The Entropy of Conditional Markov Trajectories

Mohamed Kafsi, *Student Member, IEEE*, Matthias Grossglauser, *Member, IEEE*, and Patrick Thiran, *Senior Member, IEEE*

***Abstract*—To quantify the randomness of Markov trajectories with fixed initial and final states, Ekroot and Cover proposed a closed-form expression for the entropy of trajectories of an irreducible finite state Markov chain. Numerous applications, including the study of random walks on graphs, require the computation of the entropy of Markov trajectories conditional on a set of intermediate states. However, the expression of Ekroot and Cover does not allow for computing this quantity. In this paper, we propose a method to compute the entropy of conditional Markov trajectories through a transformation of the original Markov chain into a Markov chain that exhibits the desired conditional distribution of trajectories. Moreover, we express the entropy of Markov trajectories—a global quantity—as a linear combination of *local entropies* associated with the Markov chain states.**

***Index Terms*—Entropy, Markov chains (MC), Markov trajectories.**

I. Introduction

QUANTIFYING the randomness of Markov trajectories has applications in graph theory [1] and in statistical physics [2], as well as in the study of random walks on graphs [3], [4]. The need to quantify the randomness of Markov trajectories first arose when Lloyd and Pagels [2] defined a measure of complexity for the macroscopic states of physical systems. They examine some intuitive properties that a measure of complexity should have and propose a universal measure called *depth*. They suggest that the depth of a state should depend on the complexity of the process by which that state arose, and prove that it must be proportional to the Shannon entropy of the set of trajectories leading to that state. Subsequently, Ekroot and Cover [5] studied the computational aspect of the depth measure. In order to quantify the number of bits of randomness in a Markov trajectory, they propose a closed-form expression for the entropy of trajectories of an irreducible finite state Markov chain (MC). Their expression does not allow, however, for computing the entropy of Markov trajectories conditional on the realization of a set of intermediate states. Computing the conditional entropy of Markov trajectories turns out to be very challenging yet useful in numerous domains, including the study of mobility predictability and its dependence on location side information.

Consider a scenario where we are interested in quantifying the predictability of route-choice behavior. We can model the mobility of a traveler as a weighted random walk on a graph whose vertices represent locations and edges represent possible transitions [6]. We can therefore model a route as a sample path or trajectory in a MC. If we suppose that we know where the traveler starts and ends her/his route, the randomness of the route she/he would follow is represented by the distribution of trajectories between the source and destination vertices. Consequently, the predictability of her/his route is captured by the entropy of Markov trajectories between these two states. Now, if we obtain side information stating that the traveler went (or has to go) through a set of intermediate vertices, quantifying the evolution of her/his route predictability requires the computation of the trajectory entropy conditional on the set of known intermediate states. The conditional entropy is also a way to quantify the informational value of the intermediate states revealed. For example, if the entropy conditional on the set of known intermediate states is zero, then this set reveals the whole trajectory of the traveler.

In our paper, we propose a method to compute the entropy of Markov trajectories conditional on a set of intermediate states. The method is based on a transformation of the original MC so that the transformed MC exhibits the desired conditional distribution of trajectories. We also derive an expression that enables us to compute the entropy of Markov trajectories, under conditions weaker than those assumed in [5]. Moreover, this expression links the entropy of Markov trajectories to the local entropies at the MC states.

II. Model

Let $\{X_i\}$ be a finite state irreducible and aperiodic MC with transition probability matrix P whose elements are the transition probabilities

$$\begin{aligned}P_{x_n x_{n+1}} &= p(X_{n+1} = x_{n+1} | X_n = x_n)\\ &= p(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1).\end{aligned}$$

This MC admits a stationary distribution Π, which is the unique solution of $\Pi = \Pi P$. The entropy rate $H(X)$ is a measure of the average entropy growth of a sequence generated by the process $\{X_i\}$ and is defined as

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n).$$

For the particular case of an irreducible and aperiodic MC, the limit above is equal to [7, p. 77]

$$H(X) = \sum_i \Pi(i) H(P_{i\cdot}),$$

where $P_{i\cdot}$ denotes the ith row of P and where $H(P_{i\cdot}) = -\sum_j P_{ij}\log(P_{ij})$ is the *local entropy* of state i. Note that,

Manuscript received December 12, 2012; accepted March 13, 2013. Date of publication May 13, 2013; date of current version August 14, 2013.

The authors are with the School of Computer and Communication Science, EPFL, Lausanne CH-1015, Switzerland (e-mail: mohamed.kafsi@epfl.ch; matthias.grossglauser@epfl.ch; Patrick.Thiran@epfl.ch).

Communicated by I. Kontoyiannis, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2013.2262497

throughout this paper, we use MC_P as a shorthand for the MC whose transition probability matrix is P.

