciprocal links is $g = 1/4$, because it corresponds to the deformation of the combinatorial Laplacian with maximal imaginary part. We observe empirically that, if reciprocal links are present, the value $g = 1/4$ will give a relatively large weight to reciprocal links, since they are associated to directed 2-cycles, as illustrated in Figure 4, where the reciprocal link is more emphasized that the two directed triangles of the network.

To avoid the emphasis of reciprocal links, the solution may be to choose $g = k/n < 1/2$ with $k \geq 1$, so that the edge flow on the directed links is effectively rescaled by the positive constant k. Heuristically, the result is that $a(i,j) \to \tilde{a}(i,j) = ka(i,j)$ on a directed link, and $a(i,j) = 0$ on a reciprocal link. Restricting ourselves to $n \leq 5$ and $g < 1/2$, we then recommend to choose $g = 2/5$. The results of this choice of an artificial network is displayed in Figure 5, where the directed triangles are more emphasized than the reciprocal link. The structures found with different g values are summarized in Figure 2.

Another network with a planted structure and includ-



(a) $\ell = 0$. (b) $\ell = 1$. (c) $\ell = 2$.

FIG. 5: Partitions of an artificial network with a reciprocal link for $g = 2/5$, with the normalized magnetic Laplacian.

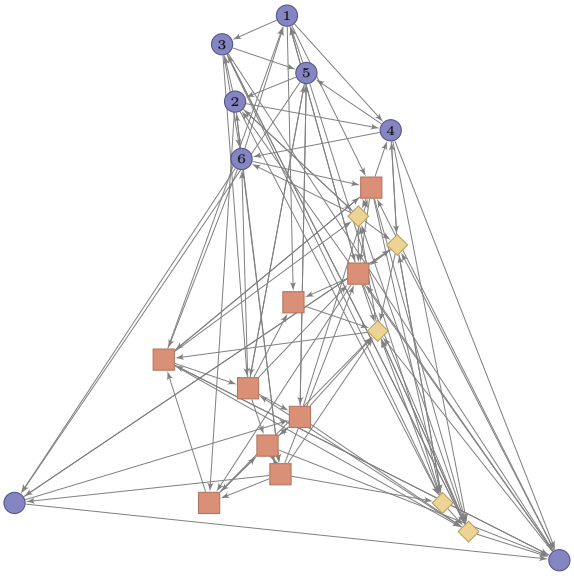


FIG. 6: Example of a directed network with a planted community [40] and reciprocal edges. The numbered nodes (1–6) are part of many directed triangles. The rest of the nodes follow an Erdòs-Rényi pattern. Our method ($\ell = 0$ and $g = 2/5$) detects this planted community and includes two other nodes which are also part of directed cycles. The same "anomalous" community is detected for $\ell = 0$ and $g = 1/3$.

ing reciprocal links was proposed in [40][2]. Our method with $g = 2/5$ is able to uncover the so-called "anomalous" community constituted of a set of 6 nodes connected by many directed triangles while the rest of the network is randomly generated, as illustrated in Figure 6.

V. FLUX COMMUNITIES AT FINITE TEMPERATURE

The methods of the previous section rely on the computation of the spectrum of the magnetic Laplacian. We may expect that the larger is the eigenvalue, the finer is the structure detected. The exploration of the multiscale structure of the network may be performed by considering the same quantum mechanical problem but at finite temperature, the latter parameter being used as a scale parameter.

A. Density operator

Let us recall the basics of quantum mechanics at finite temperature [42]. Consider a directed network and an Hamiltonian operator on the complex valued functions of the nodes. The Hamiltonian $\hat{H}$ defines a mixed state representing a statistical distribution of excitations at a certain temperature $T = 1/\beta$, given by the so-called density operator (or density matrix), i.e.

$$\hat{\rho}_\beta = \frac{e^{-\beta\hat{H}}}{Z(\beta)}, \quad \text{with } Z(\beta) = \mathrm{Tr}\Big(e^{-\beta\hat{H}}\Big). \tag{13}$$

This is well-known to be analogous to quantum mechanics with imaginary time. The statistical mixture is easily understood using the spectral representation of the Hamiltonian in Dirac notation,

$$\hat{\rho}_\beta = \sum_{\ell=0}^{N-1} \frac{e^{-\beta\lambda_\ell}}{Z(\beta)} |\chi_\ell\rangle\langle\chi_\ell|,$$

where the coefficient of each term in the sum is a probability representing the proportion of the eigenvector $|\chi_\ell\rangle$ in the mixed state. For simplicity, the eigenvalues are sorted in ascending order.

B. Correlation at finite temperature

We choose the Hamiltonian to be the positive semi-definite magnetic Laplacian $\hat{H} = \hat{\mathcal{L}}_{a,\mathrm{i}\theta}$. Let us define the matrix elements $(\rho_\beta)_{i,j} = \langle\delta_i|\hat{\rho}_\beta|\delta_j\rangle$. Incidentally, the matrix elements of the density operator are proportional to the Euclidean time Feynman propagator $K(j \to i, \beta) = \langle\delta_i|e^{-\beta\hat{H}}|\delta_j\rangle$, between the node j to i (for a reference, see [42]).

