Markov Chains with Maximum Entropy for Robotic Surveillance

Mishel George, Saber Jafarpour, *Member, IEEE,* Francesco Bullo, *Fellow, IEEE*

***Abstract*—This article provides a comprehensive analysis of the following optimization problem: maximize the entropy rate generated by a Markov chain over a connected graph of order n and subject to a prescribed stationary distribution. First, we show that this problem is strictly convex with global optimum lying in the interior of the feasible space. Second, using Lagrange multipliers we provide a closed-form expression for the maxentropic Markov chain as a function of an n-dimensional vector, referred to as the maxentropic vector; we provide a provably-converging iteration to compute this vector. Third, we show that the maxentropic Markov chain is reversible, compute its entropy rate, and describe special cases, among other results. Fourth, through analysis and simulations, we show that our proposed procedure is more computationally efficient than semidefinite programming methods. Finally, we apply these results to robotic surveillance problems. We show realizations of the maxentropic Markov chains over prototypical robotic roadmaps and find that maxentropic Markov chains outperform minimum mean hitting time Markov chains for so-called "intelligent intruders" with short attack durations. a comprehensive analysis of the following optimization problem: maximize the entropy rate generated by a Markov chain over a connected graph of order n and subject to a prescribed stationary distribution.**

***Index Terms*—Markov chain, stochastic surveillance, convex optimization, entropy rate**

I. Introduction

a) Problem description: The entropy rate of a Markov chain is a measure of information and unpredictability generated with each time-step [9]. In this paper, we study Markov chains with maximal entropy generation subject to two constraints: (i) allowable transitions are specified by a given irreducible adjacency matrix and (ii) the stationary distribution of the Markov chain is given. It is customary to refer to Markov chains with maximum entropy rate as *maxentropic*. Maxentropic Markov chains with stationary distribution constraints are of interest in surveillance strategies as they maximize the uncertainty in the path of the surveillance agent. Aside from applications to stochastic surveillance, the notion of maxentropic Markov chains is useful for example in link-prediction [18], community detection [20] and image processing [30].

b) Prior work and applications of maxentropic Markov chains: To the best of our knowledge, maxentropic Markov chains first appeared in [14] as the solution to the optimization problem of maximizing the entropy rate given the first and second moments of the Markov chain. More recently, Burda et al. [7] provide a closed form solution for maxentropic Markov chains subject solely to graph constraints. This Markov chain, referred to as the *maximal entropy random walk* (MERW), possesses the property that all walks of equal length with given start and end node are equiprobable. The solution we provide is for Markov chains subject to stationary distribution constraints in addition to graph constraints. In what follows, we discuss three applications of maxentropic Markov chains: (i) design of stochastic surveillance strategies, (ii) detection of features in images, and (iii) design of metrics on large graphs and complex networks.

Stochastic surveillance. Minimum hitting time Markov chains have been used in the design of stochastic surveillance strategies in [21] where a novel convex program formulation of the problem is considered. The notion of group hitting time for multiple random walkers is used in optimizing transition matrices for multiple agents in [22]. Furthermore, Markov chains have been used in conjunction with specific notions of intelligent intruders to design stochastic strategies [4]. The authors in [11], [26] use Markov chain Monte Carlo methods to design surveillance strategies. In [2] the mean hitting time in conjunction with multiple parallel instances of the CUSUM algorithm is used to devise a policy which ensures quickest average time to detection of anomalies. Finally, the work in [8] formulates an efficient algorithm based on Markov chains named PATROLGRAPH* which allows for effective extension to the multi-agent case.

Image analysis. Based of the notion of maximal entropy random walks in [7], several applications have been proposed in image analysis. The MERW is utilized instead of the equal neighbor random walk to detect visually salient features in [30]. The MERW has also been utilized to implement a probabilistic object localization scheme in [29]. Korus and Huang [16] successfully adopt the MERW for localizing forgeries in digital images.

Metrics on large networks. The MERW is used to design unsupervised methods for link prediction in [18]. Ochab and Burda study the feasibility of using the MERW in algorithms for community detection [20]. Furthermore, the MERW is used to study the trapping problem in dendrimers, i.e., artificial macromolecules with treelike structures [23]. More recently, a relation between entropy rate and congestion in complex networks was established and a method was proposed to mitigate congestion using MERW in [10].

c) Statement of contributions: This article makes contribution to Markov chain theory as well as to robotic surveil-

This work has been supported in part by Air Force Office of Scientific Research award FA9550-15-1-0138.

Mishel George, Saber Jafarpour, and Francesco Bullo are with the Mechanical Engineering Department and the Center of Control, Dynamical-Systems and Computation, UC Santa Barbara, CA 93106-5070, USA. {mishel,saber.jafarpour,bullo}@engineering.ucsb.edu

lance. First, we show that the novel problem of entropy rate maximization subject to graph with n nodes and visit frequency constraints is well-defined and is strictly convex. We show that the unique global solution is indeed an irreducible Markov chain. The irreducibility property implies that the solution has a well-defined stationary distribution identical to that posed in the stationary distribution constraint.

Second, as the main contribution of the paper, we provide an iterative algorithm with rigorous convergence guarantees to compute an n-dimensional vector, called the so-called *maxentropic vector*. In turn, as a function of this maxentropic vector, we provide a closed-form formula for the maximum entropy rate Markov chain, referred to as the *maxentropic Markov chain with visit frequency constraints*. In other words, we compute maxentropic chains with arbitrary stationary distributions on a graph with n nodes using an n-dimensional vector instead of optimizing transition matrices in $\mathbb{R}^{n\times n}$.

Third, we establish various additional results, including (i) the reversibility of maxentropic Markov chains with prescribed stationary distributions, (ii) a formula for the maximum entropy rate subject to the constraints, and (iii) an equiprobable path property, which, prior to this work, was only known to hold for the maximal entropy random walk. Additionally, for a few special choices of the constraints, we are able to characterize interesting special cases. For example, we show that the equal neighbor random walk on a graph is equal to the maxentropic Markov chain with visit frequency at each node proportional to the degree of the node.

Fourth, we conduct a careful comparison between our proposed procedure and standard SDP methods across a range of graph topologies. Specifically, we conduct a worst-case complexity analysis of our procedure and compare it with interior point methods used to solve semidefinite programming formulations of the entropy rate maximization problem. Empirically and analytically, we show that our proposed procedure has significantly lower runtime than an SDP method to solve the optimization problem.

Finally, we demonstrate some example realizations of these maxentropic chains in robotic scenarios. A key simulation-based result is that maxentropic Markov chains perform better than minimum hitting time Markov chains for the important case of so-called intelligent waiting intruders with short attack durations. We also conduct simulations on a partitioned graph with multiple surveillance agents and find that this result appears to hold for the multi-agent case as well.

d) Relevance to stochastic surveillance: The setup we consider is one in which the area to be surveilled has been sampled to obtain a robotic roadmap represented by a graph. The nodes of the graph designate points of high priority and the edges indicate whether it is possible to move between different nodes. (Restrictions might be imposed by obstacles, no-fly zones, etc.) The graph structure is captured by a binary adjacency matrix and the relative importance of each node is given by a normalized vector which indicates a desired visit frequency to each node. Markov chains modeled by transition matrices are well suited to designing random walks on graphs with visit frequency constraints. The left-dominant eigenvector of the transition matrix, referred to as the stationary distribution, gives the visit frequency of a random walker who moves according to the Markov chain. Graph and stationary distribution constraints are linear and hence can be enforced quite effectively in optimization problems involving cost functions with various robotic motivations such as maximizing speed of traversal, minimizing the expected reward for an intruder or convergence to a desired swarm formation [21], [4], [5].

While prior work with the same framework has emphasized the speed of the Markov chain or optimizing the probability of capture given an intruder model, the transition matrices obtained as solutions to such formulations need not necessarily be unpredictable (e.g., permutation matrices, which have zero entropy rate, are the fastest Markov chains when a Hamiltonian tour exists). The notion of maximum entropy rate Markov chains is valuable as it translates directly to maximum unpredictability in the path of the surveillance agent. The specification of a stationary distribution, which serves as a prior for where the intruder might be located, makes our approach more suited than the MERW which has a fixed stationary distribution.

e) Applications in other areas: The methods described in this paper are potentially useful for developing novel methods of conducting image analysis. The maxentropic Markov chain with visit frequency specification provides a natural way of incorporating prior knowledge of where an object or an anomaly is likely to be located within an image and hence can be used in place of the MERW in [29], [30] when such knowledge is available. Further the fact that the method described in this article scales well with the graph dimension enables its use for analysis of large images.

Finally, the methods developed in this paper could also aid in the design of novel metrics for complex networks. In refs. [16], [18], [20], novel metrics are designed by using the MERW to evaluate properties of the network. With the ability to specify visit frequencies in the random walk it becomes possible to evaluate some of these metrics in a weighted sense. For example, if one specifies that visit frequencies of the random walk be a function of the degree of each node, the metric thus obtained will incorporate such a weighting.

f) Paper organization: This paper is organized as follows. In section II we introduce notation and review known results. In section III we derive some preliminary results. In section IV we introduce the main result of this paper which is the maxentropic chain with prescribed visit frequencies. In section V we show realizations of the maxentropic Markov chain over prototypical roadmaps. Finally, in section VI we present conclusions.

II. Notation and review of known results

A. Notation

For $x \in \mathbb{R}^n$, let $\|x\|_1 = \sum_{i=1}^n |x_i|$, let $[x]$ denote the diagonal matrix with diagonal entries x, that is,

$$[x] = \begin{bmatrix} x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_n \end{bmatrix}.$$

A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if, for all partitions $\{I, J\}$ of the index set $\{1, \dots, n\}$, there exists $i \in I$ and $j \in J$ such that $a_{ij} \neq 0$. Here $\{I, J\}$ is a partition of the index set if $I \cup J = \{1, \dots, n\}$ and $I \cap J = \phi$.

Given $x, y \in \mathbb{R}^n$, we define the component-wise vector product $x \circ y \in \mathbb{R}^n$ by $(x \circ y)_i = x_i y_i$ for $i \in \{1, \dots, n\}$. We note the simple equalities:

$$[x]y = x \circ y = [y]x \quad \text{and} \quad [[x]y] = [x][y] = [x \circ y]. \tag{1}$$

Define the set of positive n-tuples by $\mathbb{R}^n_{>0} = \{x \in \mathbb{R}^n \mid x_i > 0, i \in \{1, \dots, n\}\}$ and the probability simplex of order n by $\Delta_n = \{v \in \mathbb{R}^n \mid \sum_{i=1}^n v_i = 1, v_i \geq 0 \text{ for } i \in \{1, \dots, n\}\}$.

Consider a graph G with nodeset $V = \{1, \dots, n\}$ then a walk from node i_1 to i_2 and so on until node i_k for $\{i_1, i_2, \dots, i_k\} \in V$ is denoted as $i_1 \to i_2 \to \dots \to i_k$.

The following lemma and its proof are included here for completeness.

