

ACCOUNTING FOR TOPOLOGY IN SPREADING CONTAGION IN NON-COMPLETE NETWORKS

June Zhang and José M.F. Moura

Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213 USA
Email: {junez@andrew, moura@ece}.cmu.edu

ABSTRACT

We are interested in investigating the spread of contagion in a network, G , which describes the interactions between the agents in the system. The topology of this network is often neglected due to the assumption that each agent is connected with every other agents; this means that the network topology is a complete graph. While this allows for certain simplifications in the analysis, we fail to gain insight on the diffusion process for non-complete network topology. In this paper, we offer a continuous-time Markov chain infection model that explicitly accounts for the network topology, be it complete or non-complete. Although we characterize our process using parameters from epidemiology, our approach can be applied to many application domains. We will show how to generate the infinitesimal matrix that describes the evolution of this process for any topology. We also develop a general methodology to solve for the equilibrium distribution by considering symmetries in G . Our results show that network topologies have dramatic effect on the spread of infections.

Index Terms— continuous-time Markov chain, network, stationary distribution, isomorphism, infection

1. INTRODUCTION

In the study of infection in epidemiology, trends in social networks, cascading effects in critical physical infrastructures, the topology of the network underlying the interactions between the agents is generally not accounted for [1, 2]. The implicit assumption is that this network is complete (i.e., every agent is in contact with or impacts the behavior of every other agent). This is an understandable approximation since for a complete network, under appropriate scaling conditions and in the asymptotic limit of large networks (mean field limit), the evolution of the diffusion process reduces to the study of the dynamics of a scalar statistics (e.g., fraction of infected agents in the network) [3, 4].

In this paper, we develop a methodology for studying diffusion in a network of N agents that explicitly takes into account the network topology. We model the diffusion process by a continuous-time Markov chain whose states keep track of the states of all the nodes. We refer to the state of a single agent as the micro state and the state of all the agents as the network state. The state space of our Markov model is the set of all possible network states.

First, we will show how to automatically find the infinitesimal matrix, Q , of this process for arbitrary topology based on the

characteristics of the infection process; we account for 1) healing, 2) naturally occurring infection, 3) transmitted infection from infected neighbors. Next, we look to graphs theory and isomorphism to define an equivalence relationship between the states of the Markov process. This reduces the cardinality of the state space of the Markov process to the cardinality of the state space of the equivalence classes. In particular, we show how this equivalence relationship reduces the computation complexity of finding the stationary distribution of our Markov infection model. Finally, we illustrate our approach on non-complete network topologies, namely, with the chain and cycle graph.

2. THE MODEL

2.1. Network of Agents

We represent the network of N agents by an undirected, connected, unweighted colored graph $G = (V, E, P)$ with adjacency matrix A , which is $N \times N$. The set E defines the topology, and P partitions the vertices into 2 sets: infected (1) and uninfected (0). $P = \{V_0, V_1 \mid V_0 \cup V_1 = V, V_0 \cap V_1 = \emptyset\}$ [5]. For $v \in V$, $\text{color}(v) = i$ if $v \in V_i$.

We assume that the topology remains static. The network state, \mathbf{n} , is the N -tuple of the micro states

$$\mathbf{n} = (\text{color}(v_0), \text{color}(v_1), \dots, \text{color}(v_{N-1}))$$

The i th element in \mathbf{n} is n_i , the micro state which represents the state of agent i . The set, \mathcal{N} , contains all possible network states, \mathbf{n} . Therefore, the cardinality of \mathcal{N} is 2^N .

Each network state, $\mathbf{n} \in \mathcal{N}$, induces a corresponding colored graph $G_{\mathbf{n}} = (V, E, P^{\mathbf{n}})$. $P^{\mathbf{n}} = \{V_0^{\mathbf{n}}, V_1^{\mathbf{n}}\}$ where $\{V_0^{\mathbf{n}} = i \mid n_i = 0\}$, $\{V_1^{\mathbf{n}} = i \mid n_i = 1\}$. Figure 1 shows some example network states and their induced colored graphs for a chain graph with 4 nodes. Gray nodes are infected nodes; white nodes are uninfected nodes.

