Community Detection with the z-Laplacian

Jess Banks
Cristopher Moore
Mark Newman
Pan Zhang

SFI WORKING PAPER: 2014-09-035

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author’s copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
Community Detection with the z-Laplacian

Jess Banks, Cristopher Moore, Mark Newman & Pan Zhang

September 10, 2014

Abstract

Community detection is a fundamental problem in network science, with broad applications across the biological and social arenas. A common approach is to leverage the spectral properties of an operator related to the network (most commonly the adjacency matrix or graph Laplacian), though there are regimes where these techniques are known to fail on sparse networks despite the existence of theoretically detectable community structure [3],[4]. This work introduces an operator we term the "z-Laplacian" $L_z = zA - D$, which has been observed to share important spectral properties with the non-backtracking matrix of [3],[6] and which we believe can find communities even in the sparse case. We augment tools from the theory of random matrices with message-passing and population dynamics approaches in order to study the spectrum of $L_z$.

1 Spectral Clustering

A celebrated method for finding network community structure is to classify vertices according to the eigenvectors of a linear operator associated with the graph such as the adjacency matrix or Laplacian. However, it has been observed that these spectral methods are in general ineffective when the network is very sparse [3],[4]. Krzakala et al. propose a spectral algorithm in [3] based on the non-backtracking matrix $B$, which represents a walk on the network where backtracking along an edge is prohibited; they argue that this method successfully extracts community structure even when the network is sparse. In this work, we explore the spectral properties of yet another linear operator—the z-Laplacian—which shares spectral properties with $B$ and which we believe is also suitable for detecting communities on sparse networks. A similar operator was used in [6] to study the spectral properties of $B$.

Random Graphs, Random Matrices

It is common practice to study the behavior of spectral methods when applied to various random graph ensembles; we focus on two in particular. In the Erdős-Rényi model, denoted $G(n, p)$, we choose the number of vertices $n$ as well as the average degree $c$, and include each of the possible edges with probability $p = \frac{c}{n}$ [2]. We can think of this process as inducing a probability distribution on the space of all possible graphs on $n$ vertices. Often we are concerned with the behavior of these graphs in the limit as $n$ tends to infinity, and distinguish the dense case, where $c = O(n)$, from the sparse where $c = O(1)$.

The adjacency matrix $A$ of a $G(n, p)$ graph is a random matrix where each diagonal entry is zero and each off-diagonal entry is Bernoulli-distributed with parameter $p$. This characterization allows us to bring to bear the tools of random matrix theory in our study of the spectra of $G(n, p)$ adjacency matrices. A foundational result concerns the empirical spectral density $A$, which we write

$$\rho(\zeta) = \frac{1}{n} \sum_{j=1}^{n} \delta(\zeta - \lambda_j),$$
summing over the eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_n \) of \( A \); we can apply a theorem of Wigner to show that \( \rho(\zeta) \) for the ensemble \( G(n, p) \) converges in expectation to the well-known Semicircle Law

\[
\rho_{SC}(\zeta) = \frac{1}{2\pi c} \sqrt{4c - \zeta^2}
\]

in the large-\( n \) limit [7].

Although any graph on \( n \) vertices is a possible outcome of the Erdős-Rényi model, graphs with community structure occur with vanishingly small probability. A random graph model for which the typical graph has community structure is the Stochastic Block Model, where we first fix the number of nodes \( n \), number of groups \( k \), and a vector of a priori group assignments \( q \in \{1, \ldots, k\}^n \), and include each edge with probability dependent on the group assignments of the two nodes it joins. We encode these probabilities in a matrix \( P \in [0, 1]^{k \times k} \), where \( P_{ij} \) gives the probability of an edge joining a group \( u \) and group \( j \) vertex. Again we are interested in the large-\( n \) limit, and distinguish the sparse from dense cases as follows. Let \( P = \frac{1}{n} C \), where \( C_{ij} \) gives the expected number of neighbors in group \( j \) that a group \( i \) vertex will have; taking \( C_{ij} = O(n) \) results in dense graphs and \( C_{ij} = O(1) \) sparse.

Again we can think of the adjacency matrix of an SBM graph as a random matrix with Bernoulli-distributed variables, though in this case the group assignments define a block structure and the Bernoulli parameter in each block is determined by the matrix \( P \). However, the generality of the Wigner theorem allows us to once again prove that the empirical spectral density converges to (1) [4].