Lemma 1. *Let $S \subseteq \mathbb{R}^n$ be a compact convex set, $\|.\|$ be a matrix norm on $\mathbb{R}^n$, and $h : S \to S$ be a continuously differentiable map. If $\|\partial h/\partial x(x)\| < 1$ for all $x \in S$, then h is a contraction mapping with respect to the norm $\|.\|$ and has a unique fixed point in S.*

Proof. Because S is compact and h is C^1, there exists $c \in (0, 1)$ such that

$$\|\partial h/\partial x(x)\| \leq c, \qquad \text{for all } x \in S.$$

By the Mean Value Inequality [1, Proposition 2.4.8], for every $x, y \in S$, there exists $\eta \in S$ such that

$$\|h(y) - h(x)\| \leq \|\partial h/\partial x(\eta)\|\|y - x\|.$$

Therefore, for every $x, y \in S$, we know

$$\|h(y) - h(x)\| \leq c\|y - x\|.$$

Since $0 < c < 1$, this inequality shows that $h : S \to S$ is a contraction with respect to the norm $\|\cdot\|$. By the Banach Contraction Theorem [17, Theorem 3.4.1], h has a unique fixed point in S. □

B. Review of maxentropic Markov chains

Throughout the paper we model the transition matrix of a Markov chain as a row-stochastic matrix. Given a Markov chain with an irreducible transition matrix $P \in \mathbb{R}^{n\times n}$ (i.e., an irreducible row-stochastic matrix), the entropy rate of the Markov chain is given by

$$\mathcal{H}(P) = -\sum_{i,j=1}^{n} \pi_i(P) p_{ij} \log p_{ij}, \tag{2}$$

where $\pi(P) \in \text{interior}(\Delta_n)$ is the stationary distribution of P (whose existence, uniqueness, and positivity are established by the Perron-Frobenius Theorem for irreducible matrices).

Problem 1 (Maximizing entropy rate)**.** *Given a connected undirected unweighted graph G, compute the matrix $P \in \mathbb{R}^{n\times n}$ satisfying*

$$\begin{aligned} \max \quad & \mathcal{H}(P) \\ \textit{subj. to} \quad & P \textit{ is row stochastic, i.e., } P \geq 0 \textit{ and } P\mathbb{1}_n = \mathbb{1}_n \\ & p_{ij} = 0, \textit{ if } \{i, j\} \textit{ is not an edge of } G. \end{aligned}$$

Theorem 1 (The maxentropic Markov chain [7], [14])**.** *Given a symmetric, irreducible $A \in \{0, 1\}^{n\times n}$ with associated undirected graph G, let $\lambda > 0$ and $v \in \mathbb{R}^n_{>0}$ be the dominant eigenvalue and eigenvector of A (whose existence and uniqueness are established by the Perron-Frobenius Theorem).*

Then the solution to Problem 1 is unique, is called maxentropic Markov chain *over G, and is given by*

$$P^* = \frac{1}{\lambda}[v]^{-1}A[v], \tag{3}$$

or, in components, by

$$P^*_{ij} = \frac{a_{ij}}{\lambda}\frac{v_j}{v_i}.$$

Moreover, P^* has the following properties:

(i) *its stationary distribution is $v \circ v/\|v \circ v\|_1$,*
(ii) *its paths are* equiprobable *in the following sense: pick a start node i and a path length k. The probability of traversal for a path from i of length $k \geq 1$ is*

$$\frac{1}{\lambda^k}\frac{v_j}{v_i}, \tag{4}$$

where j is the final node in the path. Note that all paths from i to j of length k have the same probability.

The following example illustrates a maxentropic chain.

Example 1. *Consider the adjacency matrix associated with a 4-node ring and the maxentropic Markov chain associated with this graph,*

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, P^* = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{bmatrix}.$$

In general one can show that the Markov chain that maximizes entropy on the ring is the transition matrix that randomizes the position of the random walker at the subsequent timestep between its current location and the two adjacent nodes on the ring.

III. Maxentropic maps and their properties

In this section we introduce and characterize two maps: the *maxentropic matrix map* and *maxentropic vector map*. These maps shall be used in the construction of Markov chains with maximum entropy subject to graph and stationary distribution constraints.

A. The maxentropic matrix map and its properties

Given a symmetric, irreducible, binary matrix $A \in \{0, 1\}^{n\times n}$, define the *maxentropic matrix map* $\Phi_A : \mathbb{R}^n_{>0} \to \mathbb{R}^{n\times n}_{\geq 0}$ by

$$\Phi_A(x) = [Ax]^{-1}A[x], \tag{5}$$

or, in components, by

$$\big(\Phi_A(x)\big)_{ij} = a_{ij}\frac{x_j}{\sum_{k=1}^n a_{ik}x_k}.$$

The maxentropic Markov chain subject to graph and stationary distribution constraints can be generated from the maxentropic matrix map for a suitable choice of x. In the remainder of this section, we only characterize the maxentropic matrix map. The connection to maxentropic Markov chains shall become clear in Section IV.

Theorem 2 (Properties of the maxentropic matrix map)**.** *Given a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ and a vector $x \in \mathbb{R}^n_{>0}$, the maxentropic matrix map has the following properties:*

(i) *$\Phi_A(x)$ is well defined, nonnegative, and row-stochastic,*
(ii) *$\Phi_A(x)$ has the same irreducible zero/positive pattern as A,*
(iii) *the left dominant eigenvector of $\Phi_A(x)$ is*

$$\pi(x) = \frac{1}{\|[x]Ax\|_1}[x]Ax, \tag{6}$$

(iv) *$\Phi_A(x)$ is reversible, i.e., $[\pi(x)]\Phi_A(x) = \Phi_A(x)^\top[\pi(x)]$.*

Proof. First, we know $Ax > 0$ because $x > 0$ and because A being irreducible implies each row of A has at least one positive entry. Hence, the diagonal matrix $[Ax]$ is invertible and $\Phi_A(x)$ is well defined and nonnegative. Finally, $[x]\mathbb{1}_n = x$ implies

$$\Phi_A(x)\mathbb{1}_n = [Ax]^{-1}A[x]\mathbb{1}_n = [Ax]^{-1}Ax = \mathbb{1}_n.$$

This concludes the proof of statement (i).

Next, note that $\Phi_A(x)$ is equal to the matrix A pre- and post-multiplied by two diagonal matrices with positive diagonal; hence $\Phi_A(x)$ has the same zero/positive pattern as A and is irreducible. This concludes the proof of statement (ii).

Regarding statement (iii), by the Perron-Frobenius Theorem for irreducible nonnegative matrices we know that $\Phi_A(x)$ has a unique left dominant eigenvector, i.e., a vector $\pi(x)$ satisfying $\pi(x)^\top\Phi_A(x) = \pi(x)^\top$ and $\mathbb{1}^\top\pi(x) = 1$.

It suffices to show $\pi(x)^\top\Phi_A(x) = \pi(x)^\top$. Recalling the equalities (1), we compute

$$\begin{aligned}\pi(x)^\top\Phi_A(x) &= \frac{1}{\|[x]Ax\|_1}\Big([x]Ax\Big)^\top\Big([Ax]^{-1}A[x]\Big)\\ &= \frac{1}{\|[x]Ax\|_1}\Big([Ax]x\Big)^\top\Big([Ax]^{-1}A[x]\Big)\\ &= \frac{1}{\|[x]Ax\|_1}x^\top[Ax][Ax]^{-1}A[x]\\ &= \frac{1}{\|[x]Ax\|_1}x^\top A[x] = \frac{1}{\|[x]Ax\|_1}\Big([x]Ax\Big)^\top.\end{aligned}$$

This concludes the proof of statement (iii).

Finally, again recalling the equalities (1) and assuming $\|[x]Ax\|_1 = 1$ without loss of generality, we compute

$$\begin{aligned}[\pi(x)]\Phi_A(x) &= [[x]Ax][Ax]^{-1}A[x]\\ &= ([x][Ax])[Ax]^{-1}A[x] = [x]A[x],\\ \Phi_A(x)^\top[\pi(x)] &= [x]A[Ax]^{-1}[[x]Ax] = [x]A[x].\end{aligned}$$

This concludes the proof of statement (iv). ☐

B. The maxentropic vector map and its properties

Next, we study the left dominant eigenvector of the row-stochastic matrix $\Phi_A(x)$. Given a binary, symmetric, irreducible matrix A with unit diagonal entries, define the *maxentropic vector map* $\phi_A : \mathbb{R}^n_{>0} \to \mathbb{R}^n_{>0}$ by

$$\phi_A(x) = [x]Ax,$$

or, in components, by

$$\big(\phi_A(x)\big)_i = x_i\sum_{k=1}^n a_{ik}x_k.$$

In what follows, we use the notion of proper maps to establish that the maxentropic vector map is a global diffeomorphism. A map $h : X \to Y$ is proper if for every compact set $C \subset Y$, the preimage $h^{-1}(C) \subset X$ is compact.

Theorem 3 (Properties of the maxentropic vector map)**.** *Given a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ with unit diagonal entries, the maxentropic vector map ϕ_A has the following properties:*

(i) *the Jacobian of ϕ_A satisfies $\partial\phi_A/\partial x(x) = [x]A + [Ax]$ and is full rank at all $x \in \mathbb{R}^n_{>0}$,*
(ii) *ϕ_A is a proper map, and*
(iii) *ϕ_A is a global diffeomorphism, in particular, for every $\pi \in \mathbb{R}^n_{>0}$, there exists a unique $x^* \in \mathbb{R}^n_{>0}$ such that $\phi_A(x^*) = \pi$.*

Proof. Regarding property (i), clearly ϕ_A is analytic. Elementary calculations based also on the equalities (1) show that $\partial\phi_A/\partial x(x) = [x]A + [Ax]$. One can show

$$\begin{aligned}\frac{\partial(\phi_A)_i}{\partial x_i}(x) &= a_{ii}x_i + \sum_{j=1}^n a_{ij}x_j\\ &> \sum_{j=1,j\neq i}^n a_{ij}x_j = \sum_{j=1,j\neq i}^n \frac{\partial(\phi_A)_i}{\partial x_j}(x),\end{aligned}$$

for all $x \in \mathbb{R}^n_{>0}$, because $a_{ii} = 1 > 0$ for all $i \in \{1,\dots,n\}$. Hence, the Jacobian matrix $\partial\phi_A/\partial x(x)$ is strictly row diagonally dominant and, therefore, invertible for all $x \in \mathbb{R}^n_{>0}$.

Before continuing, it is convenient to define the map $\widehat{\phi_A} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ by $\widehat{\phi_A}(x) = [x]Ax$, so that ϕ_A is the restriction of the map $\widehat{\phi_A}$ to $\mathbb{R}^n_{>0}$. We claim that, for every $S \subseteq \mathbb{R}^n_{>0}$, we have $\phi_A^{-1}(S) = \widehat{\phi_A}^{-1}(S)$. We establish this claim as follows. By the property of the restriction map, we can easily show that $\phi_A^{-1}(S) \subseteq \widehat{\phi_A}^{-1}(S)$. Now suppose that there exists a vector $\mathbf{v} = (v_1, v_2, \dots, v_n)^\top$ such that $\mathbf{v} \in \widehat{\phi_A}^{-1}(S)$ and $\mathbf{v} \notin \phi_A^{-1}(S)$. This implies that $\widehat{\phi_A}(\mathbf{v}) \in S$. Since $\mathbf{v} \notin \phi_A^{-1}(S)$, there exists some $i \in \{1,\dots,n\}$, such that $v_i = 0$. This implies that $([\mathbf{v}]A\mathbf{v})_i = 0$ and therefore we have $\Big(\widehat{\phi_A}(\mathbf{v})\Big)_i = 0$. However, this means that $\widehat{\phi_A}(\mathbf{v}) \notin S$. Which is a contradiction. Therefore, we have $\phi_A^{-1}(S) = \widehat{\phi_A}^{-1}(S)$.