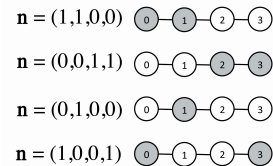


Fig. 1: Examples of Network States and Induced Colored Graphs

2.2. Markov Infection Model

We will model the evolution of spreading contagion in G using concepts from the well-known SIS (Susceptible-Infected-Susceptible) model [3]. Let $\{X(t)\}, (t \geq 0)$ be the network state at time t . At some time t , $X(t) = \mathbf{n}$, where $\mathbf{n} \in \mathcal{N}$. The process evolves based on 3 types of events:

1. Healing. Infected nodes heal in a length of time that is exponentially distributed with parameter μ .
2. Exogenous infection. Uninfected nodes may naturally become infected in a length of time that is exponentially distributed with parameter λ (assuming that there is only one type of virus).
3. Endogenous infection. Uninfected nodes become infected by transmission from infected neighbors in a length of time that is exponentially distributed with parameter γ (assuming that all infected individuals are equally contagious).

We can normalize by μ so that the process is parameterized by $\frac{\lambda}{\mu}$ and $\frac{\gamma}{\mu}$.

With these assumptions, we model $\{X(t)\}$ as a finite state, continuous-time Markov chain with state space \mathcal{N} . Adapting the notation from [6], we define 2 operators on a state of the Markov process, $\mathbf{n} = (n_0, n_1, \dots, n_j, \dots, n_k, \dots, n_{N-1})$

$$T_k \mathbf{n} = (n_0, n_1, \dots, n_k = 1, \dots, n_{N-1})$$

$$T_{j\bullet} \mathbf{n} = (n_0, n_1, \dots, n_j = 0, \dots, n_{N-1})$$

T_k defines the operation that node k is infected. If node k is already infected, the operator does nothing. $T_{j\bullet}$ defines the operation that node j is healed. If node j is already uninfected, the operator does nothing.

There are two types of state transitions in the Markov process:

- 1) $X(t)$ jumps to the network state where the k th node ($k = 0, 1, \dots, N-1$) is infected with transition rate

$$q(\mathbf{n}, T_k \mathbf{n}) = \frac{\lambda}{\mu} + \sum_{j=0}^{N-1} 1(n_j = 1) A_{jk} \frac{\gamma}{\mu}, \quad \mathbf{n} \neq T_k \mathbf{n} \quad (1)$$

where $1(\cdot)$ is the indicator function, and $A = [A_{jk}]$ is the adjacency matrix of G . The first term accounts for exogenous infections, which does not depend on G . The second term accounts for endogenous infections, which is dependent on the topology of the network.

- 2) $X(t)$ jumps to the network state where the j th node ($j = 0, \dots, N-1$) is healed with transition rate:

$$q(\mathbf{n}, T_{j\bullet} \mathbf{n}) = 1, \quad \mathbf{n} \neq T_{j\bullet} \mathbf{n} \quad (2)$$

2.3. Infinitesimal Matrix, \mathbf{Q}

Using equations (1) and (2), we can generate the infinitesimal matrix, \mathbf{Q} , which is a $2^N \times 2^N$ matrix. The i th row and j th column of \mathbf{Q} correspond to the decimal scalar representations of the network states, $\mathbf{i}, \mathbf{j} \in \mathcal{N}$, respectively.

The matrix \mathbf{Q} is not symmetric, but it has symmetric structure, meaning that the nonzero elements are in symmetric locations. Nonzero entries below the diagonal correspond to $X(t)$ transitioning to a state with one less infected individual while nonzero entries above the diagonal correspond to the Markov process jumping to a state with an additional infected node.

We can automatically generate \mathbf{Q} given a network topology G and the infection parameters $\frac{\lambda}{\mu}$ and $\frac{\gamma}{\mu}$. With \mathbf{Q} , we can find the

equilibrium distribution, π by solving $\pi \mathbf{Q} = 0$. For large N , this is computationally intensive. Our next task is to utilize the concept of isomorphism from graph theory to reduce the state space of our Markov process.

3. GRAPH ISOMORPHISM AND EQUIVALENCE CLASSES

Recall that each network state, \mathbf{n} , induces a colored graph $G^n = (V, E, P^n)$. From graph theory, two graphs are considered equivalent if they are isomorphic to each other [5].

