

A NOTE ON MARKOV NORMALIZED MAGNETIC EIGENMAPS

ALEXANDER CLONINGER

ABSTRACT. We note that building a magnetic Laplacian from the Markov transition matrix, rather than the graph adjacency matrix, yields several benefits for the magnetic eigenmaps algorithm. The two largest benefits are that the embedding becomes more stable as a function of the rotation parameter g , and the principal eigenvector of the magnetic Laplacian now converges to the page rank of the network as a function of diffusion time. We show empirically that this normalization improves the phase and real/imaginary embeddings of the low-frequency eigenvectors of the magnetic Laplacian.

1. INTRODUCTION

We consider the problem of building a diffusion operator on directed networks and graphs with adjacency matrix $W \in \mathbb{R}^{N \times N}$ that parallels Diffusion Maps on undirected networks [1]. A natural approach to this is to build the Magnetic Eigenmaps Laplacian [2],

$$L = D - T \circ W^{(s)},$$

where $(T \circ W^{(s)})_{i,j} = e^{2\pi i g (W_{j,i} - W_{i,j})} \frac{W_{i,j} + W_{j,i}}{2}$ and $D_{i,i} = \sum_j W_{i,j}^{(s)}$. This method assigns a complex rotation to asymmetric links. This approach generalizes the notion of building a diffusion map or graph embedding to work on directed graphs.

In this note, we observe that Markov normalization of the adjacency matrix,

$$P(x, y) = \frac{W(x, y)}{\sum_y W(x, y)},$$

allows us to introduce the notion of diffusion time steps, which is capable of capturing the evolution of the process. Markov normalization of the graph adjacency matrix yields several benefits:

- (1) normalizing the complex rotation matrix $e^{2\pi i g (P_{j,i} - P_{i,j})}$ by the density of the node, which emphasizes absorbing states and converges to the stationary distribution of the process,
- (2) stabilizing the phase embedding with respect to choice of the rotation parameter g ,
- (3) separating the long time trend of the process into the first eigenvector and recurrent states into subsequent eigenvectors, and
- (4) introducing diffusion time, which allows for the study of multi-step neighborhoods.

We further note that it is possible to augment the adjacency matrix by a transportation factor, as is done in the PageRank algorithm, when there exists an absorbing state [3].

For a small $\alpha > 0$, we replace $P \rightarrow (1 - \alpha)P + \alpha \mathbb{1}\mathbb{1}^\top$. This creates a small probability of jumping from any node to any other node, and thus escaping a sink in the process and making the process ergodic.

For the rest of the paper we will assume that the adjacency matrix P is ergodic, under the pretense that if there is an absorbing state, we simply add a transportation factor.

2. MARKOV PROPERTY AND DIFFUSION TIME

An important aspect of diffusion maps is the ability to separate the statistics of the sampling distribution from the geometry of the underlying manifold [4]. This is accomplished by normalizing the affinity between points by the overall transition probability,

$$P = T^{-1}W, \quad \text{where} \quad T_{x,y} = \begin{cases} \sum_y W(x,y), & x = y \\ 0, & x \neq y \end{cases}.$$

This also allows us to consider the properties of the multiple time step diffusion process P^t for $t \in \mathbb{R}^+$. For a symmetric process (i.e. undirected graph), the embedding of this process remains stable. This is because, if the eigendecomposition is $P = \Phi\Lambda\Phi^*$, then

$$P^t = (\Phi\Lambda\Phi^*)^t = \Phi\Lambda^t\Phi^*.$$

However, for a non-symmetric graph, this property no longer holds. Thus the eigenvectors of P^t will exhibit time dependence beyond a simple multiplicity factor.

The markov normalized magnetic Laplacian then takes the form of

$$L_{i,j}^{(t)} = D_{i,j} - e^{2\pi ig(P_{j,i}^t - P_{i,j}^t)} \frac{P_{i,j}^t + P_{j,i}^t}{2}, \quad \text{where} \quad D_{i,i} = \frac{1}{2} \sum_j (P_{i,j}^t + P_{j,i}^t).$$

Adjusting the weight matrix to be Markov normalized has several implications on the magnetic Laplacian. In diffusion on a symmetric graph, normalizing by the node density causes the principal eigenvector ϕ_0 to be a constant across all nodes and is thus trivial. However, when the network is not symmetric, the principal eigenvector takes on a much more important role. Namely, the principal eigenvector recovers the degree of asymmetry between nodes. This implies that, as t increases, the phase of the eigenvectors converges to the pagerank of the network.