Regarding property (ii), let C be a compact set in $\mathbb{R}^n_{>0}$. Then it is a compact set in $\mathbb{R}^n_{\geq 0}$. Therefore, C is closed in $\mathbb{R}^n_{\geq 0}$. Since $\widehat{\phi_A}$ is a continuous map, $\widehat{\phi_A}^{-1}(C)$ is closed in $\mathbb{R}^n_{\geq 0}$. We show that $\widehat{\phi_A}^{-1}(C)$ is bounded in $\mathbb{R}^n_{\geq 0}$. Since all diagonal elements of A are one, we have the following inequality

$$\|x\|_\infty^2 \leq \|[x]Ax\|_\infty, \qquad \text{for all } x \in \mathbb{R}^n_{\geq 0}.$$

Since C is compact, there exists $M \in \mathbb{R}_{>0}$ such that, for every $y \in C$, we have $\|y\|_\infty < M$. Thus, for every $x \in \widehat{\phi_A}^{-1}(C)$, we have

$$\|x\|_\infty^2 \leq \|[x]Ax\|_\infty = \|\widehat{\phi_A}(x)\|_\infty < M.$$

Therefore, $\widehat{\phi_A}^{-1}(C)$ is bounded in $\mathbb{R}^n_{\geq 0}$. This implies that $\widehat{\phi_A}^{-1}(C)$ is compact in $\mathbb{R}^n_{\geq 0}$. Recall that we established $\phi_A^{-1}(C) = \widehat{\phi_A}^{-1}(C)$. Therefore $\phi_A^{-1}(C)$ is a compact set in $\mathbb{R}^n_{>0}$.

Finally, regarding property (iii), we start by noting that property (i) implies, by the Inverse Function Theorem, that ϕ_A is a local diffeomorphism. Therefore, using property (ii) the map ϕ_A is a proper local diffeomorphism and [1, Theorem 2.5.17] implies that ϕ_A is a global diffeomorphism. □

Remark 1 (The maxentropic vector map is ill-posed without self-loops)**.** *The following example shows that the statements (ii) and (iii) in Theorem 3 do not generally hold for graphs without self-loops. Consider the adjacency matrix*

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Define the vectors, $x = [x_1 \;\; x_2]^\top$ *and* $\pi = [\pi_1 \;\; \pi_2]^\top$*. The maxentropic vector map is given by* $\phi_A(x) = [x_1x_2 \;\; x_1x_2 + x_2^2]^\top$*. One can solve for the inverse of the map* ϕ_A *explicitly in this case obtaining*

$$\phi_A^{-1}(\pi) = \begin{bmatrix} \dfrac{\pi_1}{\sqrt{\pi_2 - \pi_1}} & \sqrt{\pi_2 - \pi_1} \end{bmatrix}.$$

Consider the compact set $\Pi = \{[\pi_1 \;\; \pi_2]^\top \mid \pi_1 + \pi_2 = 1, \; 0.25 \leq \pi_1 \leq 0.5\}$*. The preimage of the set* Π *under the maxentropic vector map* ϕ_A^{-1} *is not bounded, and hence this set is not compact in* $\mathbb{R}^n$*. Also note that the* $\phi_A^{-1}(\pi)$ *is empty when* $\pi_1 > \pi_2$ *and hence the map is not a diffeomorphism.* ⋄

In what follows, we characterize the inverse function of ϕ_A at π. In other words, given a point $\pi \in \operatorname{interior}(\Delta_n)$, we compute $x = \phi_A^{-1}(\pi)$ as the solution to the algebraic equation

$$\phi_A(x) = [x]Ax = \pi. \tag{7}$$

Theorem 4 (Inverse of the maxentropic vector map)**.** *Given a symmetric, irreducible, binary matrix* $A \in \{0,1\}^{n\times n}$ *with unit diagonal entries, pick* $\pi \in \operatorname{interior}(\Delta_n)$*.*

(i) *For* $A = \mathbb{1}_n\mathbb{1}_n^\top$*, the algebraic equation (7) admits the unique solution* π*.*

(ii) *Define the constants* $\eta = \max_i\{\sum_{j=1}^n a_{ij}\sqrt{\pi_j}\}$*,* $\xi = \max_i\{\sum_{j=1}^n a_{ij}\pi_j\}$*, and the vector* $x^0 = \frac{\pi}{\sqrt{\xi}}$*. Then the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *defined by* linear iteration

$$x^{k+1} = x^k - \frac{1}{2\eta}\left([x^k]Ax^k - \pi\right), \qquad \text{for all } k \in \mathbb{N}, \tag{8}$$

converges to the unique solution of equation (7).

Proof. Regarding statement (i), note that $A = \mathbb{1}_n\mathbb{1}_n^\top$ implies $A\pi = \mathbb{1}_n$. Therefore, if we set x equal to π into equation (7), we get

$$[\pi]A\pi = [\pi]\mathbb{1}_n = \pi.$$

Regarding statement (ii), we first define the nonempty compact convex domain

$$\Omega_\pi = \left\{\mathbf{y} \in \mathbb{R}^n_{>0} \;\middle|\; \frac{\pi}{\eta} \leq \mathbf{y} \leq \sqrt{\pi}\right\}.$$

We first show that $x^0 \in \Omega_\pi$. Since A is a binary matrix, for every $i \in \{1,\dots,n\}$, we have

$$\sqrt{\sum_{j=1}^n a_{ij}\pi_j} \leq \sum_{j=1}^n a_{ij}\sqrt{\pi_j}.$$

Therefore, one can deduce that

$$\sqrt{\max_i\left\{\sum_{j=1}^n a_{ij}\pi_j\right\}} \leq \max_i\left\{\sum_{j=1}^n a_{ij}\sqrt{\pi_j}\right\}.$$

Moreover, matrix A has unit diagonal entries so that

$$\sqrt{\pi_i} \leq \sqrt{\sum_{j=1}^n a_{ij}\pi_j} \leq \sqrt{\max_i\left\{\sum_{j=1}^n a_{ij}\pi_j\right\}}.$$

Therefore, we have

$$\frac{\pi}{\eta} \leq \frac{\pi}{\sqrt{\xi}} \leq \sqrt{\pi},$$

so that $x^0 \in \Omega_\pi$. Next, define the map $f_\pi : \Omega_\pi \to \mathbb{R}^n$ by

$$f_\pi(x) = x - \frac{1}{2\eta}\left([x]Ax - \pi\right).$$

We aim to show that Ω_π is invariant under the map f_π, i.e., $f_\pi(\Omega_\pi) \subseteq \Omega_\pi$. Consider a point $x \in \Omega_\pi$. We have

$$\frac{\pi_i}{\eta} \leq x_i \leq \sqrt{\pi_i}, \qquad \text{for all } i \in \{1,\dots,n\}.$$

Therefore, for every $i \in \{1,\dots,n\}$, we compute

$$\begin{aligned}
(f_\pi(x))_i &= x_i - \frac{1}{2\eta}\left(([x]Ax)_i - \pi_i\right) \\
&= x_i - \frac{1}{2\eta}x_i^2 - \frac{1}{2\eta}\left(\sum_{j=1, i\neq j}^n a_{ij}x_j\right) + \frac{1}{2\eta}\pi_i \\
&\leq x_i - \frac{1}{2\eta}x_i^2 + \frac{1}{2\eta}\pi_i.
\end{aligned}$$

Note that, for every $i \in \{1,\dots,n\}$, we have $x_i \leq \sqrt{\pi_i} < \eta$. This implies that the maximum of the function $x_i - \frac{1}{2\eta}x_i^2$ is $\sqrt{\pi_i} - \frac{1}{2\eta}\pi_i$. Hence, we have

$$(f_\pi(x))_i \leq x_i - \frac{1}{2\eta}x_i^2 + \frac{1}{2\eta}\pi_i \leq \sqrt{\pi_i} - \frac{1}{2\eta}\pi_i + \frac{1}{2\eta}\pi_i = \sqrt{\pi_i}.$$

On the other hand, for every $i \in \{1,\dots,n\}$, we have

$$\sum_{i=1}^n a_{ij}x_j \leq \sum_{i=1}^n a_{ij}\sqrt{\pi_j} \leq \eta$$

Therefore, for every $i \in \{1,\dots,n\}$, we have

$$\begin{aligned}
(f_\pi(x))_i &= x_i - \frac{1}{2\eta}([x]Ax)_i + \frac{1}{2\eta}\pi_i \\
&= x_i\left(1 - \frac{1}{2\eta}\sum_{j=1}^n a_{ij}x_j\right) + \frac{1}{2\eta}\pi_i \\
&\geq \frac{\pi_i}{\eta}\left(1 - \frac{\eta}{2\eta}\right) + \frac{1}{2\eta}\pi_i \geq \frac{\pi_i}{\eta}.
\end{aligned}$$

This shows that $f_\pi(x) \in \Omega_\pi$ and therefore Ω_π is an invariant set for the map f_π. Next, we show that the map f_π is a contraction mapping on Ω_π. The derivative of f_π satisfies

$$\frac{\partial f_\pi}{\partial x}(x) = I_n - \frac{1}{2\eta}\left([x]A + [Ax]\right), \qquad \text{for all } x \in \Omega_\pi.$$

Also, we have

$$\begin{aligned}\Big\|\frac{\partial f_\pi}{\partial x}(x)\Big\|_1 &\\ &= \Big\| I_n - \frac{1}{2\eta}\left([x]A + [Ax]\right)\Big\|_1 \\ &= \max_i \left\{ \left| 1 - \frac{a_{ii}x_i}{2\eta} - \sum_{j=1}^n \frac{a_{ij}x_j}{2\eta}\right| + \left| \sum_{j=1, j\neq i}^n \frac{a_{ij}x_j}{2\eta} \right| \right\}.\end{aligned}$$

Since $x \in \Omega_\pi$ implies $x \leq \sqrt{\pi}$, we deduce that, for every $i \in \{1,\dots,n\}$, we have $a_{ii}x_i + \sum_{j=1}^n a_{ij}x_j \leq 2\eta$. This implies that, for every $i \in \{1,2,\dots,n\}$, we have

$$\left| 1 - \frac{a_{ii}x_i}{2\eta} - \sum_{j=1}^n \frac{a_{ij}x_j}{2\eta}\right| = 1 - \frac{a_{ii}x_i}{2\eta} - \sum_{j=1}^n \frac{a_{ij}x_j}{2\eta}.$$