Two colored graphs, $G = (V, E, P)$ and $G' = (V', E', P')$ are isomorphic, $G \sim G'$, if there is a mapping $\phi : V \rightarrow V'$ such that

1. ϕ is a bijection
2. $(v_1, v_2) \in E$ if and only if $(\phi(v_1), \phi(v_2)) \in E'$ for all $v_1, v_2 \in V$
3. $\text{color}(v) = \text{color}(\phi(v))$ for all $v \in V$

The first 2 conditions ensure that the uncolored graphs, (V, E) and (V', E') are isomorphic. The last conditional is the additional constraint required by colored graph isomorphism.

3.1. Quotient Set of \mathcal{N}

We define the equivalence class of the network state, \mathbf{n} , as

$$[\mathbf{n}] = \{\mathbf{x} \in \mathcal{N} \mid G^{\mathbf{x}} \sim G^{\mathbf{n}}\}$$

where \sim is colored graph isomorphism. The set of equivalence classes in \mathcal{N} is called the quotient set and is denoted as $\mathcal{N}/\sim = \{[\mathbf{n}_1], [\mathbf{n}_2], \dots, [\mathbf{n}_{eq}]\}$, where $eq = |\mathcal{N}/\sim|$. The cardinality of \mathcal{N}/\sim will depend on the topology of G . However we know that $|\mathcal{N}/\sim| \leq 2^N$ since $[\mathbf{n}_i] \cap [\mathbf{n}_j] = \emptyset$ for $i \neq j$ and $\mathcal{N} = [\mathbf{n}_1] \cup [\mathbf{n}_2] \cup \dots \cup [\mathbf{n}_{eq}]$.

Proposition 1: If $\mathbf{x}, \mathbf{y} \in [\mathbf{n}]$, then $\pi(\mathbf{x}) = \pi(\mathbf{y})$. We will prove this in the Appendix.

3.2. Finding the Quotient Set, \mathcal{N}/\sim

With the given network, $G = (V, E, P)$, we can find the quotient set using the following steps:

1. Initialize: $\mathcal{N}/\sim = \emptyset$
2. Find the set of mappings $\{\phi\}$ that gives the set of uncolored graphs $\{G'\}$ that is isomorphic to (V, E) . This can be done efficiently with existing algorithm [7]
3. While $(\mathcal{N} \neq \emptyset)$
 - (a) Pick $\mathbf{n} \in \mathcal{N}$. It has a corresponding colored graph $G^n = (V, E, P^n)$ where $P^n = \{V_0^n, V_1^n\}$
 - (b) For each mapping in $\{\phi\}$ and each $v \in V_1^n$, set $\text{color}(\phi(v)) = 1$ to produce colored graphs that are isomorphic to G^n . These colored graphs correspond to a set of network states, $\{\mathbf{m}\}$
 - (c) Add \mathbf{n} to \mathcal{N}/\sim
 - (d) Remove $\{\mathbf{n}, \{\mathbf{m}\}\}$ from \mathcal{N}

3.3. Equilibrium Distribution over \mathcal{N}/\sim

Since all the network states in the same equivalence class have identical equilibrium distribution, we can reduce the state space of the Markov process from \mathcal{N} to \mathcal{N}/\sim . The new infinitesimal matrix \mathbf{Q}_{eq} is a $eq \times eq$ matrix. We can find the unnormalized stationary distribution by solving $\pi_{eq} \mathbf{Q}_{eq} = 0$.

We normalize π_{eq} using the relationship

$$|[\mathbf{n}_1]|(\pi_{eq}(\mathbf{n}_1)) + \dots + |[\mathbf{n}_{eq}]|(\pi_{eq}(\mathbf{n}_{eq})) = 1$$

Depending on G , considering isomorphism can lead to great reduction in the number of network states. Figure 2 shows the log-scaling of $|\mathcal{N}/\sim|$ for different network topologies. For example, in a network where $N = 15$, $|\mathcal{N}| = 2^{15} = 32,768$. However, if the network is a chain structure like in Figure 3a), then $|\mathcal{N}/\sim| = 16,512$. A cycle graph, as shown in Figure 3b), contains even more isomorphism mappings and $|\mathcal{N}/\sim| = 1224$. Naturally, the graph with the most symmetry is the complete graph where $|\mathcal{N}/\sim| = N + 1 = 16$. For the complete graph, equivalence class is determined by the number of infected nodes. This is not true for non-complete graphs.