The connection with the approach of the previous section is more clear when writing

$$T^\theta_{i\to j}(\rho_\beta)_{i,j} = \sum_{\ell=0}^{N-1} \frac{e^{-\beta\lambda_\ell}}{Z(\beta)} \left(\chi_{\theta,\ell}(i)\, T^\theta_{i\to j}\chi^*_{\theta,\ell}(j)\right),$$

which is a weighted sum of the correlations appearing in (5) for each eigenspace. Therefore we introduce the positive matrix elements at inverse temperature β,

$$X_{\beta,\theta}(i,j) = \Big|(\rho_\beta)_{i,j}\Big| + \mathrm{Re}\Big(T^\theta_{i\to j}(\rho_\beta)_{i,j}\Big).$$

In the low temperature limit, we obtain the correlation (7) corresponding to the lowest energy state $\ell = 0$, i.e.

$$\lim_{\beta\to\infty} X_{\beta,\theta}(i,j) = X_{\ell=0,\theta}(i,j).$$

[2] Available at `https://github.com/arbenson/tensor-sc`.

As a consequence, $X_{\beta,\theta}$ can be viewed as the weighted similarity matrix of an undirected network. As before, the quality function of a partition is simply chosen to be the modularity

$$P_{\beta,\theta}(C) = \sum_{c\in C}\sum_{i,j\in c}\left(X_{\beta,\theta}(i,j) - p_{\beta,\theta}(i,j)\right),$$

with the null model $p_{\beta,\theta}(i,j)$ chosen to be the configuration model.

VI. FLUX COMMUNITIES IN ARTIFICIAL AND REAL-LIFE NETWORKS

Let us first study the effect of the normalization on an artificial example. The results obtained using the normalized and un-normalized magnetic Laplacians are compared on an example with flux communities of different sizes, illustrated in Figure 7, where the normalized method is able to distinguish all the communities.

Flow communities are often defined as structures retaining the flow of a dynamical process [12]. Here, the dynamics is not given by a Markov process, so our interpretation leads us to name them "flux" communities. A typical example is depicted in Figure 8. This toy directed network has been studied using LinkRank (directed modularity) [16], Markov Stability [12], and Infomap [10], and it is constituted of four flux communities, which are directed cycles. The difficulty to detect them is that the edges interconnecting the groups have a double weight, so that a clustering based on the symmetrized weight matrix will find four different communities of nodes connected by those strong links. In [16], this network is studied in order to describe an example where the directed modularity definition of [15] is unable to discover the communities. These structures are however highlighted by the information theory framework of [10], whereas the Markov Stability framework is able to uncover them using a random walk with teleportation as a means to explore the network.

In the framework proposed in this paper, communities are associated to a certain flux and this flux is related to a certain structure of the network. The variation of the flux is controlled by the quantised coupling constant and allows to uncover communities of different types. Let us use the finite temperature method on the network of Figure 8. Using this method for $g = 1/4$, we uncover two types of communities for the low energy states depicted in Figure 9. Since the weights are not equal in this network, the normalized Magnetic Laplacian (10) is used. First of all, at high temperature, we obtain four other groups of 4 nodes, which are more connected (see Figure 9a). This partition is easily uncovered in the absence of edge directions. Secondly, at a lower temperature, we detect four communities which are the 4-cycles illustrated in Figure 9b. They constitute flux communities, and they are also obtained in the references [10, 12, 16]. The robustness of the finite temperature method under a slight perturbation of the network is studied in Appendix C.

Actually, directed modularity optimization discovers only the partition of Figure 9a. Infomap [10], based on a random walk, is able to uncover the partition of Figure 9b, and the Markov stablility framework [12] finds it as well for a certain Markov time scale.

In order to study the same network level by level, the eigenvalues of the magnetic Laplacian are needed. The spectrum of the magnetic Laplacian provides an indication about the quality of the communities obtained. If the gap between the lowest energy level and the first excited levels is small, the significance of the community found at the first excited level is expected to be high. This is analogous to the well-known interpretation of spectral clustering using the combinatorial Laplacian. Incidentally, we can observe that the spectrum of the combinatorial Laplacian in Figure 10, built using w_{s} as affinity matrix, is qualitatively different from the spectrum of the magnetic Laplacian. Noticeably, in the case of a connected network, there is no information about community structure encoded in the eigenvector of minimal eigenvalue of the combinatorial Laplacian. The same is not true for the magnetic Laplacian. In particular, the smallest eigenvalue of the magnetic Laplacian does not vanish in general and can be degenerate, i.e. the eigenspace can be of dimension greater than one. This degeneracy is intuitively very natural since in quantum mechanics in the continuum the so-called Landau levels are infinitely degenerate.

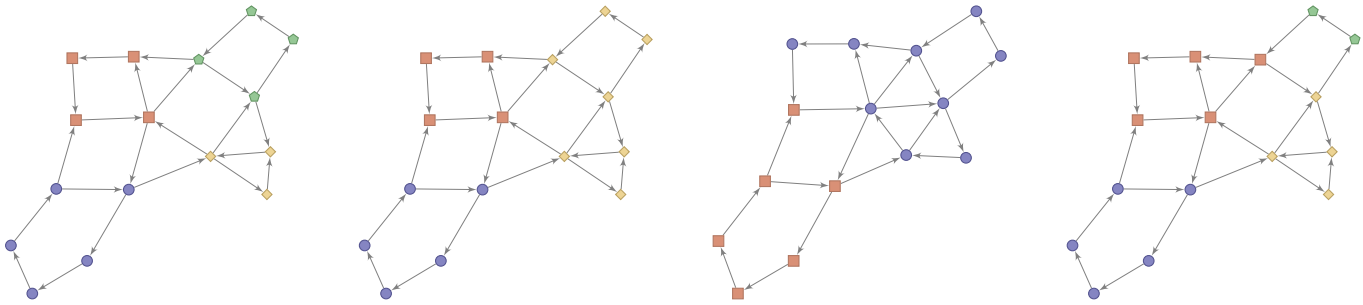
A. Comparison with directed modularity

We can compare the results given by the maximization of directed modularity with our method based on the normalized magnetic Laplacian on the two previous artificial networks. The first example in Figure 9 shows how our method is able to detect communities allowing a magnetic flux (Figure 9b) whereas directed modularity relies more on relative edge density, obtaining the partition of Figure 9a. In the second example of Figure 7, our method discovers all flux communities (Figure 7a), while directed modularity gives a different partition (Figure 7d).