Theorem 1. *Suppose there exists T such that $P_{i,j}^T > 0$ for all i, j . Let $L_{norm} = D^{-1/2}LD^{-1/2}$ be the normalized magnetic Laplacian. Then in the limit as $t \rightarrow \infty$, the principal eigenvector of the normalized Laplacian $L_{norm} = \lim_t L_{norm}^{(t)}$ (denoted ϕ) converge to*

$$\phi_i^{(g)} = c \cdot e^{2\pi gh_i} \left(\frac{1 + \sum_j P_{j,i}}{2} \right)^{1/2},$$

where $h = \text{pagerank}(P)$ (i.e. the stationary distribution of P), and $c \in \mathbb{C}$.

Proof. From Theorem 1 in [2], we know that the magnetic Laplacian L has a zero eigenvalue iff $\exists h$ such that $a_{i,j} = h_j - h_i$. Furthermore, we know that, in this case, $\phi_{k,i}^{(g)} = e^{2\pi igh_i} \phi_{k,i}^{(0)}$.

We also know that, since P is ergodic by assumption, $P_{i,\cdot}^t = h$ as $t \rightarrow \infty$, where $h = \text{pagerank}(P)$. That means $a_{i,j} = P_{i,j}^t - P_{j,i}^t = h_j - h_i$.

Let v be the principal eigenvector of $I - D^{-1/2}P_{sym}D^{-1/2}$, where $P_{sym} = \left(\frac{P+P^*}{2}\right)$ and $D_{xx} = \sum_y (P_{sym})_{xy} = \frac{1+\sum_x P_{xy}}{2}$. Then

$$\begin{aligned} (I - D^{-1/2}P_{sym}D^{-1/2})v = 0 &\implies (D^{1/2} - P_{sym}D^{-1/2})v = 0 \\ &\implies (D - P_{sym})x = 0 \\ &\implies (I - D^{-1}P_{sym})x = 0, \end{aligned}$$

where $x = D^{-1/2}v$. Because $D^{-1}P_{sym}$ is row stochastic, we know $x = \mathbb{1}$, so $v = D^{1/2}\mathbb{1}$.

Since $\phi_{k,i}^{(g)} = e^{2\pi i g h_i} \phi_{k,i}^{(0)}$, we know

$$\begin{aligned} \phi_i^{(g)} &= e^{2\pi i g h_i} (c \cdot D_i^{1/2}) \\ &= c \cdot e^{2\pi i g h_i} \left(\frac{1 + \sum_j P_{j,i}}{2} \right)^{1/2}. \end{aligned}$$

□

This only applies in the limit as $t \rightarrow \infty$. The embeddings of $L^{(t)}$ for small t , and even $t = 1$, yield valuable properties due to the Markov normalization. We show this empirically in Section 3.

3. EXAMPLES

In all examples of the Markov normalized process, we will refer to a value of $g \in [0, 1/2)$ that is on the same scale as for the unnormalized Laplacian. However, due to the weights being normalized to sum to 1, they are actually on a different scale than the unnormalized weights matrix. For that reason, we set $g \leftarrow g/\max(\max(P))$ for the Markov normalized weights so that the degree of rotation is on the same order of magnitude for both methods.

3.1. Three Cluster Example. We begin with the three cluster example from [2]. This example is particularly simple given that the complex rotation of three clusters will, more often than not, keep the clusters separated. However, for small values of g , rotation of the unnormalized asymmetric weights matrix yields insufficient separation between the clusters, as shown in Figure 1. In fact, Figure 2 shows the percentage of points that cluster correctly across 100 uniform random small values of g ($g < 0.25$), where clustering was done by k-means with $k = 3$.

By incorporating the notion of diffusion time, we are also able to observe the time dynamics of this process. Figure 8 in the appendix shows the embedding for each $t \in \{1, 2, \dots, 9\}$. Notice that after 3 and 4 time steps, the clusters become momentarily blurred. This is because, by that time, most of the energy from cluster 1 will have moved to clusters 2 and 3, etc. However, after that momentary blurring, the process stabilizes and we continue to see the three clusters rotating about each other for larger t . This tells us that there is a global mixing time of $t > 1$ for the process to satisfy $P_{i,j} > \epsilon$ for all i, j .

