**Abstract:** We examine the generalization of a fully-distributed method for computing spectral clusters to the case of aperiodic strongly-connected digraphs. In doing this we show that the sign of eigenvector components can be obtained via the Fourier series of a local wave-equation relaxation. Further we provide a runtime complexity for this variety of digraphs differing from the undirected case.

**Introduction:**

* **Spectral Clustering:** A bulk of work has been done in examining the problem of how to best determine clusters based on a similarity graph. There are many ways of doing this with varying runtimes and degrees of accuracy. One of the best documented of these is the method of spectral clustering **[1]**. The method relies on the well-known Cheeger’s inequality to approximate the cut of least conductance from the first eigenvector of the graph Laplacian. In the undirected case we say that a connected graph has the graph Laplacian where is the diagonal-degree matrix of the graph and is the adjacency matrix of . has several nice properties, among them are the fact that all eigenvalues are greater than or equal to zero, and most importantly the first non trivial eigenvalue obeys Cheeger’s inequality,

where is Cheeger’s constant and can be intuitively understood as the minimal-conductance cut on the graph defined as…

This allows for us to cluster based on the entries of the eigenvector associated to this eigenvalue . In most literature and in application this is done via the k-means algorithm and has a runtime of

* **Wave Equation Clustering:** In 2011 a paper was published outlining a fully distributed method for computing spectral clusters **[2]**. We outline the algorithm here and provide some important facts from the paper.

The main concept of the algorithm is to treat the graph as a drum-of-sorts and to locally relax the wave equation at each node. Then the first peak of the Fourier transform at each node assigns it to a cluster based on the sign of the associated Fourier coefficient. We define the discretized wave equation as follows,

where denotes the value of node at time , is the transmission speed of the wave equation, . It turns out that to ensure stability one can show through some algebra that one requires,

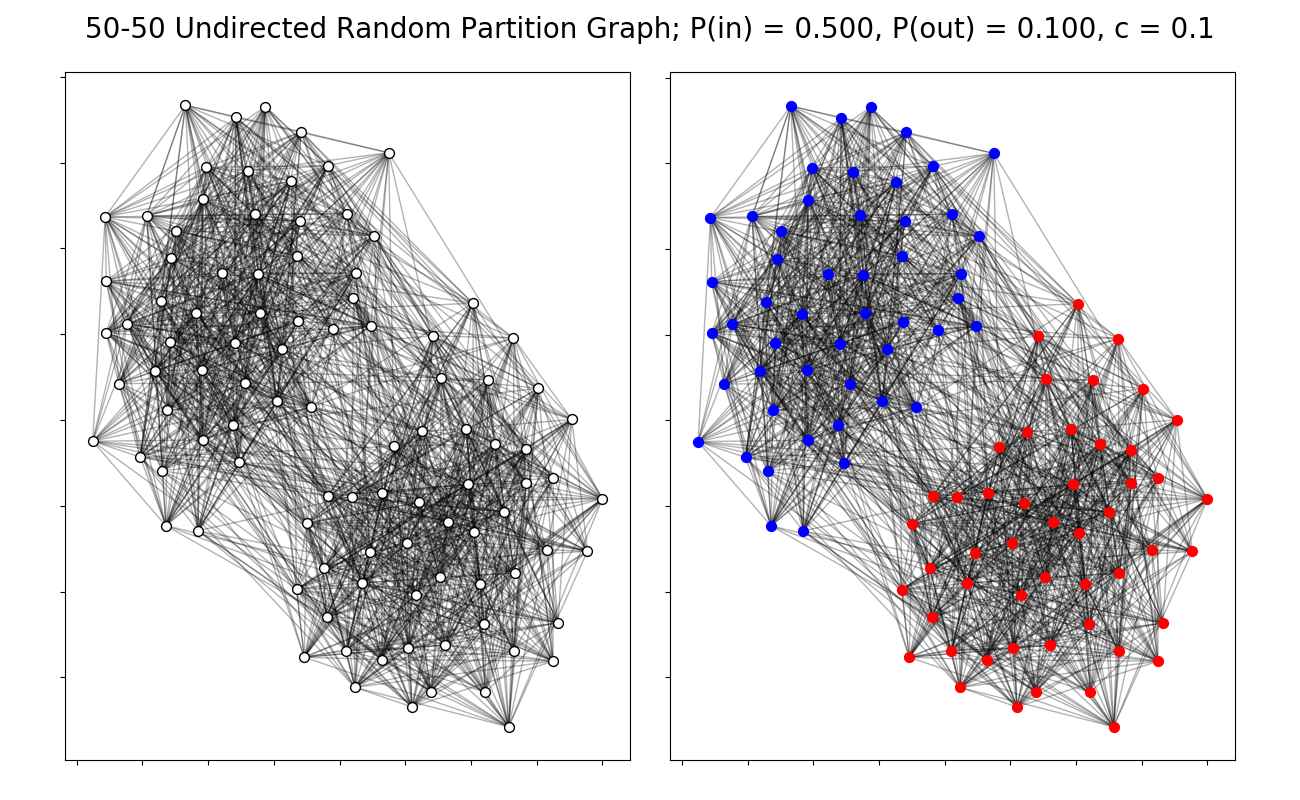
where is the spectral radius of . Recall that to assign two clusters we need to