The Detection Threshold

Because of the randomness inherent in the SBM, there is a parameter regime in which edges are empirically more likely to occur between vertices of the same group, but these groups are undetectable to any algorithm. We follow [3],[4] in considering the case where there are only two groups of equal size,

\[
C_{ij} = \begin{cases} 
    c_{in} & i = j \\
    c_{out} & i \neq j
\end{cases}
\]

and we call \( c \) the average degree of the network; it has been shown that unless

\[
c_{in} - c_{out} > 2\sqrt{c}, \tag{2}
\]

the community structure is impossible to detect.

For graphs that are sufficiently dense, a closer look at the spectrum of the adjacency matrix gives some insight into this threshold. Asymptotic convergence to (1) means that most of the spectrum—the ‘bulk’—is confined to the interval \([-2\sqrt{c}, 2\sqrt{c}]\) even for finite \( n \). However, the largest eigenvalue \( \lambda_1 \) always lies above \( 2\sqrt{c} \); its associated eigenvector encodes vertex centrality in the graph. If the graph is sufficiently dense, there is a second eigenvalue \( \lambda_c \) which is separated from the bulk; its associated eigenvector sorts vertices by community membership. This eigenvalue is

\[
\lambda_c = \frac{c_{in} - c_{out}}{2} + \frac{c_{in} + c_{out}}{c_{in} - c_{out}}
\]

we recover (2) by looking for the condition on \( c_{in} \) and \( c_{out} \) that separates \( \lambda_c \) from the bulk [3],[4].

Sadly, we are not so fortunate in the sparse case, where, as Figure 1 shows, the spectrum deviates significantly from the Wigner Semicircle and upper edge of the bulk overtakes \( \lambda_c \), making it impossible to locate among the bevy of meaningless eigenvectors even when (2) is satisfied. Krzakala et al. demonstrate in
Figure 1: (a) Spectrum of a dense SBM graph, \( n = 4000, c_{in} = c_{out} = \). Most eigenvalues are confined to the interval \([-2\sqrt{c}, 2\sqrt{c}]\), and the community eigenvalue is well-separated from the bulk. (b) Spectrum of a sparse SBM graph, \( n = 4000, c_{in} = c_{out} = \). The bulk has overrun the community eigenvalue despite theoretically detectible community structure. (c) Spectrum of \( L_z \) for the same sparse SBM graph. The bulk has been shifted to the left, leaving the community eigenvalue \( \mu^2 - 1 \) exposed.

[3] that spectral methods using the ‘non-backtracking perator’ \( B \) succeed in detecting community structure whenever (2) holds; the \( z \)-Laplacian operator that we propose in the sequel is inspired by [3] and we conjecture that it as well succeeds down to the detection threshold. Figure 1 shows that it does so by skewing the bulk toward \( -\infty \) and exposing a community-correlated eigenvalue.

**The \( z \)-Laplacian**

We motivate the \( z \)-Laplacian as follows. It has been noted ([3],[6],[5]) that every eigenvalue \( \lambda \) of the nonbacktracking matrix \( B \) different from \( \pm 1 \) satisfies

\[
\det \left[ \lambda^2 I - \lambda A + (D - 1) \right] = 0
\]

and thus we define the \( z \)-Laplacian to be, for any \( z \),

\[
L_z = zA - D.
\]

Krzakala et al. show that a particular vector \( f \) correlated with the community structure approximately satisfies the quadratic eigenvalue equation

\[
Af = \mu f + \mu^{-1}(D - 1)f,
\]

where \( \mu = \frac{c_{in} - c_{out}}{2} \) and \( D \) is the diagonal degree matrix [3]. Applying some algebra, we see that this vector is thus an eigenvalue for \( L_z \) with eigenvalue \( \mu^1 - 1 \); Figure 1 shows that this eigenvalue is not buried in the bulk for \( L_z \). Our task is then to understand the spectrum of \( L_z \), and specifically to find a regime of \( z \) where \( \mu^2 - 1 \) is well-separated from the upper edge of the bulk.

