VRDI 2019 Networks Breakout
Day 4: Network Dynamics

Daryl DeFord

June 27, 2019

1 Introduction

In order to understand the effects of topological structure on real life systems it is frequently useful to consider dynamics on networks. Many of the most powerful descriptive invariants of network theory are revealed through considering the action of some operator across a network. Most commonly the operators are assumed to be linear in some fashion so that techniques from linear algebra, such as eigenvalue analysis, can be performed. As we shall see this leads to many fascinating results, especially in the case of the graph Laplacian.

2 Markov Dynamics

The first type of dynamics we consider are those governed by Markov chains. For our purposes a Markov process is a stochastic process where the conditional probabilities of each state depend only on the current state. In the context of networks, we are usually interested in discrete state Markov chains, this is a finite collection of states, together with a (sequence of) matrix(ces) whose entries capture the transition probabilities, i.e. $A_{i,j}$ is the probability of transitioning from state j to state i.[1] Here we will generally only consider time homogeneous systems with a fixed transition matrix. This matrix is left stochastic and hence it can be shown (see next subsubsection) with Perron–Frobenius that one is the largest eigenvalue of the matrix and the corresponding eigenvector has all positive entries and can hence be interpreted as a probability vector (perhaps after normalizing). It is this steady state vector that we are most interested in.

We saw an example of this already in the context of centrality scores, when we considered the matrix AD^{-1}. This matrix represents the transition matrix of a Markov process defined on the nodes modeling a random walk on the network, where the edges are selected uniformly. This means that if the walker is at node i then proceeding to the next step the walker has $\deg(i)$ choices or equivalently each edge is chosen with probability $\frac{1}{\deg(i)}$. This is a more interesting notion in the context of directed networks, since for undirected networks the steady state proportions are just given by $\frac{D\mathbf{1}}{||D\mathbf{1}||}$.

2.1 Perron–Frobenius

It is a little bit of a pet peeve of mine that most sources about networks gloss over how Perron–Frobenius actually applies to stochastic matrices so here is an outline. To begin with, Perron–Frobenius guarantees that for certain classes of non–negative matrices there exists a unique, real maximum eigenvalue of multiplicity one, whose corresponding eigenvector has all positive entries.[2] Moreover, this is the only eigenvector of the matrix with strictly positive entries. Before proceeding, we need to discuss which types of matrices actually satisfy the hypotheses of this theorem.

The Perron–Frobenius theorem applies directly to any matrix with strictly positive entries. In our Markov example this means that at any time step any state may be reached from each other state. Additionally,

[1] It is traditional in the probability/statistics literature to have stochastic matrices act on the right of column vectors. I am not a fan of that convention.

[2] The theorem actually contains more content than this, including a very efficient formula for the projection onto the corresponding eigenspace. This portion of the theorem does not apply for operators that are irreducible and imprimitive.

the theorem applies without modification to primitive non–negative matrices. A matrix is primitive if there exists an integer m so that the m^{th} power of the matrix has all positive entries. Primitive matrices can also be characterized as irreducible, aperiodic matrices. A matrix is irreducible if for all i, j there exists an m such that the m^{th} power of the matrix has a positive entry in the i, j position. This can also be expressed by showing that the associated digraph is strongly connected.

The period of a non–negative matrix is the greatest common divisor of the lengths of closed paths in the associated digraph. An irreducible matrix is aperiodic and hence primitive if its period is equal to one. Perron–Frobenius for irreducible matrices has the same eigenvalue consequences described in the previous paragraph. Perron–Frobenius does not apply directly to reducible matrices. However, every reducible matrix may be permuted to an upper block diagonal form whose diagonal blocks are irreducible. Since the spectrum of the matrices must be the same sometimes information can be gleaned from this change of variables. This is the technique sometimes used for general stochastic matrices.

Unfortunately, many stochastic matrices, especially those attached to complex networks are neither primitve nor irreducible. In this case it is still true that they have leading eigenvalue one and non–negative (not necessarily positive) eigenvector. Since the matrix is left–stochastic the all ones vector is a right eigenvector corresponding to one. Gershogorin's Circle Theorem guarantees that one is at least a maximal eigenvalue in norm for the matrix. Additionally, since the matrix is non–negative we still have that every eigenvector corresponding to one has all entries with the same sign, since $A - I$ only has non–positive entries on the main diagonal. In this case there is no guarantee of multiplicity one. An easy example to see this failure is with a Markov chain that has several absorbing states[3]. In these cases there is an eigenvector corresponding to the eigenvalue one for each possible absorbing state. Clearly one is still the leading eigenvalue and this set of vectors is linearly independent of dimension equal to the number of absorbing states.