B. A real-file example

To illustrate the result of our method, we study the neuronal network of the C. Elegans nematode [43, 44][3], constituted of 277 neurons. In Figure 11, the communities found for different values of the electric charge g are visualized, using the physical spatial coordinates as positions of the neurons. Qualitatively, the partitions found

[3] Data obtained from `http://www.dynamic-connectome.org/`.



(a) Normalized ($\ell = 0$) (b) Not normalized ($\ell = 0$) (c) Spectral Clustering. (d) Directed modularity.

FIG. 7: Comparison of our methods for $\ell = 0$ and $g = 1/4$. The normalized magnetic Laplacian of (10) may allow to uncover communities in networks with an inhomogeneous degree distribution. In Figure 7c, we show the result of spectral clustering based on the normalized combinatorial Laplacian (Fiedler partition). Directed modularity [15] does not uncover the same communities, as illustrated in Figure 7d.

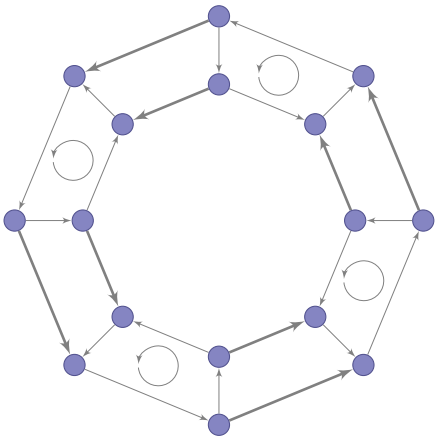
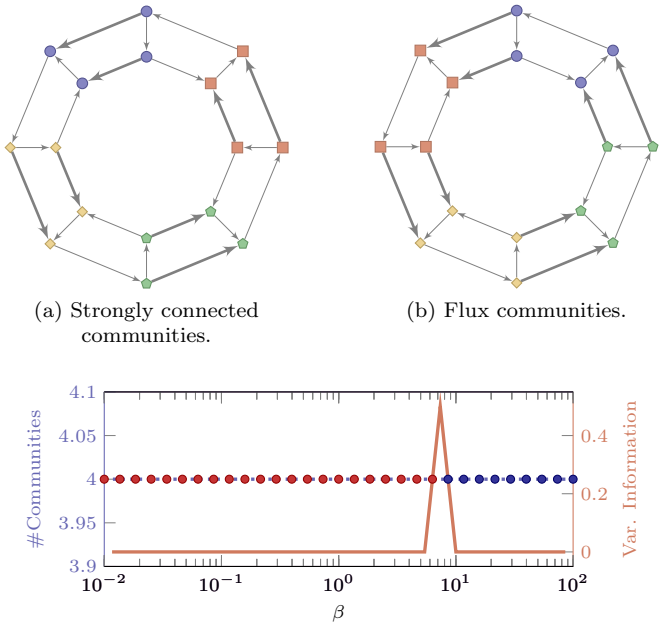


FIG. 8: A directed networks with flux communities. The network is formed by four groups of 4 nodes forming a directed cycle with unit weights. These groups are connected by edges of weight 2, displayed in bold. The four directed cycles with nodes of the same color form four flux communities of 4 nodes.



(a) Strongly connected communities. (b) Flux communities.

(c) Stability of the partitions.

FIG. 9: The communities uncovered by our method for $g = 1/4$ and using the normalized magnetic Laplacian (1). At low β, the 4 strongly connected groups form the stable partition of Figure 9a marked as •, while at large β, i.e. low temperature, the method uncovers 4 flux communities (4-cycles in Figure 9b), which is denoted by •. The variation of information is shown in order to detect the change of partition.

in Figure 11 for the lowest energy level are similar to the partition found by optimizing the directed modularity. However, one may observe that our method seems to group into the same community the neurons appearing in the middle of the figure (i.e. AVM, ALMR, ALML, BDUR, BDUL, PDER, PDEL, PVDR, PVDL, PVM), whereas directed modularity separates them in different communities.

VII. SPECTRAL CLUSTERING IN THE COMPLEX PLANE AND FLOW COMMUNITIES

In the previous section, the importance of directed cycles for detecting dense communities was emphasized. There are however many networks whose structure resides more in the flow of link directions. The magnetic eigenmaps can also reveal such features.

Let us consider a directed network such that its edge flow $a = a_h$,

$$a_h = \mathrm{d}h, \tag{14}$$

is a pure gradient, i.e. an exact 1-form satisfying $a_h(i,j) = h(j) - h(i)$ for any link $\{i,j\}$. Incidentally, in the language of combinatorial Hodge theory [33], $a = a_h$ is a consistent ranking. Based on the gauge covariance

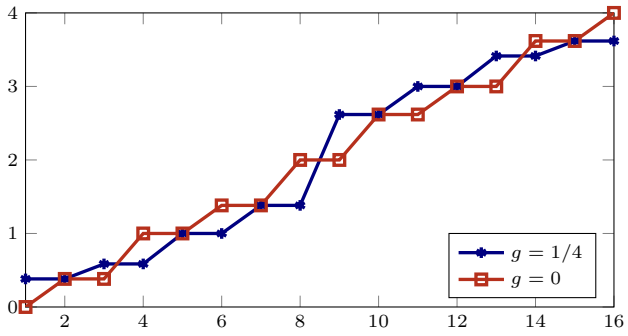


FIG. 10: Spectrum of the magnetic and combinatorial Laplacian for the network of Figure 8. Clearly, in the magnetic Laplacian the gap between the ground state and the first excited state is small, with respect to the gap between the two excited levels.