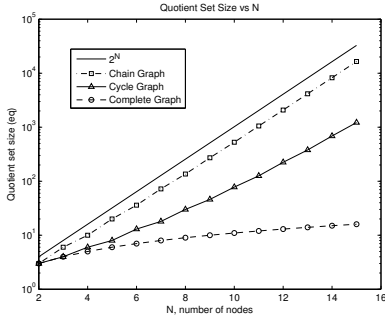


Fig. 2: $\log(|\mathcal{N}/\sim|)$ vs. $N = 2$ to 15 for Different Graph Topologies

4. RESULTS

We illustrate our model with 2 examples with parameters ($N = 4$, $\frac{\lambda}{\mu} = 0.1$, $\frac{\gamma}{\mu} = 3$) for two network topologies: a chain graph, G_1 , and a cycle graph, G_2 , as shown in Figure 3a) and Figure 3b) respectively.

Figure 3 shows the quotient sets of G_1 and G_2 . By considering equivalence classes, we can reduce the size of the state space of G_1 from 16 to 10 and of G_2 from 16 to 6. Table 1 summarizes the equilibrium distribution of G_1 and G_2 .

The evolution of the infection on the chain and cycle graph exhibit different limiting behavior. Both models have a higher rate of endogenous infection to healing, so we expect many nodes to become infected. However, the cycle topology has a higher probability of all nodes being infected than the chain node; since more nodes are connected together in the cycle graph, it is easier for the virus to spread amongst the nodes.

We can also see that the limiting probability is not determined simply by the number of infected nodes. In G_1 , it is less likely for node v_0 and v_2 to be infected simultaneously than for a single node to be infected. However, in the cycle graph, the probability of any

two nodes being infected is higher than the probability of a single infected node. The equilibrium distribution is determined by the spatial arrangement of infected and non infected nodes. For example, in the cycle graph it is more likely for connected nodes like v_0 and v_1 to be infected at the same time than for unconnected nodes such as v_0 and v_3 to be infected simultaneously. There is a predilection for infected nodes to be clustered and connected together, thereby giving rise to ‘epidemic’ behavior.

Table 1: Equilibrium distribution for chain and cycle graph ($N = 4$, $\frac{\lambda}{\mu} = 0.1$, $\frac{\gamma}{\mu} = 3$)

Network State \mathbf{n}	$[\mathbf{n}]_{G_1}$	$\pi_{G_1}(\mathbf{n})$	$[\mathbf{n}]_{G_2}$	$\pi_{G_2}(\mathbf{n})$
0000	$[n_1]$	0.1882	$[n_1]$	0.0968
1000	$[n_2]$	0.0213	$[n_2]$	0.0097
0100	$[n_3]$	0.0163	$[n_2]$	0.0097
1100	$[n_4]$	0.0498	$[n_3]$	0.0252
0010	$[n_3]$	0.0163	$[n_2]$	0.0097
1010	$[n_5]$	0.0130	$[n_4]$	0.0105
0110	$[n_6]$	0.0375	$[n_3]$	0.0252
1110	$[n_7]$	0.1031	$[n_5]$	0.0735
0001	$[n_2]$	0.0213	$[n_2]$	0.0097
1001	$[n_8]$	0.0100	$[n_3]$	0.0252
0101	$[n_5]$	0.0130	$[n_4]$	0.0105
1101	$[n_9]$	0.0391	$[n_5]$	0.0735
0011	$[n_4]$	0.0498	$[n_3]$	0.0252
1011	$[n_9]$	0.0391	$[n_5]$	0.0735
0111	$[n_7]$	0.1031	$[n_5]$	0.0735
1111	$[n_{10}]$	0.2790	$[n_6]$	0.4484