2.2 Basic Definitions

The transition matrix of a Markov chain can be interpreted as a weighted, directed graph with the states as nodes and the probabilities as edge weights. This provides convenient language for characterizing the properties of a given chain. For the remainder of this section we will not distinguish between a Markov chain and its associated digraph. This following are some basic definitions associated to Markov analysis.

1. (essential) A state i in the chain is essential if for every other state that can be reached from i there is a path returning to i.

2. (irreducible) A Markov chain is irreducible if it is strongly connected.

3. (primitve) A Markov chian is primitive if there exists an integer m such that any two (not necessarily distinct) vertices in the graph can be connected by a path of length m.

4. (steady state) A stead state of a Markov chain is an eigenvector corresponding to one. For primitive chain this can be computed as the limit of the matrix powers. This property implies that for such a chain the initial distribution has no consequences in the limit.

5. (period) The period of a state is the greatest common divisor of all lengths of cycles that begin at the state. A state is aperiodic if its period is one. Note that in bipartite graph, all periods must be even since there areno odd cycles.

6. (hitting time) The hitting time T_i of a state is the random variable associated to the chain counting the minimum number of steps until the path returns to the state.

7. (transient) A state i is transient if there is non–zero probability that a path from i never returns to i. A non–transient state is called recurrent. If we define $q_i^n = P(T_i = n)$ then the state is transient if $\sum_{n=1}^{\infty} q_i < 1$.

8. (mean recurrence time) The mean recurrence time of a node is the expected value of the hitting time $M_i = \sum_{n=1}^{\infty} n * q_i^n$. If M_i is finite then i is positive recurrent.

[3] defined below

9. (absorbing) A state is absorbing if it is impossible to leave. This occurs when a node has no out edges. These states are also known as sinks.

10. (source) A state is a source if it has no in edges.

11. (ergodic) A state is ergodic if it is aperiodic and positive recurrent.

2.3 Absorbing Markov Chains

A Markov chain is said to be absorbing if it contains at least one absorbing state and there is at least one finite path from each state to an absorbing state. We expect the steady state solutions to these chains to distribute the probability only among the absorbing states. These chains are usually analyzed in terms of a fundamental matrix $F = \sum_{n=1}^{\infty} P^n$ where P is the transition matrix relating only the transient states. Using Neumann series we can realize this matrix as $F = (I - P)^{-1}$. Then, we can interpret the i, j entry of F as the expected number of times state j is visited if the initial state is i. The variance can be computed from a matrix derived from F with the Hadamard product.

It is easy to see that the interpretation of $F_{i,j}$ implies that the expected number of steps beginning at i until the flow is absorbed is captured by $F\mathbf{1}$. Furthermore, the likelihood of reaching one transient state beginning at another is captured by $[(F - I)\operatorname{diag}(F)^{-1}]_{i,j}$, while the probability of being absorbed by a particular absorbing state is $[FR]_{i,j}$ where R is the transition matrix connecting the transient states to the absorbing states.

2.4 Uses in Complex Networks

Considering Markov chains on networks we usually use the $0, 1$ structure of the adjacency matrix of the graph to determine where to assign probabilities under the assumption that there is no possibility of flow between disconnected nodes, with the caveat that probabilities may be added along the diagonal to capture the possibility of remaining in place. The most common Markov process on networks uses only the information in the adjacency matrix to define the transition probabilities. This is sometimes known as the normalized random walk Laplcian: AD^{-1}.

This is model of a random walk on the network where the probability of moving from i to one of its neighbors j is $\frac{1}{\deg(i)}$. This weighting assumes that each edge is taken uniformly at random. In this case the interpretation of the steady state is as a limiting distribution of probabilities. We expect more highly connected nodes to have higher entries in this steady state, so the magnitudes of the steady state vector are frequently used as a measure of centrality on the network. Other probabilities on the edges can be enforced to reflect extra structure known from the application or edge weights associated to the network. In order to study flows across networks extra nodes can be added. For example, adding a source node to set the initial distribution, or adding a sink node to transform the walk into an absorbing Markov process.

This interpretation also shows up in several other network contexts. For example, normalized spectral clustering can be computed in terms of the eigenvectors of this matrix. Additionally, many other types of Markovian dynamics are defined on networks for specific applications because in general they admit simpler solutions than more general models, even if they are not fully representative of the actual application.