resolve the first peak in the Fourier series. Therefore, for some number of cycles and

frequency we know that will be sufficient for obtaining a binary cluster.

through some simple algebra from here it can be shown that,

* **Example:**



Above is an example of a given cluster from the distributed algorithm on a 100 node

undirected partition graph with inter-cluster connection probability 0.1 and intra-cluster

connection probability 0.5.

**Extension:**

* **The Directed Graph Laplacian:** We are trying to extend this wave equation clustering method to the realm of directed graphs so we clearly require a new notion of the graph Laplacian . Luckily, assuming our graph is strongly-connected and aperiodic Chung has written about an alternative Laplacian we will call that is defined as follows **[3]**,

where is the stationary matrix of the transition matrix on the graph. This matrix is

Hermitian and as such all eigenvalues again are greater than or equal to zero, but most

Importantly the first non-trivial eigenvalue of again obeys Cheeger’s inequality and as

Such can be used for spectral clustering.

* **Theory:** We now have to ensure stability of our relaxation. This entails bounding the spectral radius of our Laplacian LD. This is known to be bounded above by 2 **[4]**. However, the derivation is difficult to find in literature so we outline it here,
  + **Proof:** First we consider the vector understood as the square root of the stationary matrix applied to the vector of all 1’s. Call then,

Showing that is both a left and right eigenvector of with eigenvalue 1.

Those familiar with the Perron-Frobenius theorem **[5]** will notice that this implies the spectral radius of , . It turns out that the same property

holds for .

Now we assume and call . Then,

. Therefore, we write,

We know and , so we have reached a contradiction and know . From here we simply need a straightforward application of the triangle

Inequality,

Therefore, the same bounds on as in the undirected case maintain stability of

our relaxation.