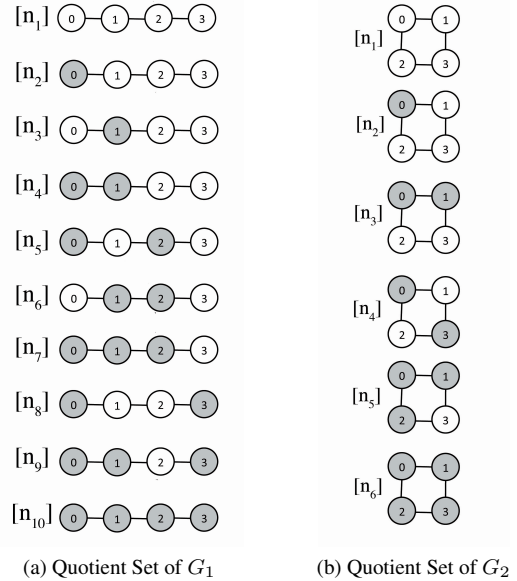


Fig. 3: Colored Graph Equivalence Classes for G_1 and G_2

5. CONCLUSION

Using our model, we have shown that the diffusion process is highly dependent on the network topology. Assuming a complete network topology will give very different results from any non-complete topology. We find the infinitesimal matrix that describes the evolution of the disease, which accounts for infection and healing of the

nodes in any topology. We use the concept of graph isomorphism to reduce the state space of the Markov process to expedite the calculation for the stationary distribution.

6. REFERENCES

- [1] I. Dobson, B.A. Carreras, and D.E. Newman, "A branching process approximation to cascading load-dependent system failure," *Hawaii International Conferences on System Sciences*, Jan. 2004.
- [2] N. Boccara, *Modeling complex systems*, Springer Verlag, 2010.
- [3] M. O. Jackson, *Social and Economic Networks*, Princeton University Press, 2008.
- [4] A. Santos and J.M.F. Moura, "Emergent behavior in large scale networks," *Conference on Decision and Control*, Dec. 2011.
- [5] A. Lal and D. van Melkebeek, "Graph isomorphism for colored graphs with color multiplicity bounded by 3," Tech. Rep., Citeseer, 2005.
- [6] F.P. Kelly, *Reversibility and stochastic networks*, vol. 40, Wiley New York, 1979.
- [7] Gabor Csardi and Tamas Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, pp. 1695, 2006.
- [8] G. Royle C. Godsil, *Algebraic Graph Theory*, Springer-Verlag, 2001.
- [9] S. Karlin H. M. Taylor, *An Introduction to Stochastic Modeling*, Academic Press, 3rd edition, 1998.

A. PROOF OF PROPOSITION 1

If $\mathbf{x}, \bar{\mathbf{x}} \in [\mathbf{n}]$, then $\pi(\mathbf{x}) = \pi(\bar{\mathbf{x}})$.

\mathbf{x} and $\bar{\mathbf{x}}$ are in the same equivalence class so there is a mapping, ϕ , such that the induced graph $G^{\mathbf{x}} = (V, E, P^{\mathbf{x}})$ is color isomorphic to $G^{\bar{\mathbf{x}}} = (\phi(V), E, P^{\bar{\mathbf{x}}})$. $G^{\mathbf{x}}$ and $G^{\bar{\mathbf{x}}}$ have the same adjacency matrix, A . Using Lemma 1.31 from [8], which states that node $v_i \in G^{\mathbf{x}}, (i = 0, 1, \dots, N-1)$ and node $\phi(v_i) \in G^{\bar{\mathbf{x}}}$ have the same number of neighbors. This means that $\sum_{k=0}^{N-1} A_{ki} = \sum_{k=0}^{N-1} A_{k\phi(i)}$.

The equilibrium distribution, π , satisfies the global balance equation:

$$\pi(\mathbf{x}) \left(\sum_{k=0}^{N-1} q(\mathbf{x}, T_k \mathbf{x}) + \sum_{j=0}^{N-1} q(\mathbf{x}, T_{j\bullet} \mathbf{x}) \right) = \sum_{k=0}^{N-1} \pi(T_k \mathbf{x}) q(T_k \mathbf{x}, \mathbf{x}) + \sum_{j=0}^{N-1} \pi(T_{j\bullet} \mathbf{x}) q(T_{j\bullet} \mathbf{x}, \mathbf{x}) \quad (3)$$

$$\pi(\bar{\mathbf{x}}) \left(\sum_{\phi(k)=0}^{N-1} q(\bar{\mathbf{x}}, T_{\phi(k)} \bar{\mathbf{x}}) + \sum_{\phi(j)=0}^{N-1} q(\bar{\mathbf{x}}, T_{\phi(j)\bullet} \bar{\mathbf{x}}) \right) = \sum_{\phi(k)=0}^{N-1} \pi(T_{\phi(k)} \bar{\mathbf{x}}) q(T_{\phi(k)} \bar{\mathbf{x}}, \bar{\mathbf{x}}) + \sum_{\phi(j)=0}^{N-1} \pi(T_{\phi(j)\bullet} \bar{\mathbf{x}}) q(T_{\phi(j)\bullet} \bar{\mathbf{x}}, \bar{\mathbf{x}}) \quad (4)$$