3 Diffusion

Diffusion on networks is modeled as a discrete version of continuous diffusion processes, such as those governed by the heat equation. The idea is that each node is affected only by its neighbors and that the quantities associated to each node "diffuse" or move from areas of high concentration to lower concentration areas. We model this by assuming that for any initial distribution φ on the nodes of the network the values change by a scalar multiple (the diffusion constant c) of the sum of the differences between the value at each node and the values at each of its neighbors. Symbolically, this is $\frac{d\varphi_i}{dt} = c\sum_j A_{ij}\varphi\varphi_i - \varphi_j$ or $\frac{d\varphi}{dt} = c(D - A)\varphi$ in matrix form.

This matrix $L = D - A$ is called the graph Laplacian. It is one of the most important operators associated to a given network and its eigenvalues have deep connections to the structure of the graph. Returning to

the diffusion model we see that $\frac{d\varphi}{dt} = cL\varphi$ is a linear differential equation and since L is symmetric it is diagonalizable with a simple solution. Since all of the eigenvalues of L are non–negative the solution converges to a steady state in the limit that corresponds to equidistribution of the initial vector across the network as we would expect from the continuous case. Since the solution or steady state is so simple we are usually more interested in the rate at which the function converges for various inputs. This rate is controlled by the eigenvalues.

3.1 Normalized Diffusion

The Laplacian $L = D - A$ defined in the previous section is not entirely standard. There are several other matrices that go by the name "graph Laplacian", usually normalized versions of the operator above. For example if we want the change at a node to be equal to the average of its neighbors instead of the sum we can define $\hat{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. This version is usually preferred mathematically for its relation to invariants of the graph. It also has a natural connection to the random walk Markov matrix associated to a network which shows up in the context of clustering.

The second eigenvalue of this normalized Laplacian provides the solution to the normalized cut problem. It is also (up to a change of coordinates) equivalent to a scalar perturbation the random walk Laplacian, so depending on context we can translate eigendata between the matrices if one of the matrices is more computationally tractable. A comparison of the bounds of the eigenvalues of this normalization and the standard Laplacian shows that their proportions remain unchanged, but the magnitudes are compressed in this realization.

3.2 Properties of the Laplacian

The graph Laplacian can be constructed as N^TN where N is an incidence matrix for the graph. This implies that the Laplacian is positive definite. It is also clear that L is symmetric from its construction so L has several nice algebraic properties such as orthogonal diagonalizability (from the spectral theorem). However, L is never invertible since $L\mathbf{1} = \mathbf{0}$. If the underlying network has k connected components, L can be permuted into a matrix with k block diagonal components representing these subnetworks. A similar decomposition is possible for the normalized Laplacian. A famous theorem of Kirchoff shows that the determinant of any $n-1$ minor of L is the number of spanning trees in the network.

3.3 Eigenalues of the Laplacian

The most interesting aspect of the graph Laplacian is that the eigenvalues of the operator have deep connections to many of the standard graph invariants. We mention just a few of the most common properties here, but the book "Spectral Graph Theory" by Fan Chung collects many more applications, although most of these properties tend to be of greater mathematical than practical interest. A gentler introduction to these topics is provided in Brualdi's book.

1. Since L is positive definite all eigenvalues are non–negative.
2. The multiplicity of zero as an eigenvalue of the Laplacian is the number of connected components in the network.
3. The next smallest eigenvalue is known as the Fiedler value or algebraic connectivity of the network.
4. The edge connectivity of the graph is always greater than or equal to the Fiedler value. Additionally, the number of edges between a set of nodes U and its complement is at least the product of the Fiedler value and $\frac{|U||V\backslash U|}{n}$.
5. The Fiedler value is bounded below by the inverse of the number of edges in the graph and the diameter of the graph.
6. The Fiedler value is also a measure of how fast the diffusion process across the network occurs.

7. The synchronization of the network is defined to be the quotient of the largest eigenvalue of the Laplacian and the Fiedler value. This is also a measure of the rate of diffusion across the network.

How can you use the graph Laplacian to determine the number of connected components of the network? Can you identify the groups of nodes in each component?

The graph Laplacian is defined as $D-A$. It is clear that zero is an eigenvalue of this matrix since all of the row sums are zero. Thus, $L\mathbf{1} = \mathbf{0}$. The number of connected components is the number of zero eigenvalues of L. This can be seen from the fact that the restriction of $\mathbf{1}$ to the subspace spanned by the nodes in each connected component is annihilated by L and that this collection of vectors is linearly independent. To see that there are no more vectors with the property we can proceed by induction using the fact that if a component is connected then the eigenvalue zero has multiplicity one. This fact can be seen by observing that if v is a corresponding eigenvector, then $0 = v^t L v = \sum_{i\sim j}(v_i - v_j)^2$ so $v_i = v_j$ since the graph is connected.