property (2), it is straightforward to prove that the eigenvalues and eigenvectors of the magnetic Laplacian $\hat{\mathcal{L}}_{\mathrm{d}h,\mathrm{i}\theta}$ are in one to one correspondence with the eigenvalues and eigenvectors of the combinatorial Laplacian, i.e.

$$\hat{\mathcal{L}}_{\mathrm{d}h,\mathrm{i}\theta} = e^{-\mathrm{i}\theta h} \circ \hat{\mathcal{L}}^{\mathrm{C}} \circ e^{\mathrm{i}\theta h}. \tag{15}$$

Actually, the function h, defined on the nodes, is a potential whose gradient gives the edge flow. In particular, in the case of a connected directed network, the lowest energy state of the magnetic Laplacian (15) is simply given by

$$\chi_{\theta,0}(i) = e^{-\mathrm{i}\theta h(i)} \chi_{0,0}(i), \quad \text{with } \chi_{0,0}(i) = \mathrm{cst},$$

where $\chi_{0,0}$ is the constant eigenvector of the combinatorial Laplacian. In this case, we propose to assign two nodes i and j in the same community if

$$\mathrm{phase}(\chi_{\theta,0}(i)) \approx \mathrm{phase}(\chi_{\theta,0}(j)),$$

which corresponds to a spectral clustering in the complex plane of the eigenfunction of lowest energy. Hence, the communities are the groups of nodes with the same potential h.

A. Communities with a running flow

In real-life networks, the edge flow rarely satisfies the exactness condition (14), and therefore, the eigenvectors of $\hat{\mathcal{L}}_{a,\mathrm{i}\theta}$ and $\hat{\mathcal{L}}^{\mathrm{C}}$ are not exactly in one to one correspondence. Nevertheless, it is still instructive to study the lowest energy state of $\hat{\mathcal{L}}_{a,\mathrm{i}\theta}$, i.e. the first magnetic eigenmap.

An example network where the lowest eigenvector of magnetic Laplacian is able to uncover a community structure is given in Figure 12. It is composed of three groups of ten nodes. Within each group there is a uniform probability 0.5 that two nodes are connected. There is also a probability 0.5 that a node connects with a node from another group. However, 90 percent of the connections between the groups point in the direction of the flow. A similar network was actually proposed in [45]. The communities can be detected thanks to mixture models [46], whereas directed modularity is expected to fail. Actually, this type of community structure is associated to the role played by each node in the network, and hence this feature is in the same spirit than the so-called Role Based Similarity [13, 47].

The phase of the lowest eigenvector is depicted in Figure 13. Indeed, the phase of $\chi_{\theta,0}$ is almost piecewise constant.

A similar situation happens for the word adjacency directed network of Figure 14, which was constructed from an English text [48][4]. The network is built by collecting adjacent nouns and adjectives in the novel *David Copperfield*. Hence, a directed link points from a word to another adjacent word, if the first one appears before the second. From the structure of English, we can expect a certain flow structure in the network. Indeed, the phase of the first eigenmaps separates the nouns from the adjectives.

The difficulty of discovering these type of communities using the phase of the first eigenvector of the magnetic Laplacian is that the number of communities has to be guessed. In order to circumvent this problem, we propose the definition of a quality function. Actually, the correlations

$$\xi^{(a)}_{\theta,0}(i,j) = \mathrm{Re}\big(\chi_{\theta,0}(i)\,\chi^{*}_{\theta,0}(j)\big),$$

for any nodes i and j, incorporate the information necessary to find the flow communities in the network. Hence, we define the positive matrix elements