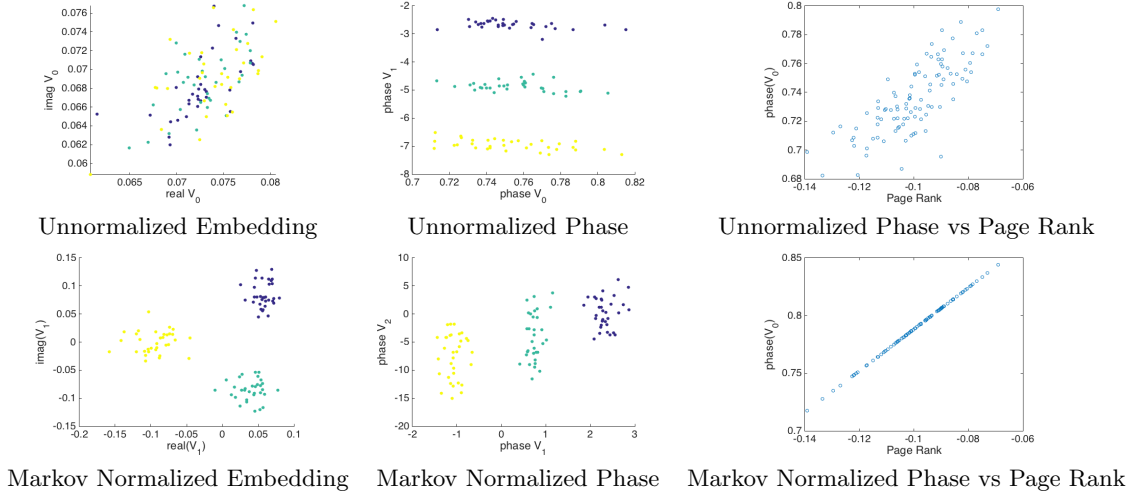


FIGURE 1. Three clusters with $P(\text{inCluster}) = 0.5$, $P(\text{outCluster}) = 0.5$, $P(\text{rotateClockwise}) = 0.9$, and with parameter $g = 0.04$. The markov normalized embeddings are with $t = 1$, though the bottom right page rank plot is for $t = 4$ to allow sufficient time to converge.

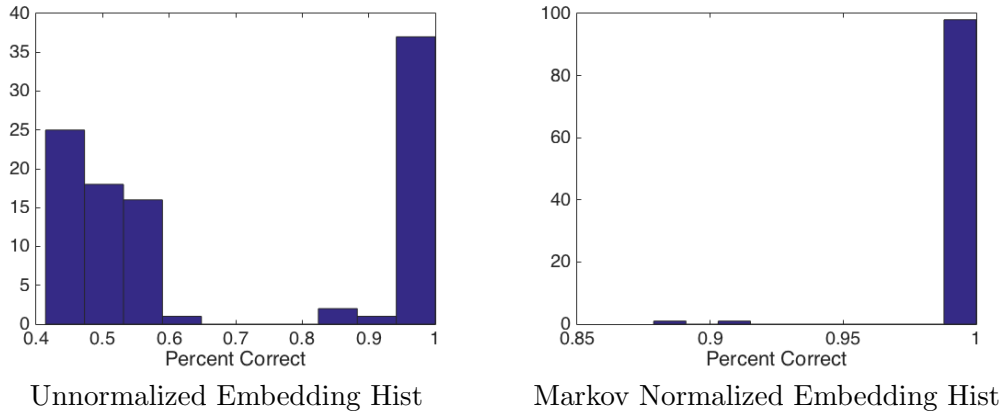


FIGURE 2. Three clusters with $P(\text{inCluster}) = 0.5$, $P(\text{outCluster}) = 0.5$, $P(\text{rotateClockwise}) = 0.9$, and with random parameter $g < 0.25$. The markov normalized embeddings are with $t = 1$.

3.2. Circle with Drift. We consider data distributed on a unit circle, with drift in the counterclockwise direction. This is an interesting process as it has a very large mixing

time. We define the affinity between points to be

$$k(x, y) = \begin{cases} e^{-\|x-y\|^2/5\sigma^2}, & \angle xy \geq 0 \\ e^{-\|x-y\|^2/\sigma^2}, & \angle xy < 0 \end{cases},$$

where $\angle xy$ measures the angle between the points. The affinity matrix is seen in Figure 9 in the appendix.

Figure 3 compares the different normalizations and eigenvectors for the circle with drift. By normalizing the weights matrix, the first eigenvector is once again the trivial eigenvector. Subsequent eigenvectors recover the geometry of the circle.