The components can be discovered from an arbitrary basis of the null space of L by identifying the components in each vector that have the same values, since we know that the vectors must be expressible as linear combinations of the characteristic vectors on the connected components. The underlying idea is the the matrices can be rearranged to be block diagonal representations of the connected components[4].

Derive the graph Laplacian and the solution for the corresponding diffusion problem.

The standard graph Laplacian can be realized in several separate ways that highlight different aspects of its usefulness. The various normalized versions of the Laplacian add even further complexity to this operator. Algebraically, it can be constructed as BB^T where B is the incidence matrix of the network, defined as a $n \times m$ matrix with each column representing an edge and signs (± 1) assigned to the columns arbitrarily. This construction gives the Laplacian many of its algebraic properties, such as symmetry and positive semi–definiteness.

On the other hand, the Laplacian arises quite naturally in the context of studying diffusion on graphs, as well as in the clustering case discussed in Problem 2, where the Laplacian arose as the constraints matrix for the relaxed cut problems. For diffusion, we consider a vector of values representing quantities on the nodes that spread along edges in the network proportionally to the difference between the values at the incident edges. This is a discretization of the continuous heat flow model that is solved by the standard Laplacian[5]. Letting φ represent our vector function of interest, this gives the following system of differential equations:

$$\frac{d\varphi_i}{dt} = -c\sum_j A_{i,j}(\varphi_i - \varphi_j)$$

Where c is the proportionality constant and the minus sign is for historical (in)convenience.

Distributing this expression, we obtain

$$\frac{d\varphi_i}{dt} = -c\varphi_i \deg(\mathrm{i}) + c\sum_j A_{i,j}\varphi_j = -c\sum_j (\delta_i(j)\deg(i) - A_{i,j})\varphi_j$$

Rewriting this expression in matrix form for the entire vector at once gives

$$\frac{d\varphi}{dt} = -k(D-A)\varphi = -cL\varphi,$$

which is exactly the form that we wanted. To solve this linear differential equation, we note that since L is symmetric it is orthogonally diagonalizable so we can write $\varphi(0) = \sum_{k=1}^n c_k v_k$, where the v_k are eigenvectors of L. Then, substituting into our matrix expression we have:

$$\begin{aligned} 0 = &\ \tfrac{d\sum_{k=1}^n c_k v_k}{dt} + cL\sum_{k=1}^n c_k v_k \\ = &\ \sum_{k=1}^n \tfrac{dc_k v_k}{dt} + cc_k\lambda_k v_k \end{aligned}$$

[4] at least for A and L.

[5] hence the name

Since the v_k are linearly independent, this implies that we have $\frac{dc_i}{dt} + c\lambda_i c_i = 0$ for all i. The solution to each of these linear equations is $c_i(t) = c_i(0)e^{-c\lambda_i t}$. We proved above that L is positive semi–definite, so the λ_i are all non–negative and hence the final solution for φ converges to a steady vector in the limit, determined by coefficients of the components of the kernel of L. Furthermore, on each component of the graph the values of φ converge to the average of the original values on that component since that is the limit of the projection onto the kernel, which is $\mathbf{1}$ times the projection onto each component.

A more general case can be considered if we allow forcing terms to act on our nodes. In this case we are allowing for the existence of constant sources or sinks across the network. This modifies the equations derived above by transforming the linear system to an affine one. We can still make use of the orthogonality of the eigenvectors of L to obtain componentwise relations $c_i(t) = c_i(0)e^{-c\lambda_i t} + \frac{\gamma_i}{\lambda_i t}$ where the γ_i is the i^{th} coordinate of the forcing vector in the eigenvector coordinates.

3.4 Epidemic Models

One of the main uses of network dynamics is modeling the flow of some quantity (information or disease) across a network. These models have much in common with both the Markov and diffusion models. The simplest model is the SI model which is simply a reinterpretation of the standard diffusion model with an initial distribution consisting of point masses. Generally in this model the outcomes at each node are assumed to be binary or probabilistic. The most commonly used model is the SIR model (susceptible, infected, recovered) in which each node in the network is labelled with one of the properties at each stage and a Markov process probabilistically updates the labels. The main concerns of this model are the basic reproduction number and the threshold number, which represent the spreading rate of the disease and the rate at which the disease is expected to die out respectively. Many tweaks of this model are possible, each leading to a different (longer) acronym.

Another type of model is a contagion (or information) flow model. This is closely related to the diffusive model discussed above. Percolation techniques are also used to study the flow of diseases across networks. In general, for epidemic dynamics, the main concerns are understanding the equilibrium behavior of the model in terms of the proportion of infected individuals as well as the cyclic behaviors associated to the disease. The basic reproduction number is defined to be the expected number of new cases generated by a single infected individual in an entirely healthy population.