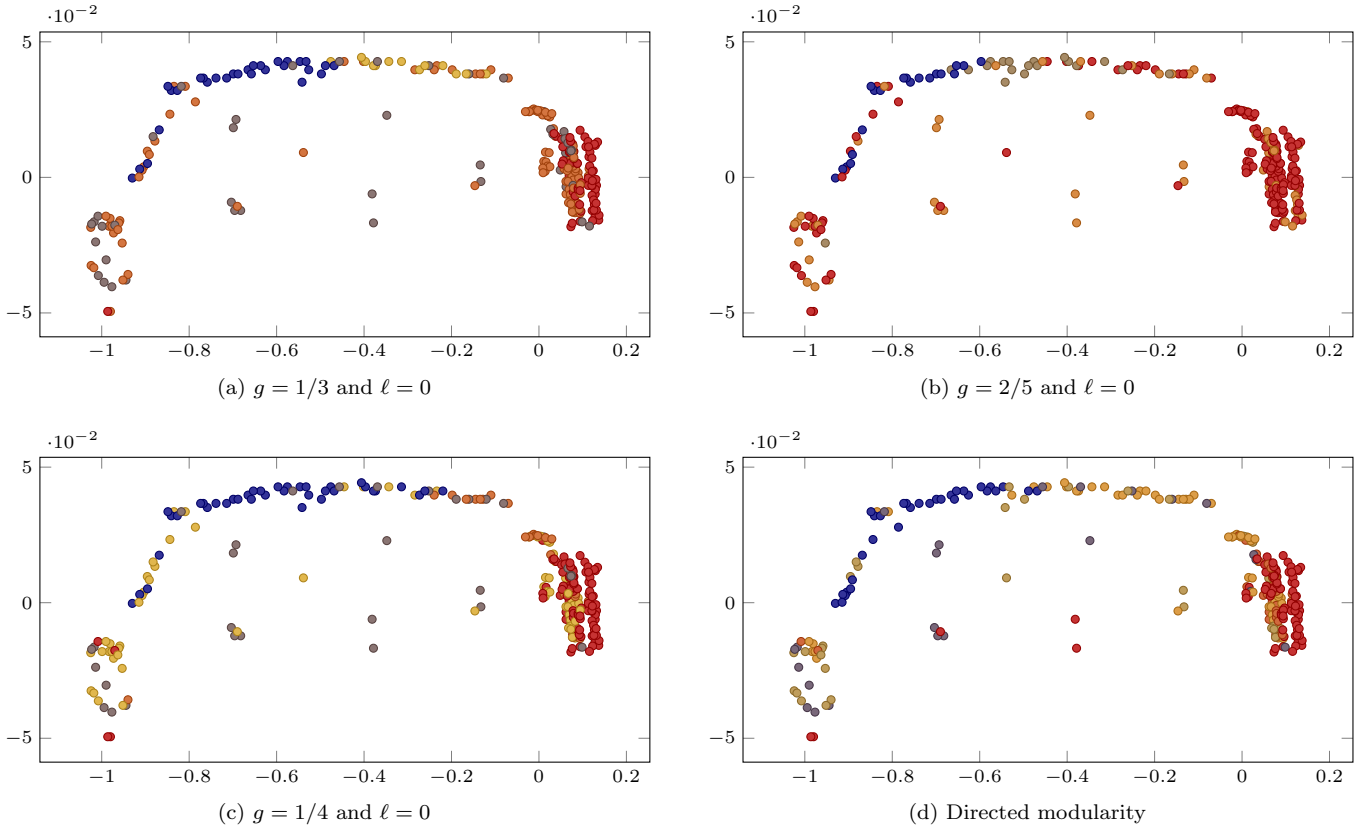
$$\Xi^{(a)}_{\theta,0}(i,j) = \left|\chi_{\theta,0}(i)\,\chi^{*}_{\theta,0}(j)\right| + \xi^{(a)}_{\theta,0}(i,j), \tag{16}$$

which define another weighted similarity matrix. The matrix $\Xi^{(a)}_{\theta,0}$ corresponds to another network containing the same nodes but with new weighted undirected links. The correlation $\Xi^{(a)}_{\theta,0}$ differs from $X^{(a)}_{\theta,0}$ only by the parallel transport factor. Noticeably, this matrix is not invariant under a gauge transformation $a \to a + \mathrm{d}h$.

B. Optimization of a quality function

A modularity optimization procedure on $\Xi^{(a)}_{\theta,0}$ allows to uncover the three flow communities of Figure 12, depicted with different colours. For simplicity, we used the Newman-Girvan modularity [2] with the configuration model, although actually we could have chosen another method in order to partition the undirected network associated to $\Xi^{(a)}_{\theta,0}$.

[4] Data obtained from `http://www-personal.umich.edu/~mejn/netdata/`.



(a) $g = 1/3$ and $\ell = 0$

(b) $g = 2/5$ and $\ell = 0$

(c) $g = 1/4$ and $\ell = 0$

(d) Directed modularity

FIG. 11: Communities in the C. Elegans directed network found by optimizing (11). The nodes are displayed according to the physical positions of the neurons, units are cm. In each figure, the nodes in the same community share the same color.

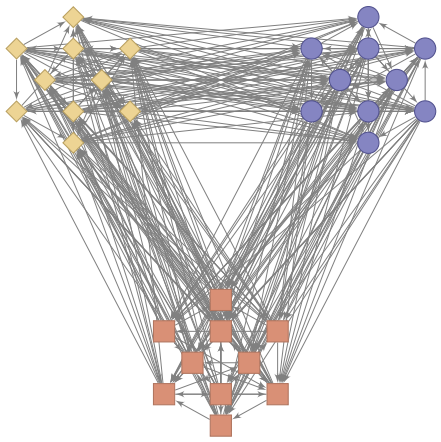
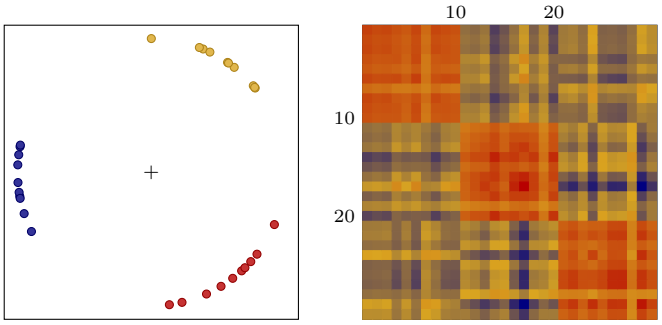


FIG. 12: Directed network with a flow running between three sets of nodes. The nodes within each group are not strongly connected.



(a) Phase of $\chi_{\theta,0}$

(b) $\Xi^{(a)}_{\theta,0}$

FIG. 13: Phase of the lowest eigenvector $\chi_{\theta,0}$ and the matrix $\Xi^{(a)}_{\theta,0}$ for $g = 1/4$, as a function of the vertex number for the network of Figure 12.

To compare this method with directed modularity on another network, we examine now a real-life example used in [15] representing the Big Ten football network. In the directed network of Figure 15, our method based on the optimization the modularity of (16) gives three communities. However, directed modularity gives only two communities. In our method, there are two communities of teams shared with the directed modularity approach, however the team of Minnesota (number 4) is singled out because its number of successes equals its number of defeats. Therefore, the partition found by our method seems to be consistent with the intuition.