**2 The Spectral Density as a Generating Function**

In order to study the spectrum of \( L_z \), we will for the moment focus on graphs generated by \( G(n, c/n) \) where \( c = O(1) \). We first record a fundamental result from the theory of random matrices relating the empirical
spectral density of a matrix $M$ to the traces of its powers. Recall that the empirical spectral density of $M$ is

$$
\rho(\zeta) = \frac{1}{n} \sum_{j=1}^{n} \delta(\zeta - \lambda_j).
$$

Expressing $\delta(\zeta)$ as a limit of Cauchy distributions and exploiting some complex arithmetic, we see that

$$
\delta(\zeta) = \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{\zeta^2 + \epsilon^2} = \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{\zeta - i\epsilon},
$$

meaning that we can rewrite the spectral density as

$$
\rho(\zeta) = \frac{1}{n\pi} \lim_{\epsilon \to 0} \Im \sum_{j=1}^{n} \frac{1}{\zeta - \lambda_j - i\epsilon}
$$

$$
= \frac{1}{n\pi} \lim_{\epsilon \to 0} \Im \sum_{j=1}^{n} \sum_{t=0}^{\infty} \left( \frac{\lambda_j + i\epsilon}{\zeta} \right)^t
$$

$$
= \frac{1}{n\pi} \lim_{\epsilon \to 0} \Im \sum_{t=0}^{\infty} \text{tr}(M + i\epsilon 1)^t \frac{\zeta^t}{t+1}.
$$

Taking the limit as $\epsilon \to 0$ from above, we can finally write

$$
\rho(\zeta) = \frac{1}{n\pi} \Im \sum_{t=0}^{\infty} g(\zeta), \quad (4)
$$

where

$$
g(\zeta) = \sum_{t=0}^{\infty} \frac{\text{tr} M^t}{\zeta^{t+1}}. \quad (5)
$$

Interpreting $M$ as a (weighted) adjacency matrix, and recalling the fact that $\text{tr} M^t$ counts the number of (weighted) loops of length $t$ on the graph which $M$ encodes, we can think of $g(\zeta)$ as the generating function which counts loops.

### 3 Message-Passing

With equations (4) and (5) in hand, we can study the spectrum of $L_z$ by interpreting $L_z$ as a new graph $G' = (V', E')$ and computing the generating function $g(\zeta)$ for loops on $G'$. Referring to the definition $L_z = zA - D$, it’s clear that we can obtain $G'$ from $G$ by adding a self-edge at every vertex $v_i$ with weight $d_i$ and giving each of the original edges a weight $z$. Moreover, we can leverage the fact that sparse Erdős-Rényi and SBM graphs converge to trees [cite] in the large-$n$ limit in order to compute $g(\zeta)$ via a message-passing approach.

Let $g^{i\leftarrow j}(\zeta)$ be the generating function for loops starting and ending at $v_i$ and not using the edge $(i, j)$. Assuming that the graph is locally treelike, we can think of such loops as being made up of two components. The first are self-edges at $v_i$ with weight $-d_i$, the second—for each neighbor $v_k$ of $v_j$ other than $v_i$—are loops $(v_k, v_j, v_k, v_j, ..., v_k, v_k)$ such that the edge $(v_k, v_j)$ does not appear in the loop $v_k \rightarrow v_k$; these have weight $z^2 g^{i\rightarrow k}(\zeta)$. Every loop counted by $g^{i\leftarrow j}$ is a concatenation of these two component loops of some arbitrary
length, so we can write $g^{i \leftarrow j}$ as

$$g^{i \leftarrow j}(\zeta) = \sum_{r=0}^{\infty} \frac{1}{\zeta + d_j - z^2 \sum_{k \in \partial(j) \setminus i} g^{j \leftarrow k}(\zeta)}.$$

Using this recursive equation, we can compute $g^{i \leftarrow j}$ for every pair of vertices $v_i$ and $v_j$ by treating $g^{i \leftarrow j}$ as a message sent from $v_j$ to $v_i$; we initialize the population of messages with random complex numbers, update the messages that each vertex sends according to the messages it receives from its neighbors, and iterate this process until the message population converges. At this point we can compute the generating function for all loops starting and ending at vertex $v_j$ by

$$g^j(\zeta) = \sum_{r=0}^{\infty} \frac{1}{\zeta + d_j - z^2 \sum_{k \in \partial(j) \setminus i} g^{j \leftarrow k}(\zeta)}.$$

making

$$\rho(\zeta) = \frac{1}{n\pi} \Im g(\zeta) = \frac{1}{n\pi} \Im \sum_{j=1}^{n} g^j(\zeta). \quad (7)$$

However, (7) means that we can locate regions where $L_z$ has zero spectral density without computing $g(\zeta)$ at all, simply by looking for values of $\zeta$ at which all of the messages $g^{i \leftarrow j}(\zeta)$ are real. Of particular salience to the community-detection properties of $L_z$ is that the upper edge of the bulk occurs at the smallest value of $\zeta$ for which the message population is real. Even for $c = O(1)$, graphs generated by $G(n, c/n)$ are in general not trees for $n < \infty$; in this case the message-passing approach allows us to make an estimate of the edge of the bulk which in the infinite limit appears to be exact (Figure 3).
Figure 3: Difference for $L_z$ between the actual edge of the bulk ($\lambda_2$) and the numerical message-passing approximation given by $\min_\zeta \left\{ \Im g^{\text{act}}(\zeta) = 0 \right\}$, for $G(n, c/n)$ with $c = 3$, $z = 2.366$, and $n$ ranging from 200 to 50,000. Differences for each $n$ are averaged over 20 graph realizations.