A. Entropy of Markov Trajectories

We follow the setting of [5] closely. We define a *random trajectory* T_{sd} of a MC as a path with initial state s, final state d, and no intermediate state d, i.e., the trajectory is terminated as soon as it reaches state d. Using the Markov property, we express the probability of a particular trajectory $t_{sd} = sx_2 \dots x_k d$ given that $X_1 = s$ as

$$p(t_{sd}) = P_{sx_2} P_{x_2 x_3} \dots P_{x_k d}.$$

Let $\mathcal{T}_{sd}$ be the set of all trajectories that start at state s and end as soon as they reach state d. As the MC defined by the matrix P is finite and irreducible, we have

$$\sum_{t_{sd} \in \mathcal{T}_{sd}} p(t_{sd}) = 1 \qquad \text{for all } s, d.$$

So the discrete random variable T_{sd} has as support the set $\mathcal{T}_{sd}$, with the probability mass function $p(t_{sd})$. Subsequently, we use $p(t_{sd})$ as a shorthand for $p(T_{sd} = t_{sd})$. We can now express the entropy of the random trajectory T_{sd} as

$$H_{sd} \equiv H(T_{sd}) = - \sum_{t_{sd} \in \mathcal{T}_{sd}} p(t_{sd}) \log p(t_{sd}).$$

We define the matrix of trajectory entropies H where $H_{ij} = H(T_{ij})$. Ekroot and Cover [5] provide a general closed-form expression for the matrix H of an irreducible, aperiodic, and finite state MC.

The entropy $H_{sd|u}$ of a trajectory from s to d given that it goes through u is defined by

$$\begin{aligned} H_{sd|u} &\equiv H(T_{sd} | T_{sd} \in \mathcal{T}^u_{sd}) \\ &= - \sum_{t_{sd} \in \mathcal{T}^u_{sd}} p(t_{sd} | T_{sd} \in \mathcal{T}^u_{sd}) \log p(t_{sd} | T_{sd} \in \mathcal{T}^u_{sd}), \end{aligned} \tag{1}$$

where $\mathcal{T}^u_{sd}$ is the set of all trajectories in $\mathcal{T}_{sd}$ with an intermediate state u

$$\mathcal{T}^u_{sd} = \{t_{sd} \in \mathcal{T}_{sd} : t_{sd} = s \dots u \dots d\}.$$

The major challenge is to compute efficiently the entropy $H_{sd|u}$. Even the costly approach of computing all the terms of the sum (1) is not always possible because the set $\mathcal{T}^u_{sd}$ has an infinite number of members in the case, where after removing state d, the transition graph of the MC is not a DAG. It is important to emphasize that the entropy $H_{sd|u}$ is not the entropy of the random variable T_{sd} given another random variable—a quantity which is easy to compute—but the entropy of T_{sd} conditional on the realization of a dependent random variable.

In Fig. 1, we show an example of a finite-state irreducible and aperiodic MC. Note that the presence of cycles implies that the set of trajectories between some pair of states might have infinite cardinality ($|\mathcal{T}_{14}| = \infty$, for example). Therefore, in addition to being complex, the naive approach of enumerating all trajectories is not always possible.

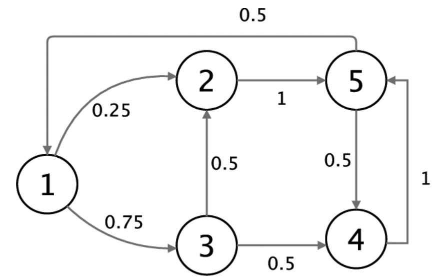

Fig. 1. Irreducible, 5-state, Markov chain annotated with the transition probabilities.

Using the results of [5], we obtain the matrix of trajectory entropies

$$H = \begin{pmatrix} 3.56 & 3.69 & 1.74 & 3.18 & 1.56 \\ 2 & 5.69 & 3.74 & 2.59 & 0 \\ 3 & 3.84 & 4.74 & 2.29 & 1 \\ 2 & 5.69 & 3.74 & 2.59 & 0 \\ 2 & 5.69 & 3.74 & 2.59 & 1.78 \end{pmatrix}.$$

The zero elements of the matrix H correspond to deterministic trajectories such as T_{25}, which is equal to the path 25 with probability 1 since no other path allows a walk to go from 2 to 5. The entropy of the random trajectory T_{15} is 1.56 bits. Now imagine that we have an additional piece of information stating that the trajectory T_{15} goes through state 4. Intuitively, we would be tempted to argue that the entropy $H_{15|4}$ of the trajectory T_{15} conditional on going through state 4 is equal to $H_{14} + H_{45}$, but this additivity property does not hold. Indeed, the conditional entropy $H_{15|4}$ is zero because the trajectory T_{15}, conditional on the intermediate state 4, can only be equal to the path 1345, whereas $H_{14} = 3.18$ bits, hence $H_{14} + H_{45} = 3.18 + 0 = 3.18 \neq H_{15|4}$ bits.

In the next section, we study the entropy of Markov trajectories conditional on *multiple* intermediate states and derive a general expression for this entropy.

III. Entropy of Conditional Markov Trajectories

Let α_{sud} denote the probability that the random trajectory T_{sd} goes through the state u at least once

$$\alpha_{sud} = p(T_{sd} \in \mathcal{T}^u_{sd}).$$

This is also equal to the probability that a walk reaches the state u before the state d, given that it started at s. In order to compute α_{sud}, the technique from [8], [9] is to make the states u and d absorbing (a state i is absorbing if and only if $P_{ii} = 1$) and compute the probability to be absorbed by state u given that the trajectory has started at state s.