Note that $T_k \mathbf{x}$ and $T_{\phi(k)} \bar{\mathbf{x}}$ are in the same equivalence class. The T_k operator represents a coloring of the k th node in $G^{\mathbf{x}}$ while $T_{\phi(k)}$ represents a coloring of the $\phi(k)$ th node in $G^{\bar{\mathbf{x}}}$. This coloring, together with the mapping ϕ , satisfy the 3 conditions of color isomorphism. Similarly, $T_{j\bullet} \mathbf{x}$ and $T_{\phi(j)\bullet} \bar{\mathbf{x}}$ are in the same equivalence class.

Consider the transition rates:

$$q(\mathbf{x}, T_{j\bullet} \mathbf{x}) = 1 = q(\bar{\mathbf{x}}, T_{\phi(j)\bullet} \bar{\mathbf{x}})$$

$$\begin{aligned} q(\mathbf{x}, T_k \mathbf{x}) &= \frac{\lambda}{\mu} + \sum_{j=0}^{N-1} 1(n_j = 1) A_{jk} \frac{\gamma}{\mu} \\ &= \frac{\lambda}{\mu} + \sum_{j=0}^{N-1} 1(n_j = 1) A_{j\phi(k)} \frac{\gamma}{\mu} \\ &= q(\bar{\mathbf{x}}, T_{\phi(k)} \bar{\mathbf{x}}) \end{aligned}$$

By the same reasoning,

$$\begin{aligned} q(T_k \mathbf{x}, \mathbf{x}) &= q(T_{\phi(k)} \bar{\mathbf{x}}, \bar{\mathbf{x}}) \\ q(T_{j\bullet} \mathbf{x}, \mathbf{x}) &= q(T_{\phi(j)\bullet} \bar{\mathbf{x}}, \bar{\mathbf{x}}) \end{aligned}$$

Equations (3) and (4) represents two linear systems with identical coefficients. Since finite state Markov process always has a unique equilibrium distribution, the solution to the linear systems must be unique as well. Therefore, $\pi(\mathbf{x}) = \pi(\bar{\mathbf{x}})$ when $\mathbf{x}, \bar{\mathbf{x}} \in [\mathbf{n}]$.

A.1. Finite State Continuous-Time Markov Process Review

We model the diffusion process by a finite state continuous-time Markov chain, $\{X(t)\} (t \geq 0)$ with states $\{0, 1, \dots, M\}$. We will only consider an aperiodic, irreducible, time homogenous, finite state Markov process [9]. The process is characterized by

$$\begin{aligned} q(i, j) &= \lim_{\tau \rightarrow 0} \frac{P_{ij}(\tau)}{\tau}, \quad j \neq i \\ &= 0, \quad j = i \end{aligned}$$

where $P_{ij}(t)$ is the transition probability of going from state i to state j in t duration. It is also the $\{i, j\}$ entry in the matrix $\mathbf{P}(t)$. Have

$$q(i) = \sum_{j=0, j \neq i}^M q(i, j)$$

The rates, $q(i)$ and $q(i, j)$, provide us with the infinitesimal description of the Markov process. The amount of time that the process will remain in state i is exponentially distributed with parameter $q(i)$. We can express the transition rates as the infinitesimal matrix, \mathbf{Q} , where

$$\mathbf{Q} = \begin{bmatrix} -q(0) & q(0, 1) & \dots & q(0, M) \\ q(1, 0) & -q(1) & \dots & q(1, M) \\ \vdots & & & \\ q(M, 0) & q(M, 1) & \dots & -q(M) \end{bmatrix}$$

The differential equation that governs the evolution of $\mathbf{P}(t)$ is

$$\mathbf{P}'(t) = \mathbf{P}(t) \mathbf{Q}$$

We can find π by solving the system of equations

$$0 = \pi \mathbf{Q} = [\pi(0) \quad \pi(1) \quad \dots \quad \pi(M)] \mathbf{Q}$$

with the additional constraint that $\sum_{i=0}^M \pi(i) = 1$.