4 Population Dynamics

In order to compute the average spectral density for an entire random graph ensemble—i.e. $G(n, p)$ for fixed $n, p$—we can alter this message-passing machinery as follows. We begin with an initial population of messages $\{g_j(\zeta)\}$ and at every step we pick $D$ from the degree distribution of random graph ensemble, select $D - 1$ messages $g_{j_1}(\zeta), \ldots, g_{j_{D-1}}(\zeta)$ from the population, compute

$$\hat{g} = \left( \zeta + D - z^2 D \sum_{k=1}^{D-1} g_{j_k} \right)^{-1}, \quad (8)$$

and use $\hat{g}$ to replace a random message from the population. This process is repeated until the distribution of messages reaches a fixed point. Although it is possible to compute the generating function $g(\zeta)$ from this limiting message distribution, once again we can locate regions of zero spectral density by finding values of $\zeta$ for which the limiting distribution is real-valued.

The $\delta$-Function Message Distribution

While the distributional fixed point of (8) is in general quite complicated, there is a critical value of $\zeta$ at which the limiting message distribution is a $\delta$-function where all the messages have the same value. If $g_j = \frac{1}{z^2}$, the update equation (8) becomes

$$\hat{g} = \left( \zeta + D - z^2 \sum_{k=1}^{D-1} \frac{1}{z^2} \right)^{-1}$$

$$= \frac{1}{\zeta + 1}.$$ 

Solving for $\zeta$ gives us the critical value $\zeta = z^2 - 1$. 

6
Although this distribution is a fixed point of (8), we need to analyze its linear stability to be sure that it is an attractor. We can do this by considering the partial derivatives
\[
\frac{\partial \hat{g}}{\partial g_{jk}} = \frac{z^2}{(\zeta + 1)^2} = \frac{1}{z^2}.
\]

Now, at every iteration we compute \( \hat{g} \) based on \( D - 1 \) messages \( g_{j1}, \ldots, g_{j(D-1)} \), so the distribution is stable so long as
\[
\sum_{k=1}^{D-1} \frac{\partial \hat{g}}{\partial g_{jk}} < 1,
\]
meaning that on average it is sufficient for
\[
\frac{\partial \hat{g}}{\partial g_{jk}} = \frac{1}{z^2} < \frac{1}{(D-1)}.
\]

Thus, when \( z > \sqrt{(D-1)} \), the message distribution at \( \zeta = z^2 - 1 \) is a real-valued \( \delta \)-function and \( L_z \) has no spectral density at this point. In the case of \( G(n, c/n) \), both \( D \) and \( D-1 \) are Poisson-distributed with mean \( c \), so the condition becomes \( z > \sqrt{c} \).

![Figure 4: First and second eigenvalues for \( L_z \) as a function of \( z \) for a single \( G(n, c/n) \) graph with \( n = 2000 \) and \( c = 3 \). For \( z \in [\sqrt{c}, c] \), the curve \( \zeta = z^2 - 1 \) appears to lie between \( \lambda_1 \) and \( \lambda_2 \).](image)

**An Upper Bound for the Edge of the Bulk**

In fact, even without an analytic expression for the fixed point distribution, we can compute an upper bound for the edge of the bulk. In particular, so long as \( D \) has an infinite tail, there is no spectral density for \( \zeta > z^2 - 1 \). First, if every \( g_{jk} \) is confined to the interval \([0, 1/z^2]\) on the real line, then
\[
0 \leq \frac{1}{\zeta + D} \leq \left( \frac{1}{\zeta} + D - z^2 \sum_{k=1}^{D-1} g_{jk} \right)^{-1} \leq \frac{1}{1 - \zeta} \leq \frac{1}{z^2},
\]