Thus, for every $i \in \{1,2,\dots,n\}$, we get

$$\left| 1 - \frac{a_{ii}x_i}{2\eta} - \sum_{j=1}^n \frac{a_{ij}x_j}{2\eta}\right| + \left| \sum_{j=1, j\neq i}^n \frac{a_{ij}x_j}{2\eta} \right| = 1 - \frac{a_{ii}x_i}{\eta}.$$

Therefore, we obtain

$$\Big\|\frac{\partial f_\pi}{\partial x}(x)\Big\|_1 = \max_i \left\{ \left| 1 - \frac{a_{ii}x_i}{\eta} \right| \right\} < 1.$$

Now, using Lemma 1, the map f_π has a unique fixed point in the domain Ω_π and, for $x^0 = \frac{\pi}{\sqrt{\xi}} \in \Omega_\pi$, the sequence defined by the linear iteration (8) converges to this unique fixed-point. The proof of the theorem is complete if one notes that x^* is the unique fixed point of f_π if and only if x^* is the unique solution to the algebraic equation (7). □

Remark 2 (Solution to the maxentropic vector map on the complete graph)**.** *If $A = \mathbb{1}_n\mathbb{1}_n^\top$, then we have $\eta = 1$ and the initial condition $x^0 = \frac{\pi}{\sqrt{\eta}} = \pi$ in statement (ii) is the fixed-point of the linear iteration* (8) *and the unique solution to the algebraic equation* (7). ◇

Remark 3 (Newton-Raphson iteration)**.** *The Newton-Raphson iteration for the nonlinear equation $\phi_A(x) = \pi$ is*

$$x^{k+1} = x^k - ([x^k]A + [Ax^k])^{-1}([x^k]Ax^k - \pi). \tag{9}$$

In simulations, this iteration appears to always converge for a wide variety of graphs, from random initial conditions, and for arbitrary choices of $\pi \in \text{interior}(\Delta_n)$ — even if we are unable to provide a convergence proof. We postpone to Section IV-C a runtime comparison between the linear iteration (8) *and this Newton-Raphson iteration* (9). ◇

IV. Maxentropic Markov chains with prescribed stationary distributions

In this section, we define the optimization problem whose solution we characterize. We then prove uniqueness and existence of the solution before we introduce the main result of the paper which is a closed-form expression for the maxentropic Markov chain at given stationary distribution, following which we perform computational comparisons with standard convex program solvers and provide proofs for the main result.

A. Problem statement

Recall that the solution to Problem 1, i.e., the maximum entropy problem subject to purely graph constraints, is the Markov chain given by equation (3) in Theorem 1. In what follows, we introduce a new optimization problem by imposing additional stationary distribution constraints on Problem 1. Before we state the problem definition, we remind the reader that given $\pi \in \text{interior}(\Delta_n)$ and given a Markov chain with an irreducible transition matrix $P \in \mathbb{R}^{n\times n}$ (i.e., an irreducible row-stochastic matrix), the entropy rate of the Markov chain P at fixed π is given by

$$\mathcal{H}_\pi(P) = -\sum_{i,j=1}^n \pi_i p_{ij} \log p_{ij}. \tag{10}$$

Problem 2 (Maximizing entropy rate with a stationary distribution constraint)**.** *Given a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ with unit diagonal entries and given a positive vector $\pi \in \text{interior}(\Delta_n)$, compute the transition matrix $P \in \mathbb{R}^{n\times n}$ satisfying*

$$\begin{aligned} \max \quad & \mathcal{H}_\pi(P) && (11)\\ \textit{subj. to} \quad & P \geq 0, && (12)\\ & p_{ij} = 0, \ \textit{if } a_{ij} = 0, && (13)\\ & P\mathbb{1}_n = \mathbb{1}_n, && (14)\\ & \pi^\top P = \pi^\top. && (15)\end{aligned}$$

Remark 4. *Problem 2 is a disciplined convex program and hence the numerical solution of this program can be computed in CVX [12].* ◇

Remark 5 (Problem 2 is ill-posed without self-loops)**.** *For given graph topologies without self-loops and for many corresponding instances of stationary distributions, CVX returns that Problem 2 is infeasible. For all such cases, we find that the linear iteration in equation* (8) *diverges (recall Remark 1). For example, consider once again the adjacency matrix*

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

CVX returns that the program is infeasible for any stationary distribution constraint (15)*, $[\pi_1 \ \pi_2]^\top P = [\pi_1 \ \pi_2]^\top$, in which $\pi_1 > \pi_2$. Additionally, for this setting the linear iteration* (8) *diverges.*

Indeed, the graph topology embodied by A dictates that, whenever the surveillance agent visits node 1, then the agent visits node 2 in the subsequent timestep. Hence, the visit frequency at node 2 is necessarily greater than or equal to the visit frequency at node 1. ◇

Several such preliminary results indicate that imposing graph constraints can restrict the set of stationary distribution achieved by Markov chains and inspired the conjecture in Section VI. We do not pursue this potentially interesting direction of research in this paper as within the framework of designing surveillance strategies, self-loops can be naturally incorporated. Hence, we proceed under the assumption that all nodes have self-loops.

Remark 6. *In the absence of self-loops in G, the set of irreducible Markov chains over G with prescribed stationary distribution might be empty; see [15] and the conjecture in Section VI for additional context.*

Before we introduce the main result, we prove that the optimizer is irreducible. The optimizer being irreducible ensures existence of the solution to Problem 2 as only irreducible stochastic matrices have well-defined stationary distributions. In addition, we also prove that the optimizer assigns a positive transition probability to every edge in the graph, which is a property that shall be utilized in the proof of the main result.

Theorem 5 (Maxentropic Markov chains are well defined)**.** *Given a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ with unit diagonal entries and given a positive vector $\pi \in \text{interior}(\Delta_n)$, Problem 2 satisfies the following properties:*

(i) *the cost function is strictly concave and the constraint set is compact and convex. Hence, its global maximum solution P^* exists and is unique;*

(ii) *the optimizer P^* satisfies $p^*_{ij} > 0$ whenever $\{i,j\}$ is an edge of the graph G associated to A. Hence, P^* is irreducible and has a well-defined stationary distribution that must be equal to π.*

Because of statement (ii), we refer to P^* as the maxentropic Markov chain over G with stationary distribution π.

Note that, for a symmetric, irreducible $A \in \{0,1\}^{n\times n}$ with unit diagonal entries, the graph associated to A is undirected, unweighted, and connected and has self-loops at each node.

Proof of Theorem 5. Regarding statement (i), the function $-p\log(p)$ is strictly concave with a strictly positive second-derivative for $p > 0$. The entropy rate is a linear combination of strictly concave functions and hence $\mathcal{H}(P)$ is strictly concave.

Regarding statement (ii), we first show, using a contradiction, that the diagonal entries of P^* can not be zero. Assume that exactly one of the diagonal elements of P^* is zero, i.e., there exists a single $k \in \{1,\dots,n\}$ such that $p^*_{kk} = 0$. We try to find a contradiction. For every $0 < \epsilon < \frac{1}{2}$, define the matrix-valued function $\tilde{P}^*(\epsilon) = (1-\epsilon)P^* + \epsilon I_n$. Note that, for every $\epsilon \in (0,\frac{1}{2})$, we have $\tilde{P}^*(\epsilon) \geq 0$. Also, for every $\epsilon \in (0,\frac{1}{2})$, we have $\tilde{p}^*_{ij} \geq 0$ if $\{i,j\}$ is an edge of G and $\tilde{p}^*_{ij} = 0$ otherwise. One can check that $\tilde{P}^*(\epsilon)\mathbb{1}_n = \mathbb{1}_n$ and $\pi^\top \tilde{P}^*(\epsilon) = \pi^\top$. These facts imply that, for every $\epsilon \in (0,\frac{1}{2})$, the matrix $\tilde{P}^*(\epsilon)$ is in the feasible set of Problem 2. By the Mean Value Theorem [24, Theorem 5.10], for every $\epsilon \in (0,\frac{1}{2})$, there exists $c_\epsilon \in (0,\epsilon)$ such that

$$\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*) = \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \epsilon}\bigg|_{c_\epsilon} \epsilon.$$

Note that, for $i,j \in \{1,\dots,n\}$, we have

$$\frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ij}} = \pi_i(\log((1-\epsilon)p^*_{ij}) + 1), \qquad \forall i \neq j,$$
$$\frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ii}} = \pi_i(\log((1-\epsilon)p^*_{ii} + \epsilon) + 1).$$

Using the chain rule, we compute

$$\begin{aligned}\frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \epsilon} &= \sum_{i=1}^n\sum_{j=1}^n \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ij}}\frac{\partial \tilde{p}^*_{ij}}{\partial \epsilon} \\ &= \sum_{i=1}^n \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ii}}\frac{\partial \tilde{p}^*_{ii}}{\partial \epsilon} + \sum_{i=1}^n\sum_{j=1,j\neq i}^n \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ij}}\frac{\partial \tilde{p}^*_{ij}}{\partial \epsilon} \\ &= -\sum_{i=1}^n \pi_i(\log((1-\epsilon)p^*_{ii}+\epsilon)+1)(1-p^*_{ii}) \\ &\quad + \sum_{i=1}^n\sum_{j=1,j\neq i}^n \pi_i(\log((1-\epsilon)p^*_{ij})+1)(p^*_{ij})\end{aligned}$$

Using the fact that $p^*_{kk} = 0$, we get

$$\begin{aligned}\frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \epsilon} &= -\pi_k(\log(\epsilon)+1) \\ &\quad - \sum_{i=1,i\neq k}^n \pi_i(\log((1-\epsilon)p^*_{ii}+\epsilon)+1)(1-p^*_{ii}) \\ &\quad + \sum_{i=1}^n\sum_{j=1,j\neq i}^n \pi_i(\log((1-\epsilon)p^*_{ij})+1)(p^*_{ij}).\end{aligned}$$

Hence we obtain

$$\begin{aligned}\frac{1}{\epsilon}\Big(\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*)\Big) &= -\pi_k(\log(c_\epsilon)+1) \\ - \sum_{i=1,i\neq k}^n \pi_i\big(\log((1-c_\epsilon)p^*_{ii}+c_\epsilon)+1\big)(1-p^*_{ii}) \\ + \sum_{i=1}^n\sum_{j=1,j\neq i}^n \pi_i(\log((1-c_\epsilon)p^*_{ij})+1)p^*_{ij}.\end{aligned}$$

Since $c_\epsilon \in (0,\frac{1}{2})$ and $p^*_{ii} \neq 0$, for every $i \neq k$, the term $\sum_{k\neq i}\pi_i(\log((1-c_\epsilon)p^*_{ii}+1)+1)(1-p^*_{ii})$ is bounded. Similarly, since for every $i \neq j$, $p^*_{ij} \neq 0$, the term $\sum\sum_{i\neq j}\pi_i(\log((1-c_\epsilon)p^*_{ij})+1)p^*_{ij}$ is bounded. Thus, by choosing ϵ small enough, one can make c_ϵ small enough and, therefore, the term $-\pi_k(\log(c_\epsilon)+1)$ large enough. Thus, there exists $\epsilon^* \in (0,\frac{1}{2})$ such that

$$\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*) > 0, \qquad \text{for all } \epsilon \in (0,\epsilon^*].$$

This is a contradiction, since we assumed that P^* is the solution to Problem 2. It is straightforward to generalize this argument to the case when we have more zeros on the diagonal of P^* Therefore, all the diagonal entries of P^* are strictly positive.