Our first step toward computing $H_{sd|u}$ is to express it as a function of quantities that are much simpler to compute. The idea is to relate the entropy of a trajectory conditional on a given state to its entropy conditional on *not* going through that state. Therefore, we define the entropy $H_{sd|\bar{u}}$ of a trajectory from s to d given that it does *not* go through u to be

$$H_{sd|\bar{u}} \equiv H(T_{sd} | T_{sd} \notin \mathcal{T}^u_{sd}).$$

Using the chain rule for entropy, we can derive the following equation which relates $H_{sd|u}$ to H_{sd}, $H_{sd|\bar{u}}$ and α_{sud}:

$$H_{sd} = \alpha_{sud} H_{sd|u} + (1 - \alpha_{sud}) H_{sd|\bar{u}} + h(\alpha_{sud}) \quad (2)$$

for all u, where $h(\alpha_{sud})$ is the entropy of a Bernoulli random variable with success probability α_{sud}.

Proof: First, we define the indicator variable I by

$$I = \begin{cases} 1 & \text{if } \mathrm{T}_{sd} \in \mathcal{T}_{sd}^{u}, \\ 0 & \text{otherwise.} \end{cases}$$

Using the chain rule for entropy, we express the joint entropy $H(T_{sd}, I)$ in two different ways

$$\begin{aligned} H(T_{sd}, I) &= H(I) + H(T_{sd}|I) \\ &= H(T_{sd}) + H(I|T_{sd}) = H(T_{sd}), \end{aligned}$$

because I is a deterministic function of T_{sd}. So the entropy of the random trajectory T_{sd} can be expressed as

$$\begin{aligned} H(T_{sd}) &= H(I) + H(T_{sd}|I) \\ &= H(I) + H(T_{sd}|I = 1)p(I = 1) \\ &\quad + H(T_{sd}|I = 0)p(I = 0) \\ &= H(I) + H(T_{sd}|T_{sd} \in \mathcal{T}_{sd}^{u})p(T_{sd} \in \mathcal{T}_{sd}^{u}) \\ &\quad + H(T_{sd}|T_{sd} \notin \mathcal{T}_{sd}^{u})p(T_{sd} \notin \mathcal{T}_{sd}^{u}). \end{aligned}$$

Since $\alpha_{sud} = p(T_{sd} \in \mathcal{T}_{sd}^{u}) = p(I = 1)$, we obtain

$$\begin{aligned} H(T_{sd}) &= \alpha_{sud} H(T_{sd}|T_{sd} \in \mathcal{T}_{sd}^{u}) \\ &\quad + (1 - \alpha_{sud}) H(T_{sd}|T_{sd} \notin \mathcal{T}_{sd}^{u}) + h(\alpha_{sud}). \end{aligned}$$

■

As we know from [5], [8], [9] how to compute H_{sd} and α_{sud}, if we are able to compute $H_{sd|\bar{u}}$, we can use (2) to find $H_{sd|u}$. However, generalizing (2) to trajectories conditional on passing through *multiple* intermediate states turns out to be difficult, hence we propose an approach that circumvents this problem. As we will see, the difficulty of our approach also boils down to computing the entropy of a trajectory conditional on *not* going through a given state.

First, we define $\mathcal{T}_{sd}^{\boldsymbol{u}}$, the set of all trajectories in $\mathcal{T}_{sd}$ that exhibit the sequence of intermediate states $\boldsymbol{u} = u_1 u_2 \dots u_l$, i.e.,

$$\mathcal{T}_{sd}^{\boldsymbol{u}} = \{t_{sd} \in \mathcal{T}_{sd} : t_{sd} = s \dots u_1 \dots u_2 \dots u_l \dots d\}.$$

For an arbitrary sequence of states $\boldsymbol{u} = u_1 u_2 \dots u_l$, satisfying $p(T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) > 0$, we prove the following lemma.

Lemma 1:

$$H(T_{sd}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) = \sum_{k=0}^{l-1} H_{u_k u_{k+1}|\bar{d}} + H_{u_l d}, \quad (3)$$

where $u_0 = s$.

Proof: First, given $T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}$, the random trajectory T_{sd} can be expressed as a sequence of random subtrajectories $(T_{su_1}, T_{u_1 u_2}, \dots, T_{u_{l-1} u_l}, T_{u_l d})$. Therefore, the conditional entropy $H(T_{sd}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}})$, which we denote by $H_{sd|u_1 \dots u_l}$, can be written as a joint subtrajectory entropy

$$H_{sd|u_1 \dots u_l} = H(T_{su_1}, T_{u_1 u_2}, \dots, T_{u_l d}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}).$$