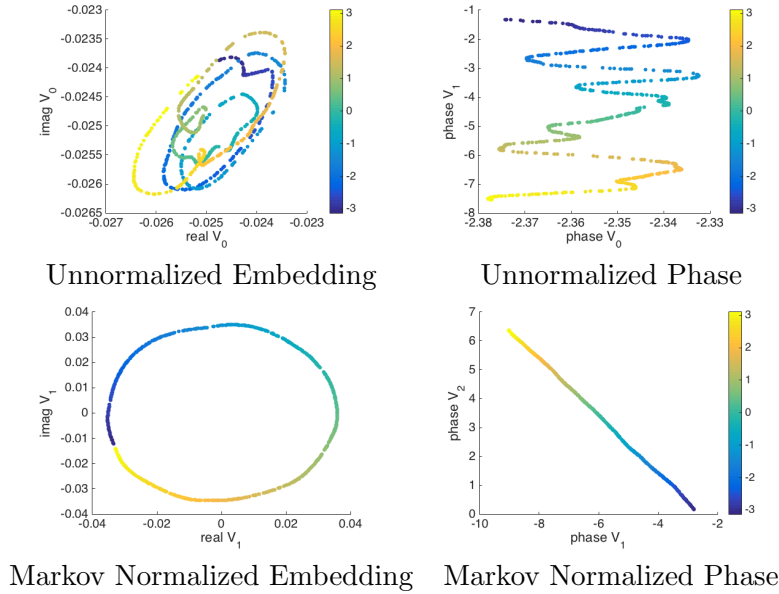


FIGURE 3. Circle with drift, colored by angle in original data. Top: Unnormalized and leading two eigenvectors, Bottom: Markov normalized and first two non-trivial eigenvectors. $g = 0.04$

The higher order eigenvectors also contain useful information, as in diffusion maps. They mimic the sines and cosines of the angle around the circle, as the Fourier functions $\{e^{inx}\}$ are eigenfunctions of the Laplacian on the circle. We wish to recover these functions independent of the sampling density on the circle. Figure 10 in the appendix shows that the Markov normalized weights matrix gives a closer approximation of the sin multiples on the circle that comes from the unnormalized weights matrix. Note that the data used here is not uniformly sampled across the circle.

3.3. Bow Tie Example. Another issue with not including a diffusion time is the need for g to implicitly assume the number of clusters. By setting g to be $1/k$ for k clusters as the authors suggest [2], the complex rotation can lose valuable information about transitions

that aren't k periodic. We demonstrate this with an example similar to the three cluster example from Section 3.1. We build a three cluster cycle among clusters 1,2, and 3, and a 5 cluster cycle among clusters 1, 4, 5, 6, and 7. By normalizing the weights, we see in Figure 4 that the non-principal eigenvectors capture the geometry of the process. Moreover, we discount the fact that cluster 1 appears in both cycles, and will end with twice the energy of the other clusters for large t .

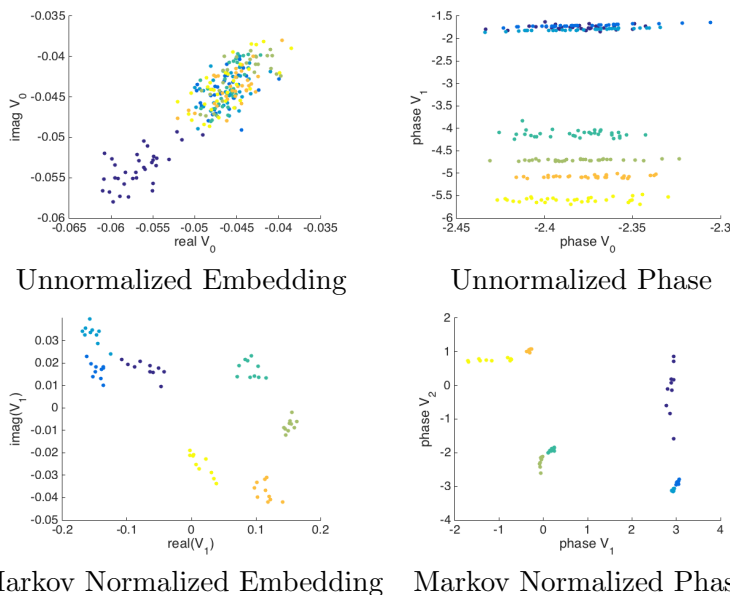


FIGURE 4. Bow tie clusters with $P(\text{inCluster}) = 0.5$, $P(\text{outCluster}) = 0.5$, $P(\text{rotateClockwise}) = 0.9$, and with $g = 0.04$. For Markov normalization, we take $t = 1$.

We are also able to observe the time dynamics of this process. Figure 11 in the appendix shows the embedding for each $t \in \{1, 2, \dots, 9\}$. Notice that after 7 and 8 time steps, the clusters become momentarily blurred. This is because, by that time, most of the energy from one end of the bow tie has moved to the other. However, after that momentary blurring, the process stabilizes and we continue to see the three clusters rotating about each other for larger t . This tells us that there is a global mixing time of $t > 6$ for the process to satisfy $P_{i,j} > \epsilon$ for all i, j . We can also see these dynamics in the affinity matrix in Figure 12. And finally, we see in Figure 12 that after a large t , the phase of the leading eigenvector converges to the page rank of the process.