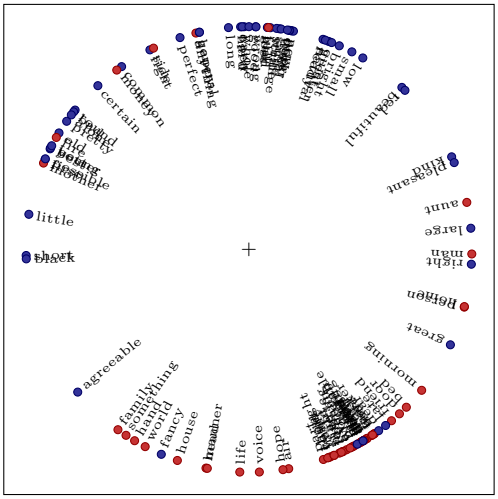
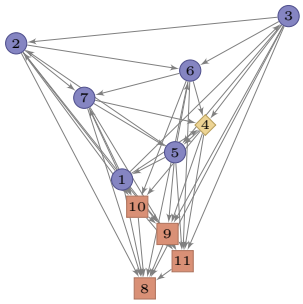
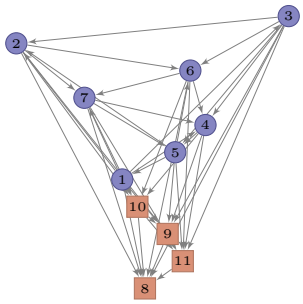


FIG. 14: The phase of the first magnetic eigenmaps for the word adjacency network and $g = 2/5$, clearly separating the nouns (denoted as •) from the adjectives (denoted as •), represented on a circle.



(a) $g = 1/4$ and $\ell = 0$.



(b) Directed modularity.

Team	Position	Team	Position
Penn State	1	Wisconsin	7
NorthWestern	2	Illinois	8
Ohio State	3	MichiganState	9
Minnesota	4	Purdue	10
Michigan	5	Indiana	11
Iowa	6		

(c) List of the teams and their associated position.

FIG. 15: Comparison of our method based on the modularity optimization of (16) for flow communities of the unnormalized magnetic Laplacian with directed modularity in the Big Ten football network, and the ranking of the teams.

VIII. CONCLUSIONS

Link directions in complex networks may contain relevant information. Accounting for this information may be done in various manners, for example several dynamical processes may be imagined to explore the geometry of the networks. In the past, much effort was devoted to the study of the geometrical structure of networks in terms of density "clusters". On the contrary, we have been interested here in other local structures which are also related to the topology of the networks, where the word "topology" is understood in the mathematical sense.

In particular, the use of the complex valued magnetic Laplacian for the problem of community detection in directed network was studied in this paper. The method was strongly inspired by quantum physics, and it generalizes known results to the complex domain. A striking feature of the magnetic Laplacian is that it is related to the topology of the network. Indeed, there is a strong relationship between discrete Hodge theory and the results presented in this paper. As we have illustrated with several experiments, this approach allows to unveil communities on directed graphs based either on cycles (flux communities) or in the role of the different nodes.

It is expected that different deformations of the combinatorial Laplacian may be relevant to answer other questions of interest for the study of complex networks.

Acknowledgments

The authors would like to thank the following organizations. • EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A–DATADRIVE-B (290923). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. • Research Council KUL: CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants. • Flemish Government: – FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grant. – iMinds Medical Information Technologies SBO 2015. – IWT: POM II SBO 100031. • Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017).

[1] M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the American Mathematical Society*, (56):1082–1097 & 1164–1166, 2009.

[2] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(026113), 2004.

[3] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103:8577–8582, 2006.

[4] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and

J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, (328):876–878, 2010.

[5] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, 2006.

[6] V. A. Traag, P. Van Dooren, and Y. Nesterov. Narrow scope for resolution-limit-free community detection. *Phys. Rev. E*, 84:016114, 2011.

[7] F. Chung. *Spectral Graph Theory*. Am. Math. Soc., 1997.

[8] U. von Luxburg. A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[9] J.C. Delvenne, R. Lambiotte, and L.E.C. Rocha. Diffusion on networked systems is a question of time or structure. *Nat. Commun.*, (1038), 2015.

[10] M. Rosvall and C. T. Bergstrom. Maps of random wals on complex networks reveal community structure. *Proc. Matl. Acad. Sci. USA*, 105(4):1118–1123, 2008.

[11] J.C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proc. Natl. Acad. Sci. USA*, 107:12755–12760, 2008.

[12] R. Lambiotte, J.C. Delvenne, and M. Barahona. Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. *IEEE Transactions on Network Science and Engineering*, 1:76–90, 2014.

[13] M. Beguerisse-Daz, G. Garduo-Hernández, B. Vangelov, S.N. Yaliraki, and M. Barahona. Interest communities and flow roles in directed networks: the Twitter network of the UK riots. *J. R. Soc. Interface*, 11(101):20140940, 2014.

[14] A. Arenas, J. Duch, A. Fernandez, and S. Gomez. Size reduction of complex networks preserving modularity. *New J. Phys.*, 9(176), 2007.

[15] E. A. Leicht and M. E. J. Newman. Community structure in directed networks. *Phys. Rev. Lett.*, 100(118703), 2008.

[16] Y. Kim, S.-W. Son, and H. Jeong. LinkRank: Finding communities in directed networks. *Phys. Rev. E*, 81(016103), 2010.

[17] F. D. Malliaros and M. Vazirgiannis. Clustering and community detection in directed networks: a survey. *Phys. Rep.*, 533:95–142, 2013.