Applying the chain rule for entropy, we obtain successively

$$\begin{aligned} H_{sd|u_1 \dots u_l} &= H(T_{su_1}, T_{u_1 u_2}, \dots, T_{u_l d}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &= H(T_{su_1}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &\quad + H(T_{u_1 u_2}|T_{su_1}; T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &\quad \vdots \\ &\quad + H(T_{u_l d}|T_{su_1}, \dots, T_{u_{l-1} u_l}; T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}). \end{aligned}$$

The Markovian nature of the process generating the trajectory T_{sd} implies that each of the subtrajectories $T_{u_k u_{k+1}}$ is independent of the preceding ones, given its starting point u_k. Since the sequence $s\boldsymbol{u} = s u_1 u_2 \dots u_l$ defines the starting point of each subtrajectory, we can therefore write that

$$\begin{aligned} &H(T_{u_k u_{k+1}}|T_{su_1}, \dots, T_{u_{k-1} u_k}; T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &= H(T_{u_k u_{k+1}}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}). \end{aligned} \quad (4)$$

Using (4), the expression for the conditional entropy becomes

$$\begin{aligned} H_{sd|u_1 \dots u_l} &= H(T_{su_1}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &\quad + H(T_{u_1 u_2}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}) \\ &\quad \vdots \\ &\quad + H(T_{u_l d}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}). \end{aligned}$$

Note that for each trajectory $T_{u_k u_{k+1}}$, the only restriction imposed by the event $\{T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}\}$ is that the final state d cannot be an intermediate state of any of the first l trajectories $T_{su_1}, T_{u_1 u_2}, \dots, T_{u_{l-1} u_l}$. As a result,

$$\begin{aligned} H_{sd|u_1 \dots u_l} &= H(T_{su_1}|T_{su_1} \notin \mathcal{T}_{su_1}^{d}) \\ &\quad + H(T_{u_1 u_2}|T_{u_1 u_2} \notin \mathcal{T}_{u_1 u_2}^{d}) \\ &\quad \vdots \\ &\quad + H(T_{u_l d}) \\ &= \sum_{k=0}^{l-1} H_{u_k u_{k+1}|\bar{d}} + H_{u_l d}, \end{aligned}$$

where $u_0 = s$. ■

Now, if we are able to compute $H_{u_k u_{k+1}|\bar{d}}$, we can use (3) to derive $H(T_{sd}|T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}})$. The following lemma shows how the conditional entropy $H_{u_k u_{k+1}|\bar{d}}$ can be obtained by a simple modification of the MC.

We consider a MC whose transition probability matrix is P, and s, u, and d three distinct states such that $\alpha_{sud} = p(T_{sd} \in \mathcal{T}_{sd}^{u}) < 1$. Let $\bar{P}$ be the transition matrix of the same MC but where both states u and d are made absorbing, and whose entries are thus

$$\bar{P}_{ij} = \begin{cases} 0 & \text{if } \mathrm{i = u,\ d \ and\ i} \neq j, \\ 1 & \text{if } \mathrm{i = u,\ d\ and\ i = j}, \\ P_{ij} & \text{otherwise.} \end{cases} \quad (5)$$

Next, we define a second matrix P', obtained by a transformation of the matrix $\bar{P}$

$$P'_{ij} = \begin{cases} \frac{1-\alpha_{jud}}{1-\alpha_{iud}}\bar{P}_{ij} & \text{if } \alpha_{iud} \neq 1, \\ \bar{P}_{ij} & \text{otherwise.} \end{cases} \tag{6}$$

Lemma 2: (i) The matrix P' is stochastic and (ii) If T'_{sd} is a random trajectory defined on the MC whose transition probability matrix is P' then

$$H(T_{sd}|T_{sd} \notin \mathcal{T}^u_{sd}) = H(T'_{sd}).$$

Proof: (i) The matrix $\bar{P}$ is the transition probability matrix of a MC where the states u and d are absorbing. We can therefore introduce the vectors of absorption probability $\boldsymbol{a_u} = (a_{1u}, a_{2u}, \ldots, a_{nu})$ and $\boldsymbol{a_d} = (a_{1d}, a_{2d}, \ldots, a_{nd})$ where a_{iu} and a_{id} are, respectively, the probability of being absorbed by u and d, given that the trajectory starts at i. These vectors are eigenvectors of $\bar{P}$ associated with the unit eigenvalue [8, p. 227]

$$\bar{P}\boldsymbol{a_u} = \boldsymbol{a_u} \qquad \bar{P}\boldsymbol{a_d} = \boldsymbol{a_d}. \tag{7}$$

Moreover, as $\text{MC}_{\bar{P}}$ has only two absorbing states u and d, for all i, $a_{iu} = 1 - a_{id}$. Recall that for all i, $\alpha_{iud} = a_{iu}$ hence (6) can be written as

$$P'_{ij} = \begin{cases} \frac{a_{jd}}{a_{id}}\bar{P}_{ij} & \text{if } a_{id} \neq 0, \\ \bar{P}_{ij} & \text{otherwise.} \end{cases}$$