Next, assuming that all diagonal elements of P^* are positive, we show that, for every i,j with $i \neq j$, if $a_{ij} > 0$, then $p^*_{ij} > 0$. Assume that there exists exactly one pair (k,l) such that $a_{kl} > 0$ but $p^*_{kl} = 0$. We try to find a contradiction. Define the matrix $\Gamma \in \mathbb{R}^{n\times n}$ with all zero entries except for:

$$\Gamma_{kk} = -1, \quad \Gamma_{kl} = 1, \quad \Gamma_{lk} = \frac{\pi_k}{\pi_l}, \quad \text{and } \Gamma_{ll} = -\frac{\pi_k}{\pi_l}.$$

Define $\eta = \min\{\frac{1}{1+p^*_{kk}}, \frac{\pi_l}{\pi_l p^*_{ll}+\pi_k}\}$. For every $\epsilon \in [0, \eta)$, we define the matrix function $\tilde{P}^*(\epsilon) = (1-\epsilon)P^* + \epsilon\Gamma$. One can show that, for every $\epsilon \in (0, \eta)$, we have $\tilde{P}^*(\epsilon) \geq 0$. Moreover, for every $\epsilon \in (0, \eta)$, we have $\tilde{p}^*_{ij} \geq 0$ if $\{i, j\}$ is an edge of G and $\tilde{p}^*_{ij} = 0$ otherwise. One can check that $\tilde{P}^*(\epsilon)\mathbb{1}_n = \mathbb{1}_n$ and $\pi^\top \tilde{P}^*(\epsilon) = \pi^\top$. This implies that, for every $\epsilon \in (0, \eta)$, the matrix $\tilde{P}^*(\epsilon)$ is in the feasible set of Problem 2. By the Mean Value Theorem [24, Theorem 5.10], for every $\epsilon \in (0, \eta)$, there exists $c_\epsilon \in (0, \epsilon)$ such that

$$\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*) = \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \epsilon}\bigg|_{c_\epsilon} \epsilon.$$

Using the chain rule, we compute

$$\begin{aligned}
\frac{1}{\epsilon}\Big(\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*)\Big) &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \mathcal{H}_\pi(\tilde{P}^*)}{\partial \tilde{p}^*_{ij}} \frac{\partial \tilde{p}^*_{ij}}{\partial \epsilon}\bigg|_{c_\epsilon} \\
= -\pi_k(\log(c_\epsilon)+1) &+ \pi_k(\log((1-c_\epsilon)p^*_{kk} - c_\epsilon) + 1)(1+p^*_{kk}) \\
&- \pi_l(\log((1-c_\epsilon)p^*_{lk} + \frac{\pi_k}{\pi_l}c_\epsilon) + 1)\left(\frac{\pi_k}{\pi_l} - p^*_{lk}\right) \\
&+ \pi_l(\log((1-c_\epsilon)p^*_{ll} - \frac{\pi_k}{\pi_l}c_\epsilon) + 1)\left(p^*_{ll} + \frac{\pi_k}{\pi_l}\right) \\
&+ \sum_{i\notin\{k,l\}} \sum_{j\notin\{k,l\}} \pi_i(\log((1-c_\epsilon)p^*_{ij}) + 1)p^*_{ij}. \qquad (16)
\end{aligned}$$

Note $p^*_{ij} = 0$ if and only if $(i, j) = (k, l)$. Since $c_\epsilon \in (0, \eta)$, by choosing ϵ small enough, one can make $-\pi_k(\log(c_\epsilon) + 1)$ large enough while the remaining terms in the right hand side of (16) are bounded. Therefore, there exists $\epsilon^* \in (0, \eta)$ such that

$$\mathcal{H}_\pi(\tilde{P}^*(\epsilon)) - \mathcal{H}_\pi(P^*) > 0, \qquad \text{for all } \epsilon \in (0, \epsilon^*].$$

This contradicts the fact that P^* is the solution to Problem 2. The generalization of this argument to the case where we have more zeros in P^* is straightforward. Hence, $p^*_{ij} = 0$ if and only if $\{i, j\}$ is not an edge of the graph G. □

B. Main result

Having motivated the problem of finding the maximum entropy Markov chain subject to graph and stationary distribution constraints and having obtained some preliminary results, we finally present the solution to Problem 2.

Theorem 6 (Maxentropic Markov chains with prescribed stationary distribution)**.** *Consider a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ with unit diagonal entries and a positive vector $\pi \in \operatorname{interior}(\Delta_n)$. Let $x = \phi_A^{-1}(\pi)$ denote the solution to $[x]Ax = \pi$ (whose existence, uniqueness, positivity, and computation algorithm are given in Theorems 3 and 4).*

Then the maxentropic Markov chain over G with stationary distribution π is

$$P^* = \Phi_A(\phi_A^{-1}(\pi)) = [Ax]^{-1}A[x]. \qquad (17)$$

Moreover, P^* is reversible and its entropy rate is

$$\mathcal{H}(P^*) = -2x^\top A[x]\log(x) + \pi^\top \log(\pi). \qquad (18)$$

We postpone the proof of this theorem to Section IV-D.

Remark 7. *Theorem 6 implies the following result: if G has self-loops at each node, then, for all $\pi \in \operatorname{interior}(\Delta_n)$, there exists at least one Markov chain over G with stationary distribution π.* ⋄

We provide a corollary describing notable choices of the maxentropic vector in Theorem 6.

Corollary 1 (Remarkable special cases)**.** *Given a symmetric, irreducible, binary matrix $A \in \{0,1\}^{n\times n}$ with unit diagonal entries, let $d = A\mathbb{1}_n$ and $D = [A\mathbb{1}_n]$ denote its degree vector and matrix, and let v and λ denote its dominant eigenvector and eigenvalue. Then*

(i) *the maxentropic Markov chain with stationary distribution $(\mathbb{1}_n^\top d)^{-1}d$ is*

$$P^* = \Phi_A(\mathbb{1}_n) = [A\mathbb{1}_n]^{-1}A,$$

with entropy rate

$$\mathcal{H}(P^*) = (\mathbb{1}_n^\top d)^{-1}d^\top \log(d);$$

(This is the so-called equal neighbor random walk.)

(ii) *the maxentropic Markov chain with stationary distribution $v \circ v/\|v \circ v\|_1$ is*

$$P^* = \Phi_A(v) = \frac{1}{\lambda}[v]^{-1}A[v],$$

with entropy rate

$$\mathcal{H}(P^*) = \log \lambda;$$

(This is the maxentropic Markov chain characterized in Theorem 1 as the solution to Problem 1.)

(iii) *if $A = \mathbb{1}_n\mathbb{1}_n^\top$ and π is arbitrary, then the maxentropic Markov chain over the complete graph with stationary distribution π is*

$$P^* = \Phi_A(\pi) = \mathbb{1}_n\pi^\top,$$

with entropy rate

$$\mathcal{H}(P^*) = -\pi^\top \log(\pi).$$

(The maxentropic vector for the complete graph is shown to be π in Theorem 4.)

Finally, we present an interesting property associated with maxentropic Markov chains with prescribed stationary distribution, that is an extension of Theorem 1(ii).

Lemma 2 (All allowed permutations of a walk are equiprobable)**.** *Under the same assumptions as in Theorem 6, consider a start node i and a final node j on the graph G for which there exists a path $i, l_1, l_2, \dots, l_k, j$ and a path $i, \sigma(l_1), \sigma(l_2), \dots, \sigma(l_k), j$ for a permutation σ. The following equiprobable path traversal property holds for maxentropic Markov chains:*

$$\begin{aligned}
&\mathbb{P}[i \to l_1 \to l_2 \to \dots \to l_k \to j] \\
&= \mathbb{P}[i \to \sigma(l_1) \to \sigma(l_2) \to \dots \to \sigma(l_k) \to j].
\end{aligned}$$

We postpone the proof of this lemma to Section IV-D.

C. Computational complexity

In this subsection we show how our proposed procedure to compute maxentropic chains with prescribed stationary distributions is useful not only to reveal their structure and properties but also serves as a valuable method in terms of reducing computational complexity. In short our claim is that:

> *To compute maxentropic Markov chains (as in Problem 2), the linear iteration (8) and equality (17) (as stated in Theorem 6) are in general computationally faster than general-purpose convex program solvers.*

We establish this claim in two ways. First, we consider a variety of graphs, we fix a given tolerance, and we observe empirically that that our proposed method has significantly smaller runtimes than the standard CVX solver, see Table I.

Graph	**Linear iteration (8) & equality (17)**	**Newton-Raphson iteration (9)**	**CVX**
Line	0.02s	0.01s	44.24s
Star	0.23s	0.01s	54.97s
Ring	0.02s	0.01s	37.53s
Lattice	0.01s	0.02s	40.42s
Complete*	0.01s	0.03s	575.58s

TABLE I: Average runtimes of various methods over 100 runs on standard graph topologies with 100 nodes to compute maxentropic Markov chains with a randomly chosen stationary distribution for each run. Tolerance is fixed as 10^{-8} in all cases. Computations were performed on a 2.9GHz processor using MATLAB.
*We delete one edge from the complete graph as the iteration in Theorem 4 starts with the solution to the complete graph.

Second, we analyze the computational complexity of the competing algorithms in their two parts: the cost per iteration, and the number of iterations required to get to within a specific tolerance of the optimal solution. In what follows we analyze each algorithm and report the results in Table II.

Method	**Cost per iteration**	**No. of iterations**
Linear iteration (8) & equality (17)	$O(n) - O(n^2)$	$O(1)$-$O(n)$
Newton-Raphson iteration (9)	$O(n^3)$	$O(1)^*$
CVX	$O(n^3)$-$O(n^6)$	$O(\sqrt{n})$

TABLE II: Computational complexity of various method to compute maxentropic Markov chains with given stationary distribution.
* We can prove this bound for sparse graphs and in simulations the bound holds for complete graphs.

For the linear iteration (8) in Theorem 4, each iteration consists of only matrix multiplications with the adjacency matrix or a diagonal matrix, whose cost per iteration is $O(n)$ when the adjacency matrix A is sparse and $O(n^2)$ when A is dense. Also, a careful study of the Banach Fixed Point Theorem and the estimates in Theorem 4 shows that the number of iterations for a fixed tolerance depends on the maximum degree of nodes in the graph. In particular, it can be shown that for sparse graphs such as ring graphs and lattice graph, where the maximum degree does not change with the size of the graph, the number of iterations is $O(1)$. However, for star graphs and dense graphs such as the complete graph, the number of iterations is of order $O(n)$. In short, the effective worst case complexity across graph topologies for a fixed tolerance is $O(n^3)$.

For the Newton–Raphson iteration (9), the factorization of the Jacobian at each step leads to $O(n^3)$ number of operations for each iteration. It can be shown that the number of iterations necessary to obtain a solution within a fixed tolerance is $O(1)$ for sparse graphs. In simulations it is observed that the number of iterations is only weakly dependent on the problem size even for dense graphs and is essentially a constant. We are unable to provide an effective worst-case analysis for the number of iterations necessary when the topology is dense, but across different graph topologies it appears safe to assume that the worst case complexity for fixed tolerance is $O(n^3)$.