Thus for \( \zeta > z^2 - 1 \), there is a real-valued solution to (9) with support confined to this interval. As above, we can analyze the linear stability of this distribution by taking partial derivatives. To do so, we’ll write \( \hat{g} \).
in terms of its real and imaginary parts as \( \hat{g} = \hat{a} + i\hat{b} \), where

\[
\hat{b} = \frac{z^2 \sum_{k=1}^{D-1} b_{jk}}{\left( \zeta + D - z^2 \sum_{k=1}^{D-1} a_{jk} \right)^2 + \left( z^2 \sum_{k=1}^{D-1} b_{jk} \right)^2},
\]

and \( g_{jk} = a_{jk} + i b_{jk} \). Now, taking partial derivatives at the real line,

\[
\left. \frac{\partial \hat{b}}{\partial b_{jk}} \right|_{b_{jk}=0} = \frac{z^2}{\zeta + D - z^2 \sum_{k=1}^{D-1} a_{jk}},
\]

and since \( a_{jk} = g_{jk} \in [0, \frac{1}{z^2}] \) and \( \zeta > z^2 - 1 \),

\[
\left. \frac{\partial \hat{b}}{\partial b_{jk}} \right|_{b_{jk}=0} \leq \frac{z^2}{(\zeta + 1)^2} \leq \frac{1}{z^2}.
\]

Just as above, then, this distribution is stable so long as \( z > \sqrt{\langle D - 1 \rangle} \), or \( z > \sqrt{\zeta} \) in the \( G(n, c/n) \) case where the degrees are Poisson distributed with mean \( c \). Figure 4 compares \( \zeta = z^2 - 1 \) with \( \lambda_1 \) and \( \lambda_2 \) for a particular \( G(n, c/n) \) graph; for \( z \in [\sqrt{c}, \zeta] \), \( \zeta = z^2 - 1 \) lies between the first and second eigenvalues.

**Toward the True Bulk Edge?**

Our numerical results indicate that for \( G(n, c/n) \) the \( z \)-Laplacian in fact has zero spectral density well below \( \zeta = z^2 - 1 \) for \( z > \sqrt{\zeta} \). However, Figure 8 shows that the stable real-valued distribution is very different below this point; it is no longer confined to the interval \( [0, \frac{1}{z^2}] \), but instead appears to have infinite support. Because of this, it is unlikely that a similar argument to above will suffice to locate the true edge of the bulk. Instead, we turn to a study of the real-valued fixed point distribution with the hope that it will give us insight into the true value for the edge of the bulk.

**Figure 5:** Real fixed point message distribution, \( n = 10000, c = 3, z = c, \zeta = 18.9; \) computed from population dynamics iterations 101 – 1000.
The Fixed Point Message Distribution

Following [1], we can think of the fixed point message distribution as defining a random variable $Y(\zeta)$ which satisfies

$$Y(\zeta) = \mathbb{d} \left( \zeta + D + \sum_{j=1}^{D-1} Y_j(\zeta) \right)^{-1},$$  

(9)

where $D$ is distributed according to the degree distribution of the random graph ensemble, the $Y_j(\zeta)$ are i.i.d. copies of the random variable $Y(\zeta)$, and $\mathbb{d}$ indicates that the left- and right-hand sides have the same distribution. Bordenave and Lelarge prove that an equivalent equation to (9) is guaranteed to have a solution under some mild assumptions [1].

A closer examination of (9) give some insight into the structure of the real-valued distribution in Figure 5. We can think of $Y(\zeta)$ as a discrete convolution with the degree distribution; in the case of $G(n, p)$, we have

$$Y(\zeta) = \sum_{k=0}^{\infty} \frac{\text{Poisl}(k)}{\zeta + k + \sum_{j=1}^{k-1} Y_j(\zeta)},$$

(10)

where the $k$th term can be interpreted as the distribution of incoming messages to a vertex of degree $k$. Numerical results bear this out, and are captured in Figure 7, where we see that each of the humps in Figure 5 corresponds to a single term in (10).

Moreover, as in [1] we can derive an expression for the moment generating function of $Y(\zeta)$. Let $f(u, \zeta) = \mathbb{E} \left( e^{uY(\zeta)} \right) = \sum_{m=0}^{\infty} \frac{u^m \mathbb{E}[Y(\zeta)^m]}{m!}$ be the moment generating function; using the identity

$$e^{tw} = 1 - \sqrt{t} \int_{0}^{\infty} \frac{J_1(\sqrt{ut})}{\sqrt{t}} e^{-ltw} dt,$$