Note that all transitions leading to state u in $\text{MC}_{\bar{P}}$ will have zero probability in $\text{MC}_{P'}$. In fact, consider a state i such that $\bar{P}_{iu} > 0$ and $a_{id} > 0$. In the new matrix P', the probability of transition from i to u will be $P'_{iu} = a_{ud}\bar{P}_{iu}/a_{id}$, which is zero because $a_{ud} = 0$. Proving that P' is stochastic is now straightforward: first, the entries of P' are positive. Second, they are properly normalized and sum up to one. Indeed, if we consider a state i such that $a_{id} = 0$, we have that $\sum_j P'_{ij} = \sum_j \bar{P}_{ij} = 1$ whereas if $a_{id} \neq 0$, we have that

$$\begin{aligned} \sum_j P'_{ij} &= \sum_j \frac{a_{jd}}{a_{id}}\bar{P}_{ij} \\ &= \frac{1}{a_{id}}\sum_j \bar{P}_{ij}a_{jd} \\ &= \frac{1}{a_{id}}(\bar{P}\boldsymbol{a_d})_i = \frac{1}{a_{id}}a_{id} = 1 \end{aligned}$$

because of (7).

(ii) Let p and p' be the probability measures defined, respectively, for MC_P and $\text{MC}_{P'}$ on the same sample space $\mathcal{T}_{sd}$. Any trajectory from the set $\mathcal{T}_{sd}$ has the form $t_{sd} = sx_2 \ldots x_k d$.

If $t_{sd} \in \mathcal{T}^u_{sd}$,

$$p'(t_{sd}) = 0 \tag{8}$$

since we have constructed $\text{MC}_{P'}$ such that all transitions leading to state u have zero probability.

If $t_{sd} \notin \mathcal{T}^u_{sd}$, we have

$$\begin{aligned} p'(t_{sd}) &= P'_{sx_2}P'_{x_2x_3}\ldots P'_{x_kd} \\ &= \frac{a_{x_2d}}{a_{sd}}\bar{P}_{sx_2}\frac{a_{x_3d}}{a_{x_2d}}\bar{P}_{x_2x_3}\cdots\frac{a_{dd}}{a_{x_kd}}\bar{P}_{x_kd} \\ &= \frac{a_{dd}}{a_{sd}}\bar{P}_{sx_2}\bar{P}_{x_2x_3}\ldots\bar{P}_{x_kd}, \end{aligned} \tag{9}$$

but $a_{dd} = 1$ as the probability to be absorbed by state d, given that we have started at the same state, is 1. Moreover, we know from (5) that $P_{ij} = \bar{P}_{ij}$, for all $i \neq u, d$. As we have supposed that the trajectory t_{sd} does not admit either u or d as intermediate states, $\bar{P}_{sx_2}\bar{P}_{x_2x_3}\ldots\bar{P}_{x_kd} = P_{sx_2}P_{x_2x_3}\ldots P_{x_kd}$. Rewriting (9) yields

$$\begin{aligned} p'(t_{sd}) &= \frac{1}{a_{sd}}P_{sx_2}P_{x_2x_3}\ldots P_{x_kd} \\ &= \frac{p(t_{sd})}{1-a_{su}} \\ &= \frac{p(t_{sd})}{1-p(T_{sd}\in\mathcal{T}^u_{sd})} = p(t_{sd}|T_{sd}\notin\mathcal{T}^u_{sd}). \end{aligned} \tag{10}$$

Combining (8) and (10), we have therefore proven, for all $t_{sd} \in \mathcal{T}_{sd}$, that

$$p'(t_{sd}) = p(t_{sd}|T_{sd}\notin\mathcal{T}^u_{sd}). \tag{11}$$

Consequently, if the random variable T'_{sd} describes the trajectory between s and d in $\text{MC}_{P'}$, (11) implies that

$$H(T_{sd}|T_{sd}\notin\mathcal{T}^u_{sd}) = H(T'_{sd}).$$

■

For the particular case where $s = d$, we still can use Lemma 2 to express the conditional entropy $H_{ss|\bar{u}}$: we modify the MC by removing the incoming transitions of s and creating a new state s' that will inherit them. The conditional entropy $H_{ss|\bar{u}}$ in the original MC is equal to $H_{ss'|\bar{u}}$ in the modified one and, since $s \neq s'$, we can use Lemma 2 to express it.

Building on Lemmas 1 and 2, we can now state the main result of this paper: a general expression for the entropy of Markov trajectories conditional on multiple intermediate states.

Theorem 1: Let P be the transition probability matrix of a finite MC and $s\boldsymbol{u}d = su_1\ldots u_l d$ a sequence of states such that $p(T_{sd}\in\mathcal{T}^{\boldsymbol{u}}_{sd}) > 0$. Then, we have the following equality:

$$H(T_{sd}|T_{sd}\in\mathcal{T}^{\boldsymbol{u}}_{sd}) = \sum_{k=0}^{l-1} H(T'_{u_ku_{k+1}}) + H(T_{u_ld}), \tag{12}$$

where $u_0 = s$ and $T'_{u_ku_{k+1}}$ is a random trajectory defined on the MC whose transition probability matrix P'_k is defined as follows:

$$(P'_k)_{ij} = \begin{cases} 0 & \text{if } i = u_{k+1}, d \text{ and i} \neq j, \\ 1 & \text{if } i = u_{k+1}, d \text{ and } i = j, \\ P_{ij} & \text{if i} \neq u_{k+1}, d \text{ and } \alpha_{idu_{k+1}} = 1, \\ \frac{1-\alpha_{jdu_{k+1}}}{1-\alpha_{idu_{k+1}}}P_{ij} & \text{if i} \neq u_{k+1}, d \text{ and } \alpha_{idu_{k+1}} < 1. \end{cases} \tag{13}$$