3.4. Hidden Circle. As a different example, we consider a mix of absorbing and recurrent states. The data we consider is a unit square with drift from left to right via

$$k(x, y) = \begin{cases} e^{-\|x-y\|^2/3\sigma^2}, & x_1 < y_1 \\ e^{-\|x-y\|^2/\sigma^2}, & x_1 \geq y_1 \end{cases}.$$

On top of that, we add an annulus in the middle of the unit square with counterclockwise flow. The results of the phase and embedding are in Figure 5.

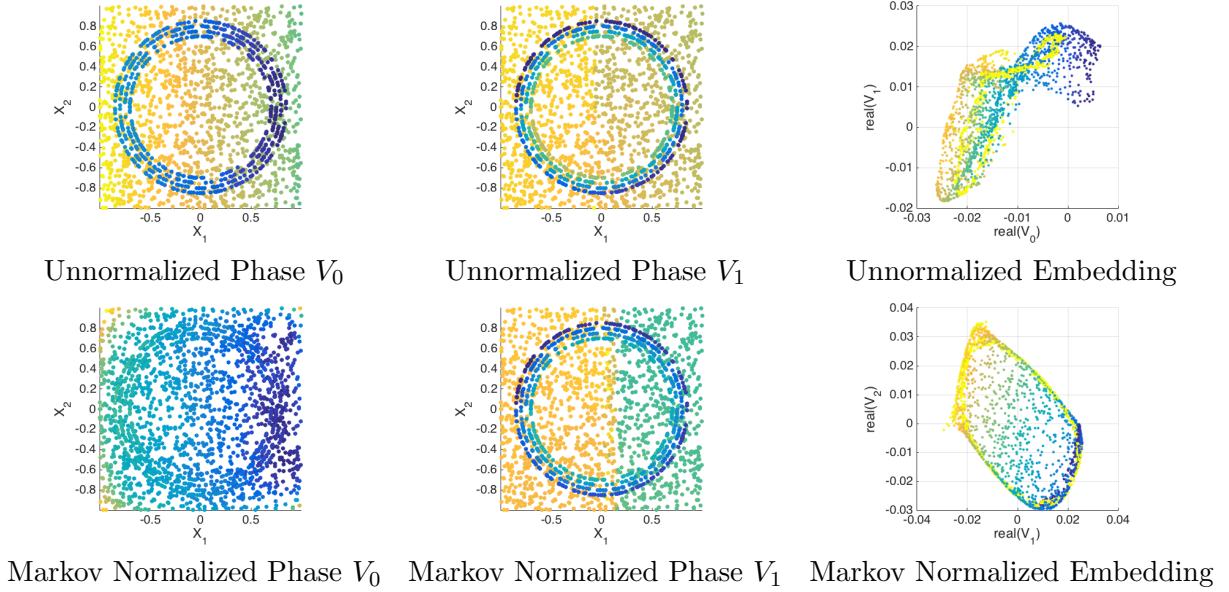


FIGURE 5. $g = 0.24, t = 4$

Also, by not normalizing the phase rotation, low density points become separated in the phase space (see Figure 6). This leads to discontinuity in the torus projection of the phase, because the top and bottom edges of the square do not have much energy coming in.

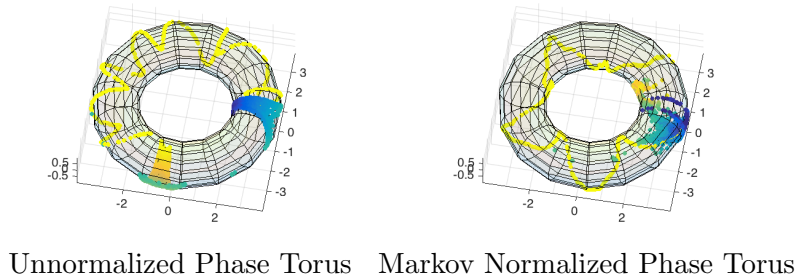


FIGURE 6. $g = 0.24, t = 1$

3.5. Absorbing States. As a final example, we return to the three cluster example from Section 3.1. We change this by removing one node’s out going edges, making that node an absorbing state. We augment the weights matrix for the normalized weights matrix by adding a teleportation parameter of $\alpha = 0.1$.