In general using a convex program solver would be computationally more expensive as the search space for the convex program is $\mathbb{R}^m_{\geq 0}$, where $m = n^2 - (2n-1) - n_e$, where n_e is the number of edge constraints. Note that the stationary and stochastic constraints in equations (14) and (15) effectively sum up to $2n-1$ constraints (it can be shown that one of the constraints is redundant). When the graph is sparse $m = O(n)$, otherwise $m = O(n^2)$. Interior point methods used by convex program solvers would need to compute the factorization of an $O(m) \times O(m)$ matrix at every iterations resulting in a runtime complexity for each iteration of $O(n^3)$ for sparse graphs and $O(n^6)$ when the graph is dense (see refs. [19], [27]). The worst-case dependence on problem size is $O(\sqrt{n})$ [28]. Even assuming a constant dependence on problem size as is observed in practice in most semidefinite program interior point solvers, the effective worst-case runtime complexity for fixed tolerance is $O(n^3)$ for sparse graphs and $O(n^6)$ for dense graphs. Also, note that CVX uses a successive approximation scheme to approximate exponential and logarithmic functions [12, Section 11.3]. While this does not affect the computational complexity of the procedure, there are no theoretical guarantees for convergence to the optimal solution for such an approximation.

Although this article presents numerical comparisons only with general purpose convex program solvers, we note that convex programs with linear constraints can be solved efficiently using first-order methods such as mirror descent [6]. Such methods have the same worst-case computational complexity as our proposed linear iteration (8). We expect our linear iteration to have lower constant factors than first-order methods for dense graphs for the following reason: as our linear iteration operates on an n-dimensional manifold whereas any first-order convex programming method operates on the space of transition probabilities which is $O(n^2)$ for the case of dense graphs.

D. Proofs

Consider a Markov chain with transition matrix P on a graph G with binary adjacency matrix A. Let the random variable Y_t denote the observed transition on the graph G at time t which can assume values on $\{1, \ldots, m\}$, where $m = \sum_i \sum_j a_{ij}$ is the total number of edges in the graph. If P is an irreducible Markov chain with stationary distribution π, then for very large times t the probability that a transition from node i to node j occurs is given by

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}[Y_t=\{i,j\}] &= \lim_{t\to\infty}\left(\mathbb{P}[X_{t+1}=j\,|\,X_t=i]\mathbb{P}[X_t=i]\right)\\ &= \mathbb{P}[X_{t+1}=j\,|\,X_t=i]\lim_{t\to\infty}\mathbb{P}[X_t=i]\\ &= \pi_i p_{ij}.\end{aligned} \tag{19}$$

This calculation motivates the following definition.

Definition 1. *For an irreducible transition matrix P with stationary distribution π, define the* ergodic flow matrix *by*

$$Q=[\pi]P. \tag{20}$$

Remark 8. *The ergodic flow matrix Q is symmetric if and only if the associated Markov chain P is reversible.* ◇

Let q_{ij} denote the entries of Q. Note that q_{ij} is the probability associated with observing a transition along an edge (i,j) at very large times t according to the calculation in equation (19) and the sum of this probability over all edges is 1. The entropy associated with this random variable is

$$\mathcal{H}(Q)=-\sum_{i,j=1}^{n} q_{ij}\log(q_{ij}). \tag{21}$$

Lemma 3 (Relation between entropy rate and entropy of ergodic flow matrix)**.** *For an irreducible transition matrix P with stationary distribution π,*

$$\mathcal{H}_\pi(P)=\mathcal{H}(Q)-\mathcal{H}(\pi), \tag{22}$$

where $\mathcal{H}(\pi)=\sum_i^n \pi_i\log(\pi_i)$.

Proof. The entropy rate of an irreducible Markov chain P with a stationary distribution π is given by

$$\begin{aligned}\mathcal{H}_\pi(P) &= -\sum_{i,j=1}\pi_i p_{ij}\log p_{ij}\\ &= -\sum_{i,j=1}^{n}\pi_i p_{ij}(\log(\pi_i p_{ij})-\log(\pi_i))\\ &= -\sum_{i,j=1}^{n} q_{ij}\log(q_{ij})+\sum_{i}^{n}\pi_i\log(\pi_i)\\ &= \mathcal{H}(Q)-\mathcal{H}(\pi).\end{aligned}$$

□

Consider the convex program which maximizes the entropy of the random variable associated with the ergodic flow matrix.

Problem 3 (Maximize entropy of ergodic flow with a stationary distribution constraint)**.** *Given a connected undirected unweighted graph G and a positive vector $\pi\in$ interior(Δ_n), compute the ergodic flow matrix $Q\in\mathbb{R}^{n\times n}$ satisfying*

$$\begin{aligned}\max\quad & \mathcal{H}(Q) && (23)\\ \textit{subj. to}\quad & Q\geq 0, && (24)\\ & q_{ij}=0,\ \textit{if } \{i,j\} \textit{ is not an edge of } G, && (25)\\ & Q\mathbb{1}_n=\pi, && (26)\\ & Q^\top\mathbb{1}_n=\pi. && (27)\end{aligned}$$

Note that the matrix Q is well-defined only when its associated transition matrix has a stationary distribution π. Hence an optimization algorithm might encounter instances where $q_{ij}=0$ when $a_{ij}=1$ and hence its associated transition matrix is possibly reducible. In such a case the matrix Q might not have the correct interpretation as the ergodic flow matrix associated with its transition matrix P. However, as a result of Theorem 5 we are guaranteed that the optimal solution Q^* will have the appropriate interpretation as the ergodic flow matrix associated with its transition matrix P^*. Further, since the ergodic flow matrix and its associated transition matrix are closely related we have the following result.

Lemma 4 (Equivalence of Problem 2 and Problem 3)**.** *Given a stationary distribution π, Problem 2 is equivalent to Problem 3 in the following sense:*

(i) *if p^*_{ij} is the optimal solution to Problem 2, then $q^*_{ij}=\pi_i p^*_{ij}$ is the optimal solution to Problem 3, and*
(ii) *if q^*_{ij} is the optimal solution to Problem 3, then $p^*_{ij}=q^*_{ij}/\pi_i$ is the optimal solution to Problem 2.*

Proof. First, we shall show that the constraints (12)-(15) in Problem 2 are equivalent to constraints (24)-(27) in Problem 3. Note that equation (20) and the fact that $\pi\in$ interior(Δ_n) implies that Q has the same zero/positive pattern as P. Hence constraints (12), (13) are equivalent to constraints (24), (25) respectively. Note that

$$P\mathbb{1}_n=\mathbb{1}_n \quad\Longrightarrow\; [\pi]P\mathbb{1}_n=[\pi]\mathbb{1}_n \quad\Longrightarrow\; Q\mathbb{1}_n=\pi.$$

Hence constraint (14) is equivalent to constraint (26). Also constraint (15) is equivalent to

$$P^\top\pi=P^\top\pi \quad\Longrightarrow\; P^\top[\pi]\mathbb{1}_n=\pi \quad\Longrightarrow\; Q^\top\mathbb{1}_n=\pi.$$

Hence constraint (15) is equivalent to constraint (27). This completes the proof of equivalency of constraints.

Second, we shall show that the maximization of the objective function in Problem 2 is equivalent to the maximization of the objective function in Problem 2 subject to the same constraints. For a given stationary distribution π, as a result of Lemma 3 $\mathcal{H}_\pi(P)$ and $\mathcal{H}(Q)$ differ by a constant quantity $\mathcal{H}(\pi)$. Hence the maximization of the objective functions in the two problems are equivalent. Given an optimal solution P^* to Problem 2 one can construct an ergodic flow matrix Q^* using equation (20) and vice-versa. Thus (i) and (ii) hold. □

Problem 4 (Relaxed convex program to maximize entropy of ergodic flow)**.** *Given a connected undirected unweighted graph G and a positive vector $\pi\in$ interior(Δ_n), compute Q such that*

$$\begin{aligned}\max\quad & \mathcal{H}(Q) && (28)\\ \textit{subj. to}\quad & Q\geq 0, && (29)\\ & q_{ij}=0,\ \textit{if } \{i,j\} \textit{ is not an edge of } G, && (30)\\ & Q\mathbb{1}_n+Q^\top\mathbb{1}_n=2\pi. && (31)\end{aligned}$$

One can show that when the graph G has self-loops at each node, the optimal values of Problem 3 and Problem 4 are the same.

Theorem 7 (Equality of solutions to Problem 3 and Problem 4)**.** *Let G be a connected undirected graph with self-loop at each node, A be the binary adjacency matrix associated to G, and $\pi \in \text{interior}(\Delta_n)$ be a positive vector. Denote the optimal value of Problem 4 by Q_r^* and the optimal value of Problem 3 by Q^*. Then the following statements hold:*

(i) $Q_r^* = Q^*$,
(ii) *there exists a vector* $x \in \mathbb{R}^n_{>0}$ *such that* $Q^* = [x]A[x]$.

Proof. Note that graph and stationary constraints are identical in both problem formulations. Further constraint (31) in Problem 4 is obtained by adding (26) and (27). Therefore, the feasible set of Problem 4 is larger than the feasible set of Problem 3 and thus we have $\mathcal{H}(Q^*) \leq \mathcal{H}(Q_r^*)$.

Using similar arguments to the proof of Theorem 5, one can show that if $Q_r^* = [q_{ij}^*]$ is the solution for Problem 4, then we have $q_{ij}^* > 0$ if and only if $a_{ij} = 1$. This implies that Q_r^* is the critical point of the Lagrange dual function $\mathcal{L} : \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathcal{L}(Q, \lambda) = -\sum_{\{i,j\}\in\mathcal{E}} q_{ij} \log q_{ij} - \sum_{i=1}^{n} \sum_{\{i,j\}\in\mathcal{E}} \lambda_i (q_{ij} + q_{ji} - 2\pi_i),$$

where $\mathcal{E}$ is the edge set of the graph G. Setting the partial derivatives of $\mathcal{L}$ to zero, for every $\{i, j\} \in \mathcal{E}$, we obtain

$$\frac{\partial \mathcal{L}}{\partial q_{ij}} = 1 + \log q_{ij} - \lambda_i - \lambda_j.$$

Introducing new Lagrange multipliers $\tilde{\lambda}_i = \lambda_i + 1/2$, the solution $Q_r^* = [q_{ij}^*]$ satisfies $q_{ij}^* = a_{ij} \exp^{-\tilde{\lambda}_i} \exp^{-\tilde{\lambda}_j}$.