Proof: The matrix P'_k is obtained from P using (13), which is equivalent to applying successively (5) and (6) where the starting, intermediate, and ending states are, respectively, u_k, d, and u_{k+1}. Therefore, using Lemma 2, we have

$$H(T'_{u_ku_{k+1}}) = H(T_{u_ku_{k+1}}|T_{u_ku_{k+1}}\notin\mathcal{T}^d_{u_ku_{k+1}})$$

for all $0 \leq k \leq l-1$. Consequently, we can write that

$$\begin{aligned} &\sum_{k=0}^{l-1} H(T'_{u_ku_{k+1}}) + H(T_{u_ld}) \\ &= \sum_{k=0}^{l-1} H(T_{u_ku_{k+1}}|T_{u_ku_{k+1}}\notin\mathcal{T}^d_{u_ku_{k+1}}) + H(T_{u_ld}), \end{aligned}$$

where $u_0 = s$. Using Lemma 1, we finally obtain

$$\sum_{k=0}^{l-1} H(T'_{u_k u_{k+1}}) + H(T_{u_l d}) = H(T_{sd} | T_{sd} \in \mathcal{T}_{sd}^{\boldsymbol{u}}).$$

■

Now, that we have derived a general expression for the entropy of Markov trajectories conditional on multiple states, we introduce, in the next section, a method that allows us to compute this expression.

IV. Entropy Computation

The closed-form expression for the entropy of Markov trajectories proposed by Ekroot and Cover [5] is valid only if the MC studied is irreducible. However, the Markov chain $\mathrm{MC}_{P'}$ obtained from MC_P after the transformations (5) and (6) is not necessarily irreducible: all transitions leading to state u have zero probability, which implies that possibly many states do not admit any path leading to d. Therefore, we need an expression for the entropy of Markov trajectories that is valid under milder conditions. In order to identify these conditions, we study the properties of $\mathrm{MC}_{P'}$. Let $\mathcal{S}$ be the set of all states in $\mathrm{MC}_{P'}$ and let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two subsets that partition $\mathcal{S}$ in the following manner:

$$\mathcal{S}_1 = \{i \in \mathcal{S} : a_{id} > 0\} \qquad \mathcal{S}_2 = \{i \in \mathcal{S} : a_{id} = 0\}.$$

The set $\mathcal{S}_1$ is closed as no one-step transition is possible from any state in $\mathcal{S}_1$ to any state in $\mathcal{S}_2$. In fact, if $i \in \mathcal{S}_1$ and $j \in \mathcal{S}_2$, (6) yields that $P'_{ij} = \bar{P}_{ij} a_{jd} / a_{id} = 0$. Clearly, all trajectories leading to state d are composed of states belonging to $\mathcal{S}_1$. Now, we propose a closed-form expression for the entropy of Markov trajectories that is valid under the weaker condition that the destination state d can be reached from any other state of the MC. Moreover, we prove that the trajectory entropy can be expressed as a weighted sum of local entropies. We also provide an intuitive interpretation of the weights.

Lemma 3: Let P be the transition probability matrix of a finite state MC such that there exists a path with positive probability from any state to a given state d. Let Q_d be a submatrix of P obtained by removing the d^{th} row and column of P

$$P = \left(\begin{array}{c|c} \mathbf{Q}_{\mathrm{d}} & \begin{array}{c} P_{1d} \\ \vdots \end{array} \\ \hline P_{d1} \quad \cdots & P_{dd} \end{array} \right).$$

For any state $s \neq d$, the trajectory entropy H_{sd} can be expressed as

$$H_{sd} = \sum_{k \neq d} ((I - Q_d)^{-1})_{sk} H(P_{k\cdot}), \tag{14}$$

where $H(P_{k\cdot})$ is the local entropy of state k.

Proof: First, observe that the matrix Q_d is a submatrix of P corresponding to all states except state d and that we use Q_d to derive the entropy of all trajectories ending at d. Applying the chain rule for entropy, we express the entropy of a trajectory as the entropy of the first step plus the conditional entropy of the rest of the trajectory given this first step

$$H_{sd} = H(P_{s\cdot}) + \sum_{k \neq d} P_{sk} H_{kd}.$$

We expand this equality further by recursively expanding the entropy H_{kd} as follows:

$$\begin{aligned} H_{sd} &= H(P_{s\cdot}) + \sum_{k \neq d} P_{sk} \left(H(P_{k\cdot}) + \sum_{k' \neq d} P_{kk'} H_{k'd} \right) \\ &= H(P_{s\cdot}) + \sum_{k \neq d} P_{sk} H(P_{k\cdot}) \\ &\quad + \sum_{k \neq d} P_{sk} \sum_{k' \neq d} P_{kk'} H_{k'd} \\ &= H(P_{s\cdot}) + \sum_{k \neq d} P_{sk} H(P_{k\cdot}) + \sum_{k \neq d} P_{sk} \sum_{k' \neq d} P_{kk'} \\ &\quad \cdot \left(H(P_{k'\cdot}) + \sum_{k'' \neq d} P_{k'k''} \left(H(P_{k''\cdot}) + \dots \right) \right) \\ &= H(P_{s\cdot}) + \sum_{k \neq d} \left(\sum_{i=1}^{\infty} ({Q_d}^i)_{sk} \right) H(P_{k\cdot}) \\ &= \sum_{k \neq d} \left(\sum_{i=0}^{\infty} ({Q_d}^i)_{sk} \right) H(P_{k\cdot}), \end{aligned} \tag{15}$$

with ${Q_d}^0 = I$.

Observe that the matrix Q_d describes the MC as long as it does not reach state d. Moreover, the matrix Q_d has a finite number of states and there is a path with positive probability from each state to state d. As a consequence, the Markov process will enter state d with probability 1, i.e., $\lim_{n \to \infty} {Q_d}^n = O$ (zero matrix). In addition, since

$$(I - Q_d)(I + Q_d + {Q_d}^2 + \dots + {Q_d}^{n-1}) = I - {Q_d}^n,$$

we can easily verify that

$$\sum_{i=0}^{\infty} {Q_d}^i = (I - Q_d)^{-1}. \tag{16}$$

Replacing (16) in (15), we have

$$H_{sd} = \sum_{k \neq d} ((I - Q_d)^{-1})_{sk} H(P_{k\cdot}).$$

■

We have shown that the entropy of a family of trajectories can be expressed as a weighted sum of the states' local entropies. The weights are given by the matrix $(I - Q_d)^{-1}$. In the Markovian literature, the matrix $(I - Q_d)^{-1}$ is referred to as the fundamental matrix [8], [9]. In fact, the $(sk)^{th}$ element of the fundamental matrix (defined with respect to the destination state d) can be seen as the expected number of visits to the state k before hitting the state d, given that we started at state s. As a result, the entropy of the random trajectory T_{sd} is the sum over the chain states of the expected number of visits to each state multiplied by its local entropy. This is a remarkable observation since it links a global quantity, which is the trajectory entropy, to the local entropy at each state.

Recall that in the example shown in Fig. 1, we found that the entropy of the trajectory T_{15} is equal to 1.56 bits. We can retrieve this result by computing the fundamental matrix with respect to state 5. The $(ij)^{th}$ element of this matrix is equal to the expected number of visits to state j before hitting state 5, given that we started at state i. Multiplying the first row of the fundamental matrix $(1, 0.625, 0.75, 0.375)$ by the column vector of local entropies $(0.81, 0, 1, 0)$ yields $H_{15} = 1 \times 0.81 + 0.75 \times 1 = 1.56$ bits.

A. Algorithm

The following algorithm defines the set of steps to compute the entropy of Markov trajectories conditional on a set of intermediate states:

Intput: Matrix of transition probability P, source state s, destination state d, sequence of intermediate states $\boldsymbol{u} = u_1 \dots u_l$

Output: $H_{sd|u_1 \dots u_l}$

```
u_0 ← s
for k = 0 to l − 1 do
    Compute P'_k from P using (13)
    Compute H(T'_{u_k u_{k+1}}) from P'_k using Lemma 3
    H_{u_k u_{k+1}|d̄} ← H(T'_{u_k u_{k+1}}) {Lemma 2}
end for
Compute H_{u_l d} from P using Lemma 3
H_{sd|u_1...u_l} = Σ_{k=0}^{l−1} H_{u_k u_{k+1}|d̄} + H_{u_l d} {Lemma 1}
return H_{sd|u_1...u_l}
```

The worst-case running time for the algorithm is $\mathcal{O}(lN^3)$ where N is the number of states of MC$_P$, and l the length of the sequence of intermediate states $\boldsymbol{u}$. This complexity is dominated by the cost of computing the inverse of the matrix $(I - Q_d)$, which is needed to compute the entropy H_{sd} in (14). However, since we need only the s^{th} row of the matrix $(I - Q_d)$ to compute the trajectory entropy H_{sd}, we can solve a system of—potentially sparse—linear equations. Moreover, many iterative methods [10, p. 508] take advantage of the structure of the matrix representing the system of linear equations in order to solve them efficiently.

Coming back to the example shown in Fig. 1, we use the algorithm above to compute the conditional entropy $H_{15|3} = 1$ bit. We leave no ambiguity about the trajectory T_{15} when we condition on both states 3 and 2 and find that $H_{15|3,2} = H_{13|\bar{5}} + H_{32|\bar{5}} + H_{25} = 0$ bits.