[18] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486:75–174, 2010.

[19] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.*, 5, 2014.

[20] G. Berkolaiko. Nodal count of graph eigenfunctions via magnetic perturbations. *Analysis and PDE*, 6(5):1213–1233, 2013.

[21] Y. Colin de Verdière. Magnetic interpretation of the nodal defect on graphs. *Analysis and PDE*, 6(5):1235–1242, 2013.

[22] M. A. Shubin. Discrete magnetic Laplacian. *Comm. Math. Phys.*, 164:259–275, 1994.

[23] Y. Aharonov and D. Bohm. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.*, 115:485–491, 1959.

[24] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[25] C. Lange, S. Liu, N. Peyerimhoff, and O. Post. Frustration index and cheeger inequalities for discrete and continuous magnetic laplacians. *Calculus of Variations and Partial Differential Equations*, 54(4):4165–4196, 2015.

[26] Y. Li and Z.-L. Zhang. Digraph Laplacian and the Degree of Asymmetry. *Int. Math.*, 8(4):381–401, 2012.

[27] R. Kenyon. Spanning forests and the vector bundle Laplacian. *The Annals of Probability*, 39(5):1983–2017, 2011.

[28] R. Forman. Determinants of Laplacians on graphs. *Topology*, 32(1):35–46, 1993.

[29] A. Singer and H.T. Wu. Vector Diffusion Maps and the Connection Laplacian. *Commun Pure Appl Math.*, 65(8):1067–1144, 2012.

[30] F. C. Graham and W. Zhao. Ranking and sparsifying a connection graph. In *WAW*, pages 66–77, 2012.

[31] M. Cucuringu. Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and sdp synchronization. *IEEE Transactions on Network Science and Engineering*, 3(1):58–79, Jan 2016.

[32] F. R. K. Chung and S. Sternberg. Laplacian and vibrational spectra for homogeneous graphs. *Journal of Graph Theory*, 16(6):605–627, 1992.

[33] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial Hodge theory. *Math. Program.*, 127(1):203–244, 2011.

[34] M. Fanuel and J. Govaerts. Dressed fermions, modular transformations and bosonization in the compactified Schwinger model. *Journal of Physics A: Mathematical and Theoretical*, 45(3):035401, 2012.

[35] P. Blanchard and D. Volchenkov. *Random Walks and Diffusions on Graphs and Databases: An Introduction*. Springer, 2011.

[36] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[37] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. *Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization*, chapter 48, pages 559–570.

[38] B. Serrour, A. Arenas, and S. Gómez. Detecting communities of triangles in complex networks using spectral optimization. *Comput. Commun.*, 34(5):629–634, 2011.

[39] A. Prat-Pérez, D. Dominguez-Sal, J. M. Brunat, and J.-L. Larriba-Pey. Shaping communities out of triangles. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1677–1681, 2012.

[40] A. R. Benson, D. F. Gleich, and J. Leskovec. Tensor spectral clustering for partitioning higher-order network structures. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 118–126, 2015.

[41] C. Klymko, D. F. Gleich, and T. G. Kolda. Using triangles to improve community detection in directed networks. In *Proceedings of the ASE BigData Conference*, Stanford, CA, 2014.

[42] A. Wipf. *Statistical Approach to Quantum Field Theory*. Number 864 in Lecture Notes in Physics. Springer, 2013.

[43] S. Varier and M. Kaiser. Neural development features: Spatio-temporal development of the caenorhabditis elegans neuronal network. *PLoS Comput Biol*, 7(1):1–9, 01 2011.

[44] M. Kaiser and C. C. Hilgetag. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput Biol*, 2(7):1–11, 07 2006.

[45] A. Lancichinetti and S. Fortunato. Benchmarks for test-

ing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E*, 80:016118, 2009.
[46] E. A. Leicht and M. E. J. Newman. Mixture models and exploratory analysis in networks. *Proc. Natl. Acad. Sci. USA*, 104:956–9569, 2007.
[47] K. Cooper and M. Barahona. Role-based similarity in directed networks, 2010.
[48] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, 2006.
[49] A. H. Al-Mohy and N. J. Higham. A new scaling and squaring algorithm for the matrix exponential. *SIAM Journal on Matrix Analysis and Applications*, 31(3):970–989, 2010.

Appendix A: Relation with previous work

1. Discrete Vector Bundle Laplacian

From a gauge theory perspective, the each of the factors $\exp\left(\mathrm{i}\theta a\left(i,j\right)/2\right)$ may be understood intuitively as a unitary "parallel transport" along the edge from i to the mid-point between i and j. A parallel transport is an isomorphism between the fibres of a vector bundle. Hence, the magnetic Laplacian (1) has a covariance property [21] under $a \to a + \mathrm{d}h$. Let us outline this idea.

In the reference [27], Kenyon defines a vector bundle V_G on a graph as the choice of a vector space V_v, called fibre, for each vertex $v \in V$. Here, for simplicity we choose V_v to be isomorphic to $\mathbb{C}$. A section of the vertex bundle is in fact given by one complex number for each vertex, hence it is a 0-form in Ω_0.