Let $x_i = \exp^{-\tilde{\lambda}_i}$ then $q_{ij}^* = a_{ij} x_i x_j$ or in matrix notation $Q_r^* = [x]A[x]$. Substituting this solution into the constraints in Problem 4 and using the fact that $a_{ij} = a_{ji}$,

$$\begin{aligned}
&\sum_j a_{ij} x_i x_j + \sum_j a_{ji} x_j x_i = 2\pi_i \\
&\implies x_i \sum_j a_{ij} x_j = \pi_i \\
&\implies [x]Ax = \pi.
\end{aligned}$$

Note that A is symmetric, binary matrix with unit diagonal entries. Thus, by Theorem 3, there exists a unique $x^* \in \mathbb{R}^n_{>0}$ such that $[x^*]Ax^* = \pi$. Therefore the global maximum of the concave function $\mathcal{H}$ is given by

$$Q_r^* = [x^*]A[x^*]. \tag{32}$$

One can verify that the solution Q_r^* also satisfies constraints (26) and (27) in Problem 3. This, together with the fact that the feasible set of Problem 4 is larger than the feasible set of the Problem 3, implies that $Q_r^* = Q^*$. This completes the proof of the part (i). Part (ii) of the theorem follows from part (i) and equation 32. ∎

Now we have the requisite results to prove Theorem 6.

Proof of Theorem 6. Using Lemma 4 and Theorem 7, the solution $P^* = [p_{ij}^*]$ to Problem 2 is given by $p_{ij}^* = (Ax)_i^{-1} a_{ij} x_j$ or in matrix notation as $P^* = [Ax]^{-1} A[x] = \Phi_A(x)$. Also as a result of (iv) in Theorem 5 P^* is reversible.

The entropy of the ergodic flow matrix Q^* is given by

$$\begin{aligned}
\mathcal{H}(Q^*) &= -\sum_{i,j=1}^{n} q_{ij}^* \log q_{ij}^* \\
&= -\sum_{i,j=1}^{n} a_{ij} x_i x_j (\log x_i + \log x_j) \\
&= -2 \sum_i x_i \sum_j a_{ij} x_j \log(x_j) = -2x^\top A([x] \log(x)).
\end{aligned}$$

The entropy rate of P^* is given by $\mathcal{H}_\pi(P^*) = \mathcal{H}(Q^*) - \mathcal{H}(\pi)$. The quantity $\mathcal{H}(\pi) = -\pi^\top \log \pi$ in vector notation and hence the result in equation (18). ∎

Proof of Lemma 2. Note that from equations (20) and (32) we can write the probability of transition from s to t as

$$p_{st} = \frac{a_{st} x_s x_t}{\pi_s}.$$

For the sake of brevity let $\sigma_m = \sigma(l_m)$ for every $1 \leq m \leq k$. Consider the probability of any valid permutation of the path $i, l_1, l_2, \ldots, l_k, j$ being traversed. This is given by

$$\begin{aligned}
p_{i\sigma_1} p_{\sigma_1\sigma_2} \ldots p_{\sigma_k j} &= a_{i\sigma_1} a_{\sigma_1\sigma_2} \ldots a_{\sigma_k j} \frac{x_i x_{\sigma_1} x_{\sigma_1} x_{\sigma_2} \ldots x_{\sigma_k} x_j}{\pi_i \pi_{\sigma_1} \pi_{\sigma_2} \ldots \pi_{\sigma_k}} \\
&= \frac{x_i x_j}{\pi_i} \frac{x_{\sigma_1}^2 x_{\sigma_2}^2 \ldots x_{\sigma_k}^2}{\pi_{\sigma_1} \pi_{\sigma_2} \ldots \pi_{\sigma_k}} \\
&= p_{ij} p_{\sigma_1\sigma_1} p_{\sigma_2\sigma_2} \ldots p_{\sigma_k\sigma_k}.
\end{aligned}$$

The quantity $p_{\sigma_1\sigma_1} p_{\sigma_2\sigma_2} \ldots p_{\sigma_k\sigma_k}$ is invariant to permutations of the sequence $\{l_m\}_{1\leq m\leq k}$. Hence all such paths are equiprobable. ∎

V. Application to Robotic Surveillance

In this section, we apply maxentropic chains with non-uniform stationary distributions to the design of robotic surveillance strategies over graphs.

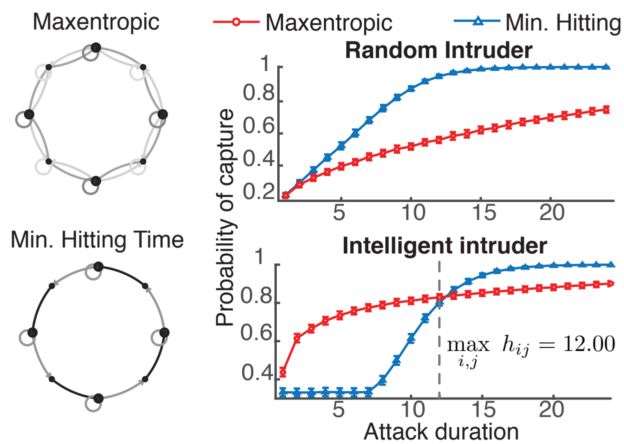
A. Setup

We consider scenarios in which (i) surveillance agents move on a roadmap (i.e., an undirected graph) according to a discrete-time random walk, (ii) intruders appear at random locations on the roadmap at random times, (iii) intruders can observe the local presence/absence of the surveillance agent(s) and decide when to attack and (iv) the intruder attack is detected precisely when a surveillance agent and the intruder are at the same location during the intruder attack. We consider the following settings: a single agent on a ring, a single agent on a lattice (see Fig. 1), and multiple agents on a partitioned map of a realistic environment (see Fig. 2).

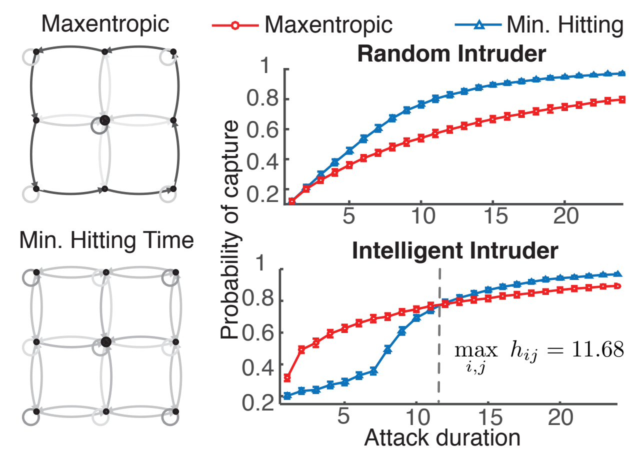
B. Intruder models

Given a probability vector $\pi \in \Delta_n$, we consider the following intruders models.

(i) *The Random Intruder*: The random intruder has no knowledge of the position of the surveillance agent(s). Such an intruder selects a node i with probability π_i. The attack takes an arbitrary duration which is quantified by the number of transitions performed by the agent(s).



(a) Ring graph



(b) Lattice graph

Fig. 1: Comparison of maxentropic Markov chain strategy with minimum hitting time strategy. The worst hitting time for the minimum hitting time Markov chain is denoted by $\max_{i,j} h_{ij}$.

(ii) *The Intelligent Intruder*: The intelligent intruder selects a node i with probability π_i, waits for a/the surveillance agent to arrive at the node, and commences an attack lasting for an arbitrary duration in the timestep immediately following the visit of the surveillance agent. The attack duration is quantified by the number of transitions performed by the agent(s). (Intelligent intruders have been previously studied for example by [25], [4].)

We visualize and design the probability vector π as follows. In Fig. 1 the size of the nodes depicts the importance of the node and hence the desired visit frequency. For the ring graph, the north, east, west and south nodes have been assigned twice the priority of the remaining nodes. For the lattice graph, the central node has twice the priority of the peripheral nodes. In Fig. 2, which depicts the multi-agent case, we prepartition the graph for four agents and specify visit frequencies such that all nodes have the same priority. Equal priority with overlapping subgraphs is achieved by specifying a nonuniform visit frequency for each agent on their individual subgraphs. We do this by splitting the visit frequency load equally between the agents for shared nodes.

C. Surveillance strategies

First, we assume that the intruders and surveillance agents assign the same level of priority to nodes in the graph. In other words, visit frequencies by surveillance agents are biased in a manner so as to be proportional to intruder attacks. Second, we consider two policies for the surveillance agent:

(i) *The maxentropic agent*: The surveillance agent adopts a policy which is the maxentropic Markov chain with visit frequencies proportional to the importance of the node, i.e., the solution described in Theorem 6.

(ii) *The minimum hitting time agent*: Let $\{h_{ij}(P)\}_{ij}$ denote the matrix of mean hitting times for the Markov chain modeled by the transition matrix P. Consider the following optimization program.

Problem 5 (Nonlinear program to minimize mean hitting time)**.** *Given a connected undirected unweighted graph G and a positive vector $\pi \in \mathrm{interior}(\Delta_n)$, compute P such that*

$$\begin{aligned} \min \quad & \sum_i \sum_j \pi_i \pi_j h_{ij}(P) \\ \textit{subj. to} \quad & P \geq 0, \\ & p_{ij} = 0, \ \textit{if } a_{ij} = 0, \\ & P\mathbb{1}_n = \mathbb{1}_n, \\ & \pi^\top P = \pi^\top. \end{aligned}$$

The solution to this nonlinear program is the Markov chain adopted by the minimum hitting time agent. The numerical optimization is conducted using a sequential quadratic programming solver as implemented by the KNITRO/TOMLAB package; for the graph sizes of interest here, this package reliably computes the global minimum solutions. This nonlinear program is identical to the formulation in [22, Problem 1] for a single agent.

For the multi-agent case, each agent performs either the maxentropic Markov chain strategy or the minimum hitting time strategy on their respective subgraphs. There is no actual coordination among the agents (except the joint specification of individual visit frequencies).

D. Simulation Results

Results for Random Intruders. For the random intruder, for all choices of visit frequencies on a variety of graph topologies, we find that the minimum hitting time agent outperforms the maxentropic agent for all attack durations. The minimum hitting time Markov chain results in faster travel times through the graph. In the absence of knowledge of attack durations, simulations indicate that a strategy with emphasis on fast travel times (small hitting times) performs better than one with emphasis on unpredictability such as the maxentropic chain. Maxentropic Markov chains by their reversible nature have mixing times of $O(D^2)$ where D is the diameter of the graph [13]. For reversible Markov chains bounds exist on the mixing time and the mean hitting time showing that these notions are equivalent [3]. Thus the maxentropic agent has

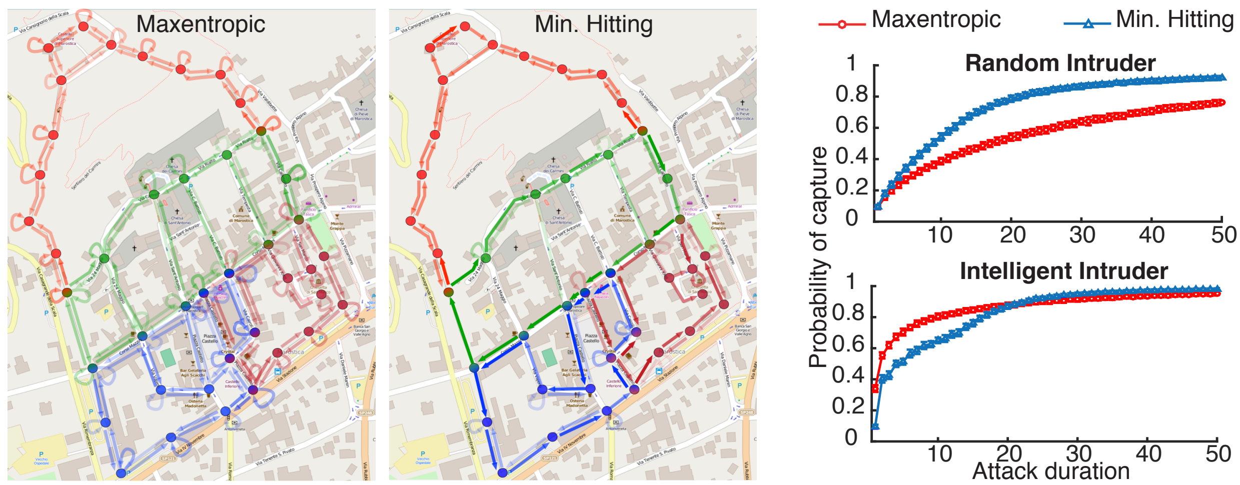


Fig. 2: Comparison of maxentropic Markov chain strategy with minimum hitting time strategy for the multi-agent case on a partitioned graph. The color(s) of the node indicates which agent(s) are surveilling the node and the opacity of the arrows indicate transition probabilities.

mean hitting time $O(D^2)$ whereas it is likely that minimum hitting time agent achieves hitting times of $O(D)$ in all cases, though such a result remains to be proved.