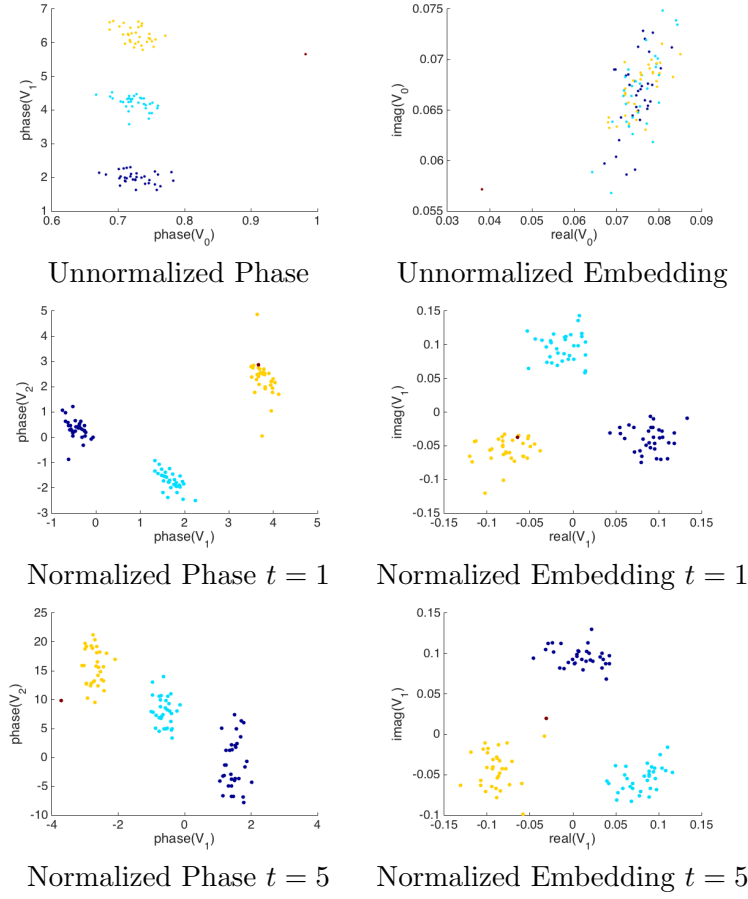


FIGURE 7. Three cluster data with absorbing state from orange cluster. Absorbing state in red. Top: diffusion time of 1, Bottom: diffusion time of 5. $g = 0.04$

We see in Figure 7 that, as time progresses, the state becomes more central to the Markov normalized process while still maintaining the geometry of the three clusters. Moreover, the phase of the principal eigenvector converges to the page rank of the process, correctly placing the absorbing state in middling importance as in Figure 13 in the appendix. This is compared to the unnormalized magnetic Laplacian, which captures the absorbing nature of the red node, but loses the geometry of the other clusters or the relationship between the absorbing state and cluster 3.

4. CONCLUSION

We note that Markov normalization of the graph adjacency matrix yields several benefits. Normalization of the complex rotation matrix T converges to the page rank of the graph

for a large time, as in Section 3.1. The Markov normalization also stabilizes the phase embedding over a larger choice of g , as in most examples in Section 3. By emphasizing the absorbing states in the first eigenvector, this normalization separates the recurrent states into subsequent eigenvectors as in Section 3.4. And finally, we can introduce diffusion time into the process and observe the multi-step neighborhood dynamics of the graph, as in most examples in Section 3.

REFERENCES

- [1] Coifman, Ronald R., and Stéphane Lafon. "Diffusion maps." *Applied and computational harmonic analysis* 21.1 (2006): 5-30.
- [2] Fanuel, Michaël, et al. *Magnetic Eigenmaps for Visualization of Directed Networks*. arXiv preprint arXiv:1606.08266 (2016).
- [3] Page, Lawrence, et al. *The PageRank citation ranking: bringing order to the web*. (1999).
- [4] Lafon, Stéphane S. *Diffusion maps and geometric harmonics*. Diss. Yale University, 2004.