Furthermore, Kenyon extends the construction to the edge space, so that there is a fibre isomorphic to $\mathbb{C}$, for each edge. If the edges are oriented, a 1-form is a section of the edge bundle, i.e. a skew symmetric function of Ω_1 in our notations. Then, it is possible to define a connection isomorphism $\phi_{ve} = \phi_{ev}^{-1}$ for each edge $e = [v, v']$ and vertex at the boundary of the edge. In our case, we choose

$$\phi_{je} = e^{\mathrm{i}\theta a(i,j)/2}, \quad \text{for } e = [i,j]\,.$$

Still, following [27], we introduce the map $\mathrm{d} : \Omega_0 \to \Omega_1$

$$\left(\mathrm{d}\psi\right)(e) = \phi_{je}\psi\left(j\right) - \phi_{je}\psi\left(i\right).$$

Subsequently, the operator $\mathrm{d}^\star : \Omega_1 \to \Omega_0$ is introduced by

$$\left(\mathrm{d}^\star\omega\right)(i) = \sum_{j|e=\{i,j\}} w_{\mathrm{s}}\left(e\right)\phi_{ej}\omega(e).$$

Hence, $\mathrm{d}^\star$ is the adjoint of d only if $\phi_{ve} = \phi_{ev}^* = \phi_{ev}^{-1}$. However, because we wish a self-adjoint Laplacian, the construction of the Laplacian of [27] as $\mathrm{d}^\star\mathrm{d}$ is the one of this paper only if $\phi_{ev}^* = \phi_{ev}^{-1}$, leading to the magnetic Laplacian of (1) and studied in [20–22].

2. Connection Laplacian

Let us outline the relation between the discrete connection Laplacian [29] and the magnetic Laplacian. Since U(1) and SO(2) are isomorphic, another definition of the magnetic Laplacian can be

$$\left(\hat{\mathcal{L}}_{a,\theta}\boldsymbol{\psi}\right)(i) = \sum_j w_{\mathrm{s}}\left(i,j\right)\left(\boldsymbol{\psi}\left(i\right) - \boldsymbol{\rho}_{j\to i}^{\theta}\boldsymbol{\psi}\left(j\right)\right),$$

with the matrix of the lowest dimensional representation of SO(2),

$$\begin{aligned}\boldsymbol{\rho}_{j\to i}^{\theta} &= e^{\mathrm{i}\theta a(j,i)\mathfrak{m}} \\ &= \begin{pmatrix}\cos\theta a(j,i) & -\sin\theta a(j,i)\\ \sin\theta a(j,i) & \cos\theta a(j,i)\end{pmatrix}\end{aligned}$$

with the skew-symmetric matrix

$$\mathfrak{m} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},$$

while $a\left(i,j\right)$ takes the values 0, -1 or $+1$.

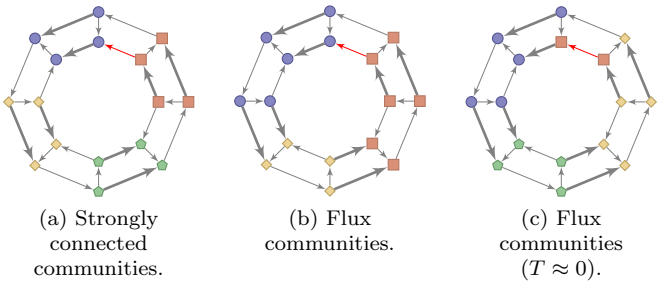
A major difference with respect to the reference [29] is that, here, we choose the representation of U(1) in order to detect specific structures in directed network, whereas Singer and Wu simply build from a dataset an undirected network with a matrix of $\mathrm{O}(d)$ on each link. The question of choosing a representation of $\mathrm{O}(d)$ is therefore not relevant.

Appendix B: Computational aspects

The methods presented here include two steps. The methods focusing on individual eigenvectors requires the computation of the smallest eigenvalue. Because we have noticed that the normalized magnetic Laplacian $\hat{\mathcal{L}}_{a,\mathrm{i}\theta}^{\mathrm{N}}$ was empirically more successful, we can merely compute the maximal eigenvalues of the following operator $\hat{\mathcal{S}}_{a,\mathrm{i}\theta} = \mathbb{I} - \hat{\mathcal{L}}_{a,\mathrm{i}\theta}^{\mathrm{N}}$, for instance thanks to the power method. The matrix exponential (13) can be computed thanks to a Padé approximant method, see for instance the scaling and squaring method [49]. For the maximization of the quality functions, we rely on the generalized Louvain method [4, 36].

Appendix C: Robustness of the finite temperature method

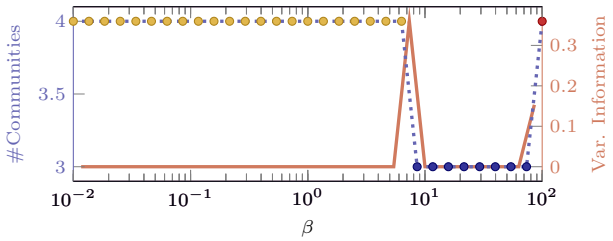
In order to test the robustness of the partitions found in the network of Figure 8, we have simply slightly modified the direction of one link of weight 1 in one of the 4-cycles of this network, which is marked in red in Figure 16. Hence, one of the four 4-cycles is broken. As a result, we find at high temperature the same partition as



(a) Strongly connected communities.

(b) Flux communities.

(c) Flux communities ($T \approx 0$).



(d) Stability of the partitions.

FIG. 16: The communities uncovered by our method for $g = 1/4$ and using the normalized magnetic Laplacian (1) (i.e. ● correspond to the partition of Figure 16a, ● to Figure 16b, and ● to the partition of Figure 16c). The red directed link in the network is the flipped link compared to Figure 8.

in the case of Figure 8, which is due to the presence of the links of weight 2. In that case, the cycles are broken. At lower temperature, the stable partition is given by groups containing the three unbroken 4-cycles, showing that the method reacts well when one 4-cycle is perturbed. Finally, at a very small temperature, where the importance of the flux is the largest, another partition preserving the cycles is obtained, with a fourth community found at the perturbed link.