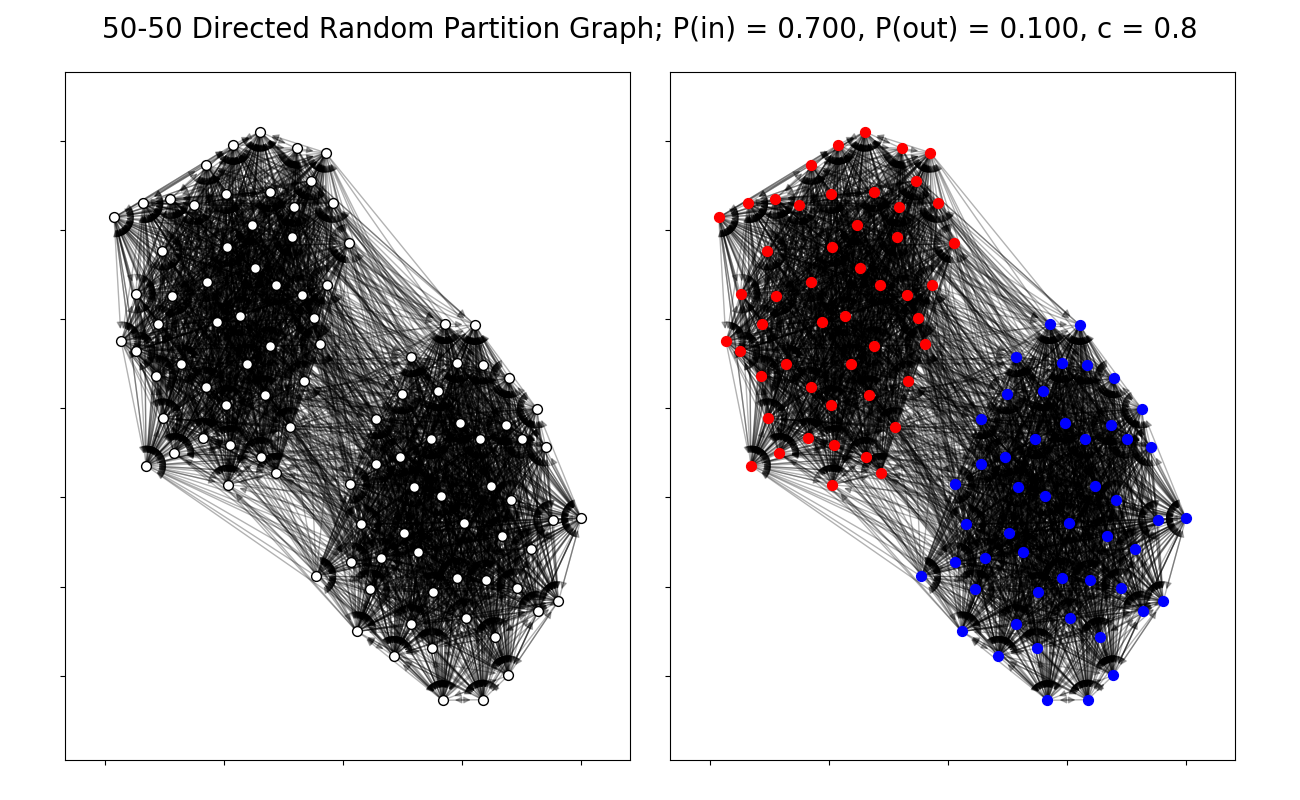
* **Runtime Complexity:** We now analyze the runtime complexity of the algorithm on our class of directed graphs. Recall our runtime complexity in the undirected case,

Where is the first non-trivial eigenvalue of L(G). Recalling the analysis from earlier this comes from some algebra based on the term . This same analysis works in the directed case, so we acquire the same time complexity.

Luckily, we can classify a “worst case scenario” for our runtime based on a lower bound for provided by Sinan Aksoy’s thesis **[6].**

where is the diameter of our digraph, and is the out-degree of node . This gives us the following run time complexity,

* **Cluster Example:**



Above is an example of a given cluster from our extended algorithm on a 100-node directed partition graph with inter-cluster connection probability 0.1 and intra-cluster connection probability 0.7.

**Conclusions:**In this paper we have outlined an extension of the wave-equation based method for spectral clustering to the case of directed, aperiodic, strongly-connected graphs. We have illustrated why the stability analysis remains the same, and why we obtain the new runtime complexity of .

**Citations:**

1. Chung, “Spectral Graph Theory”, AMS, 1997
2. Sahai, Speranzon, Banaszuk, “Hearing the Clusters in a Graph”, United Technology Research Center, 2011
3. Chung, “Laplacians and the Cheeger inequality for directed graphs”, UCSD, 2004
4. Chung, Kenter “Discrepancy inequalities for directed graphs”, Elsevier, 2014
5. Chung, Lu, “Complex Graphs and Networks”, AMS, 2006
6. Aksoy, Chung, “Random walks on directed graphs and orientations of graphs”, UCSD, 2017