Results for Intelligent Intruders. For the intelligent intruder with relatively short attack durations, we note that the maxentropic strategy outperform the minimum hitting time strategy. It stands to reason that, for attack durations larger than the worst hitting time of the chain, the capture of the intelligent intruder is very probable for the minimum hitting time strategy (capture is certain for cases on the ring with uniform stationary distribution where the minimum hitting time strategy is a clockwise or a counterclockwise traversal). In simulations, it is observed that for attack durations which are larger than the worst hitting times of the minimum hitting time chain, the minimum hitting time strategy performs better (see Fig. 1). Analysis of the hitting times of the maxentropic chain might reveal an exact condition of the regime of attack durations wherein each strategy leads to higher capture rates.

We summarize these results in Table III. In short, these results indicate that introducing unpredictability into surveillance strategies is appropriate in the important and realistic setting where (i) the intruder uses knowledge of the agents' locations to plan its attacks (e.g., attacking as soon as an agent leaves), and (ii) attack have sufficiently short duration so that they are not detectable by simple fast surveillance agents.

	Random Intruder	**Intelligent Intruder**
Maxentropic Agent	Low capture rate	High capture rate when attack duration is low
Min. Hitting Agent	High capture rate	High capture rate when attack duration is high

TABLE III: Qualitative summary of results for intruder and agent models.

VI. Conclusion

In this article we considered the optimization problem of maximizing the entropy rate of a Markov chain with prescribed stationary distribution. We showed this problem is strictly convex with a unique global optimizer. We provided a fast iterative algorithm with rigorous convergence guarantees to compute the so-called entropic vector; as a function of this entropic vector, we provide a closed-form formula for the maximum entropy Markov chain with prescribed stationary distribution. We then characterized several properties of maxentropic chains. The interest for Markov chains with maximum entropy and prescribed stationary distributions arises naturally in robotic surveillance; accordingly we showed some realizations of optimal chains for prototypical robotic roadmaps.

Numerous future research directions remain open. First, it is potentially important to extend our analysis to more general graph settings, including graphs without a complete set of self-loops and directed graphs with asymmetric adjacency matrices. For graphs without a complete set of self-loops, we present a mathematical conjecture inspired by our results on the maxentropic matrix and vector maps.

Conjecture Given a connected graph G with binary adjacency matrix A, the set of stationary distributions for all irreducible Markov chains over G is $\{[x]Ax/\|[x]Ax\|_1 \mid x \in \mathbb{R}^n_{>0}\}$. Numerical simulations indicate that the set of feasible stationary distributions over sparse graphs without self-loops is of measure zero (for an appropriately defined measure).

Second, in the robotic surveillance context, it is of interest to combine notions of unpredictability with speed of traversal of graphs; see recent related work in [21]. Appropriate notions of unpredictability for the multi-agent case are yet to be developed and could lead to the design of effective strategies against intruders with advanced planning and sensing capabilities.

References

[1] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer, 2 edition, 1988.

[2] P. Agharkar and F. Bullo. Quickest detection over robotic roadmaps. *IEEE Transactions on Robotics*, 32(1):252–259, 2016. doi:10.1109/TRO.2015.2506165.

[3] D. Aldous and J. A. Fill. Reversible Markov Chains and Random Walks on Graphs, 2002. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[4] A. B. Asghar and S. L. Smith. Stochastic patrolling in adversarial settings. In *American Control Conference*, pages 6435–6440, Boston, USA, July 2016. doi:10.1109/ACC.2016.7526682.

[5] S. Bandyopadhyay, S. J. Chung, and F. Y. Hadaegh. Probabilistic and distributed control of a large-scale swarm of autonomous agents. *IEEE Transactions on Robotics*, 33(5):1103–1123, 2017. doi:10.1109/TRO.2017.2705044.

[6] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.

[7] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw. Localization of the maximal entropy random walk. *Physical Review Letters*, 102:160602, 2009. doi:10.1103/PhysRevLett.102.160602.

[8] G. Cannata and A. Sgorbissa. A minimalist algorithm for multirobot continuous coverage. *IEEE Transactions on Robotics*, 27(2):297–312, 2011. doi:10.1109/TRO.2011.2104510.

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[10] Y. Fan and H. Liu. Mitigating congestion in complex transportation networks via maximum entropy, 2017. URL: https://arxiv.org/pdf/1701.04974.

[11] J. Grace and J. Baillieul. Stochastic strategies for autonomous robotic surveillance. In *IEEE Conf. on Decision and Control and European Control Conference*, pages 2200–2205, Seville, Spain, December 2005. doi:10.1109/CDC.2005.1582488.

[12] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, October 2016. URL: http://cvxr.com/cvx.

[13] K. Jung, D. Shah, and J. Shin. Distributed averaging via lifted Markov chains. *IEEE Transactions on Information Theory*, 56(1):634–647, 2010. doi:10.1109/TIT.2009.2034777.

[14] J. Justesen and T. Hoholdt. Maxentropic Markov chains. *IEEE Transactions on Information Theory*, 30(4):665–667, 1984. doi:10.1109/TIT.1984.1056939.

[15] A. F. Karr. Markov chains and processes with a prescribed invariant measure. *Stochastic Processes and their Applications*, 7(3):277–290, 1978. doi:10.1016/0304-4149(78)90047-9.

[16] P. Korus and J. Huang. Improved tampering localization in digital image forensics based on maximal entropy random walk. *IEEE Signal Processing Letters*, 23(1):169–173, 2016. doi:10.1109/LSP.2015.2507598.

[17] S. G. Krantz and H. R. Parks. *The Implicit Function Theorem*. Birkhäuser, 2013.

[18] R.-H. Li, J. X. Yu, and J. Liu. Link prediction: The power of maximal entropy random walk. In *ACM Int. Conf. on Information and Knowledge Management*, pages 1147–1156, Glasgow, UK, October 2011. doi:10.1145/2063576.2063741.

[19] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994. doi:10.1137/1.9781611970791.

[20] J. K. Ochab and Z. Burda. Maximal entropy random walk in community detection. *The European Physical Journal Special Topics*, 216(1):73–81, 2013. doi:10.1140/epjst/e2013-01730-6.

[21] R. Patel, P. Agharkar, and F. Bullo. Robotic surveillance and Markov chains with minimal weighted Kemeny constant. *IEEE Transactions on Automatic Control*, 60(12):3156–3167, 2015. doi:10.1109/TAC.2015.2426317.

[22] R. Patel, A. Carron, and F. Bullo. The hitting time of multiple random walks. *SIAM Journal on Matrix Analysis and Applications*, 37(3):933–954, 2016. doi:10.1137/15M1010737.

[23] X. Peng and Z. Zhang. Maximal entropy random walk improves efficiency of trapping in dendrimers. *Journal of Chemical Physics*, 140(23):234104, 2014. doi:10.1063/1.4883335.

[24] W. Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 3 edition, 1976.

[25] T. Sak, J. Wainer, and S. Goldenstein. Probabilistic multiagent patrolling. In G. Zaverucha and A. L. da Costa, editors, *Brazilian Symposium on Artificial Intelligence*, volume 5249 of *Lecture Notes in Computer Science*, pages 124–133. Springer, 2008. doi:10.1007/978-3-540-88190-2_18.

[26] K. Srivastava, D. M. Stipanovič, and M. W. Spong. On a stochastic robotic surveillance problem. In *IEEE Conf. on Decision and Control*, pages 8567–8574, Shanghai, China, December 2009. doi:10.1109/CDC.2009.5400569.

[27] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.

[28] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996. doi:10.1137/1038003.

[29] L. Wang, J. Zhao, X. Hu, and J. Lu. Weakly supervised object localization via maximal entropy random walk. In *IEEE Int. Conf. on Image Processing*, pages 1614–1617, Paris, France, October 2014. doi:10.1109/ICIP.2014.7025323.

[30] J.-G. Yu, J. Zhao, J. Tian, and Y. Tan. Maximal entropy random walk for region-based visual saliency. *IEEE Transactions on Cybernetics*, 44(9):1661–1672, 2014. doi:10.1109/TCYB.2013.2292054.

Mishel George received a B.Tech. degree in Engineering Physics from the Indian Institute of Technology, Bombay in 2011. From 2011 through 2012 he was a research assistant at the Indian Institute of Science, Bangalore where he studied bio-inspired multi-agent algorithms. Currently, he is a graduate student at the University of California, Santa Barbara pursuing a Ph.D. in mechanical engineering. His research interests include the design of stochastic surveillance strategies and the study of biological networks.

Saber Jafarpour (M'16) is a Postdoctoral researcher with the Department of Mechanical Engineering at the University of California, Santa Barbara. He received his Ph.D. in 2016 from the Department of Mathematics and Statistics at Queen's University. His research interests includes stability analysis of network systems with application to power grids and geometric control theory.

Francesco Bullo (S'95-M'99-SM'03-F'10) is a Professor with the Mechanical Engineering Department and the Center for Control, Dynamical Systems and Computation at the University of California, Santa Barbara. He was previously associated with the University of Padova, the California Institute of Technology, and the University of Illinois. His research interests focus on network systems and distributed control with application to robotic coordination, power grids and social networks. He is the coauthor of "Geometric Control of Mechanical Systems" (Springer, 2004) and "Distributed Control of Robotic Networks" (Princeton, 2009); his forthcoming "Lectures on Network Systems" is available on his website. He received best paper awards for his work in IEEE Control Systems, Automatica, SIAM Journal on Control and Optimization, IEEE Transactions on Circuits and Systems, and IEEE Transactions on Control of Network Systems. He is a Fellow of IEEE and IFAC. He has served on the editorial boards of IEEE, SIAM, and ESAIM journals, and will serve as IEEE CSS President in 2018.