where $J_1$ is the Bessel function of the first kind of order one, we have
we can write
\[
   f(u, \zeta) = \mathbb{E} \left( e^{uY(\zeta)} \right) = \mathbb{E} \left( e^{iuY(\zeta)/i} \right)
   = 1 - \sqrt{u} \int_0^\infty \frac{I_1(\sqrt{ut})}{\sqrt{t}} \mathbb{E} \left( e^{-it(i/Y(\zeta))} \right) \, dt
   = 1 - \sqrt{u} \int_0^\infty \frac{I_1(\sqrt{ut})}{\sqrt{t}} \mathbb{E} \left( e^{i(\zeta+D+1-z^2\sum_{j=1}^D Y_j(\zeta))} \right) \, dt
   = 1 - \sqrt{u} \int_0^\infty \frac{I_1(\sqrt{ut})}{\sqrt{t}} e^{i(\zeta+1)D} \mathbb{E} \left( e^{-tz^2Y(\zeta)} \right) \, dt
   = 1 - \sqrt{u} \int_0^\infty \frac{I_1(\sqrt{ut})}{\sqrt{t}} e^{i(\zeta+1)} \phi \left( e^{tf(-t^2, \zeta)} \right) \, dt,
\]
where \( \phi(x) = \mathbb{E}(x^D) \). As above, for \( G(n, p) \) both vertex degree and excess degree are Poisson distributed with mean \( c \); in this case \( \phi(x) = e^{(x-1)} \), giving
\[
   f(u, \zeta) = 1 - \sqrt{u} \int_0^\infty \frac{I_1(\sqrt{ut})}{\sqrt{t}} \exp \left( t(\zeta + 1) + c \left( e^{tf(-t^2, \zeta)} - 1 \right) \right) \, dt.
\]
We obtain an expression for the \( n \)th moment of \( Y(\zeta) \), albeit one in terms of the moment generating...
function $f(u, \zeta)$, by taking

$$
\mathbb{E}(Y(\zeta)^n) = \frac{\partial}{\partial u^n} [f(u, \zeta)]_{u=0}

= - \int_0^\infty \frac{\partial}{\partial u} \left[ \sqrt{\frac{u}{t}} I_1(\sqrt{ut}) \right]_{u=0} e^{\ell(\zeta + 1)} \phi \left( e^\ell f(-t z^2, \zeta) \right) dt

= - \int_0^\infty \frac{\partial}{\partial u} \left[ \sqrt{\frac{u}{t}} \sum_{m=0}^\infty \frac{(-1)^m}{m! (m+1)!} \left( \frac{\sqrt{ut}}{2} \right)^{2m+1} \right]_{u=0} e^{\ell(\zeta + 1)} \phi \left( e^\ell f(-t z^2, \zeta) \right) dt

= - \int_0^\infty \frac{\partial}{\partial u^n} \left[ \sum_{m=0}^\infty \frac{(-1)^m m^{m+1}}{m! (m+1)! 2^{2m+1}} \right]_{u=0} e^{\ell(\zeta + 1)} \phi \left( e^\ell f(-t z^2, \zeta) \right) dt

= \int_0^\infty \frac{(-1)^{n-1}}{2^{2n-1} (n-1)!} e^{\ell(\zeta + 1)} \phi \left( e^\ell f(-t z^2, \zeta) \right) dt.

$$

The above equation drives home the recursive, self-similar nature of the limiting distribution, in that each moment of the fixed point distribution depends on every other moment; we cannot even hope to obtain an analytic or closed form expression for any of the moments independent of an analytic expression for the distribution as a whole.

5 Conclusion

The $z$-Laplacian is a promising addition to the current toolbox of spectral community detection algorithms. $L_z$ shares spectral properties with the non-backtracking operator that has been shown to vastly outperform other spectral protocols on sparse networks, and experimental observations of the $z$-Laplacian’s spectrum indicate that it too should succeed in extracting community structure down to the universal detectability threshold [3],[4]. Although we have not yet produced a rigorous argument to this effect, our study of the spectral properties of $L_z$ has employed message-passing and population dynamics approaches potentially usable on sparse random matrices more broadly, and we believe that continued study of the fixed point distribution of (9) will give us further insight into the spectrum of this novel operator.

Acknowledgments

This work was done in conjunction with the Santa Fe Institute Research Experience for Undergraduates Program.

References


Figure 8: Real (a) and complex (b) parts of the fixed point message distribution for $n = 1000$, $c = 3$, and $z = 2.36$ for $\zeta$ between 4.07 and 4.58 = $z^2 - 1$. $D$ was Poisson distributed with mean 3, and the distributions are taken over the 100th to 200th iterations. The stable real distribution persists well below $\zeta = z^2 - 1$. 