APPENDIX

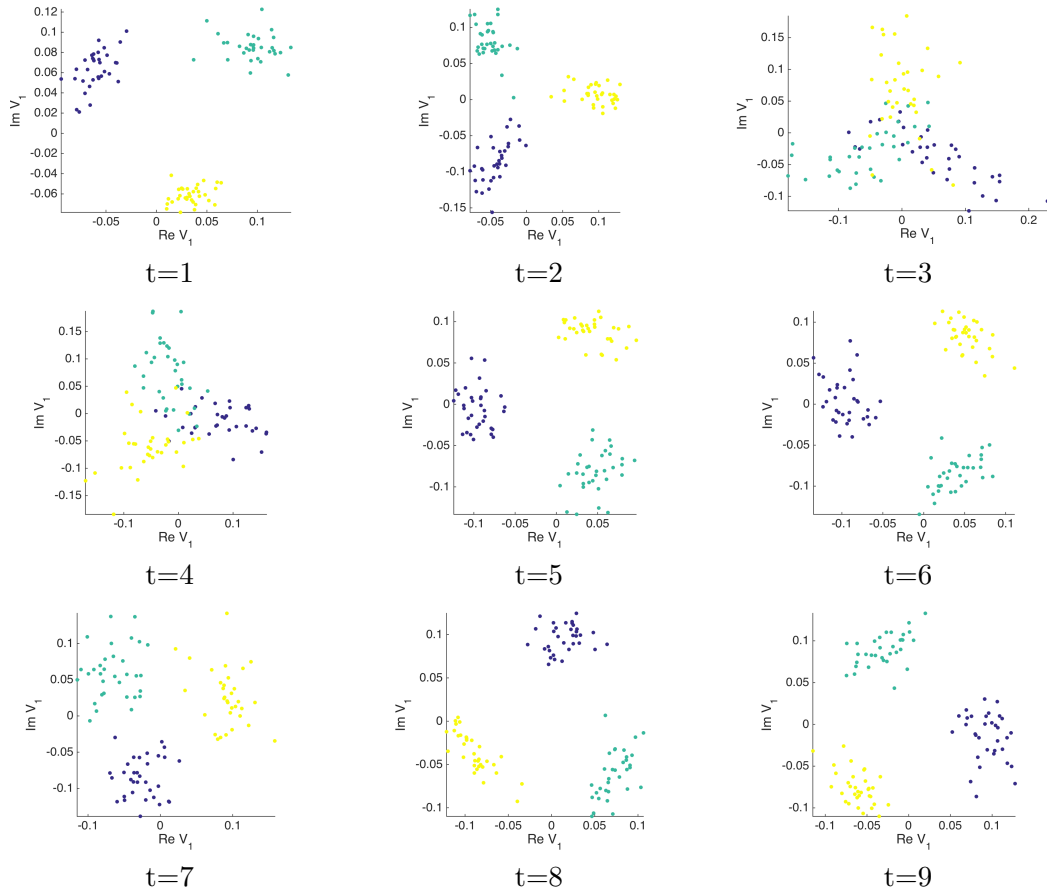


FIGURE 8. Three clusters with $P(\text{inCluster}) = 0.5$, $P(\text{outCluster}) = 0.5$, $P(\text{rotateClockwise}) = 0.9$, and with $g = 1/4$. Evolve cluster from $t \in \{1, \dots, 9\}$.

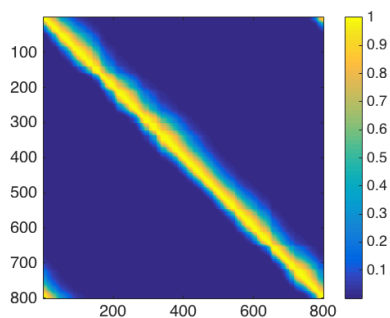


FIGURE 9. Affinity matrix for circle with drift

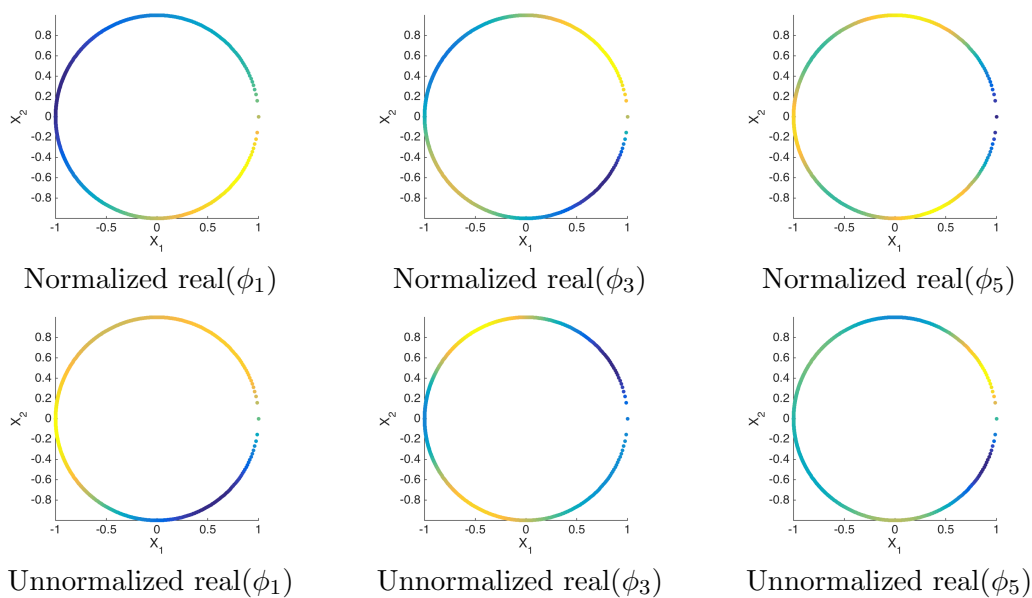


FIGURE 10. Circle with drift, colored by corresponding real part of $\text{real}(\phi_i)$. Top: Markov normalized, Bottom: Unnormalized. The eigenvectors correspond to multiples of sinusoids, the eigenvectors of the DFT matrix.

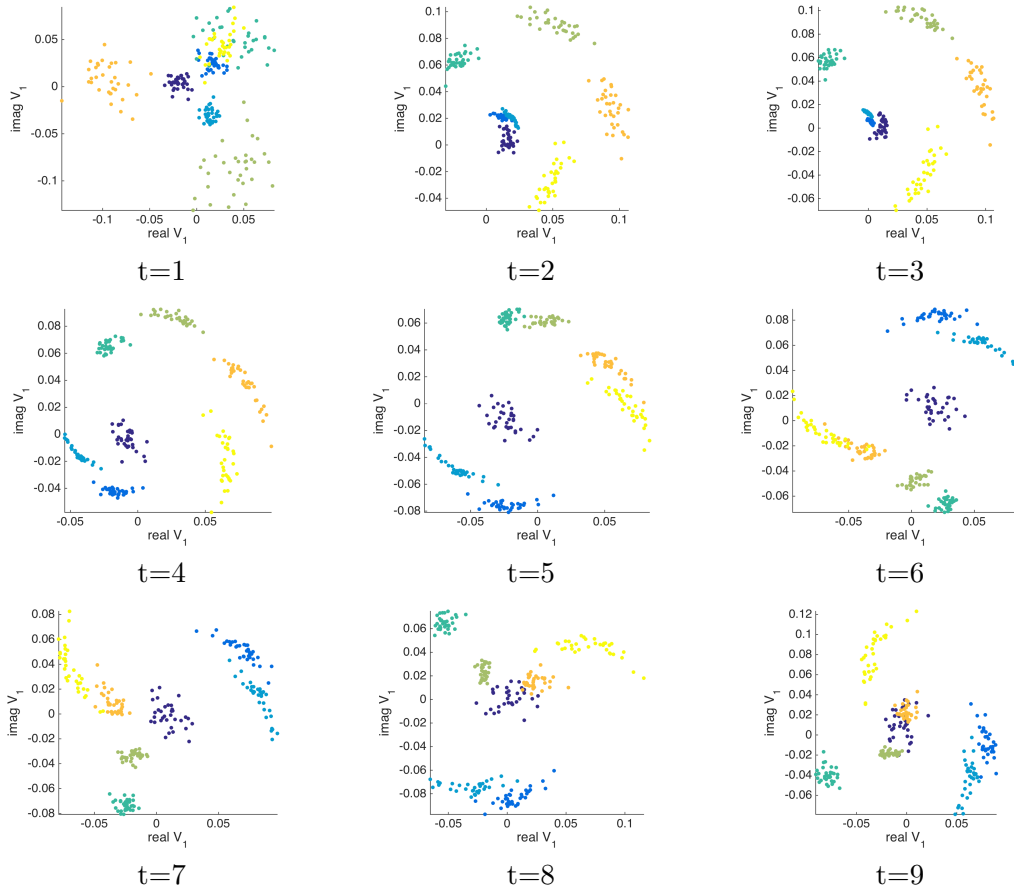


FIGURE 11. Bow tie clusters with $P(inCluster) = 0.5$, $P(outCluster) = 0.5$, $P(rotateClockwise) = 0.9$, and with $g = 1/4$. Evolve cluster from $t \in \{1, \dots, 9\}$.

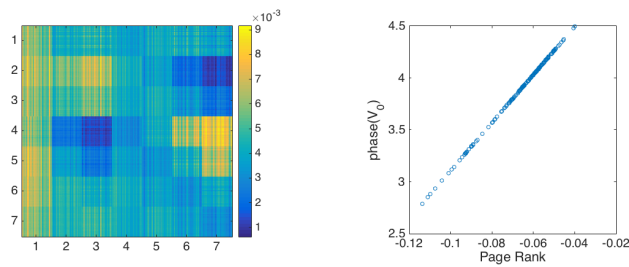


FIGURE 12. Bow tie clusters affinity at $t = 7$ (left), and bow tie clusters $phase(V_0)$ at $t = 10$ (right).

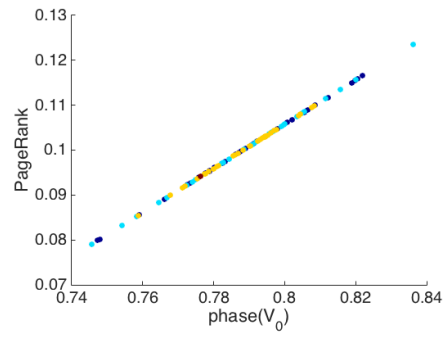


FIGURE 13. Absorbing state principal eigenvector phase compared to the page rank of the process.