Conditioning on a Set of States: In this paper, we focused on computing the entropy of Markov trajectories conditional on a *sequence* of states. A natural extension is the computation of this entropy conditional on a *non ordered* set of states. Finding a general expression for this conditional entropy appears very hard and there is no simple relation linking it to the entropy conditional on a sequence. We provide an example, shown in Fig. 2, that illustrates an interesting and counter-intuitive result about conditioning on a set of states. Intuitively, we would expect that the entropy of a random trajectory conditional on a sequence of states is always less than the entropy of the same trajectory conditional on the set formed by these states. However, this is not true. We take the MC shown in Fig. 2 as an example and we compute, using Theorem 1, the entropy of the random trajectory T_{15} conditional on going through the sequence of intermediate states $(3, 2)$

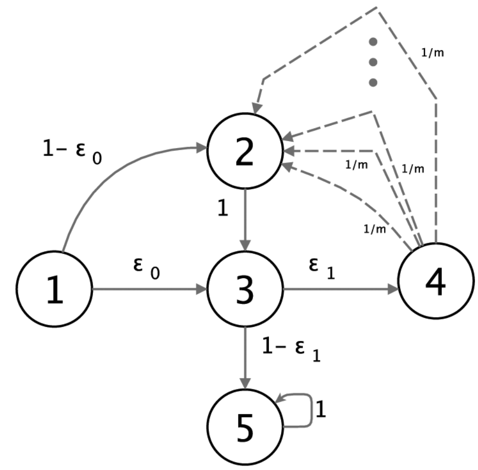

Fig. 2. Markov chain annotated with the transition probabilities. The dashed lines between states 4 and 2 represent the m equiprobable paths leading from state 4 to state 2. We choose $0 < \epsilon_1 < 1$ and $m \geq 1$ to guarantee that $|\mathcal{T}_{15}| > 0$ and that $p(T_{15} \in \mathcal{T}_{15}^{3,2}) > 0$.

$$\begin{aligned} H_{15|32} &= H_{13|\bar{5}} + H_{32|\bar{5}} + H_{25} \\ &= h(\epsilon_0) + \log m + H_{35}, \end{aligned} \tag{17}$$

where $h(\epsilon_0)$ is the entropy of a Bernoulli random variable with success probability ϵ_0. To compute the entropy of the random trajectory T_{15} conditional on going through the set of states $\{2, 3\}$, we apply the chain rule for entropy and express the entropy of a trajectory as the entropy of the first two steps plus the conditional entropy of the rest of the trajectory given these first two steps

$$\begin{aligned} H_{15|\{2,3\}} = &\, h\left(\frac{\epsilon_0\epsilon_1}{1 - \epsilon_0(1 - \epsilon_1)}\right) + \frac{\epsilon_0\epsilon_1}{1 - \epsilon_0(1 - \epsilon_1)} H_{45} \\ &+ \frac{1 - \epsilon_0}{1 - \epsilon_0(1 - \epsilon_1)} H_{35}. \end{aligned}$$

Since $H_{45} = \log m + H_{25} = \log m + H_{35}$, we have that

$$\begin{aligned} H_{15|\{2,3\}} = &\, h\left(\frac{\epsilon_0\epsilon_1}{1 - \epsilon_0(1 - \epsilon_1)}\right) + \frac{\epsilon_0\epsilon_1}{1 - \epsilon_0(1 - \epsilon_1)} \log(m) \\ &+ H_{35}. \end{aligned} \tag{18}$$

Using (17) and (18), we can write

$$\begin{aligned} H_{15|32} - H_{15|\{2,3\}} = &\, h(\epsilon_0) - h\left(\frac{\epsilon_0\epsilon_1}{1 - \epsilon_0(1 - \epsilon_1)}\right) \\ &+ \frac{1 - \epsilon_0}{1 - \epsilon_0(1 - \epsilon_1)} \log m. \end{aligned}$$

This difference can therefore be lower bounded by

$$H_{15|32} - H_{15|\{2,3\}} \geq -1 + \frac{1-\epsilon_0}{1-\epsilon_0(1-\epsilon_1)} \log m.$$

As a consequence, if $\log m > 1 + \epsilon_0\epsilon_1/1 - \epsilon_0$, the entropy of the random trajectory T_{15} conditional on going through the sequence $(3, 2)$ is strictly greater than the entropy of the same trajectory conditional on going through the set of states $\{2, 3\}$. The reason is that conditioning on the sequence $(3, 2)$ implies that the random trajectory T_{15} is composed of a random subtrajectory T_{42} whose entropy can be made arbitrary large by increasing the parameter m. More generally, this example illustrates the absence of a simple relation between the entropy of random trajectories conditional on a sequence of states and the entropy of the same trajectory conditional on the set formed by these states.

V. Conclusion

In this paper, we address the problem of computing the entropy of conditional Markov trajectories. We propose a method based on a transformation of the original Markov chain into a Markov chain that yields the desired conditional entropy. We also derive an expression that allows us to compute the entropy of Markov trajectories, under conditions weaker than those assumed in [5]. Furthermore, this expression links the entropy of Markov trajectories—a global quantity—to the local entropy of states.

These results have applications in various fields including mobility privacy of the users of online services. In fact, using our framework, we are able to quantify the predictability of a user's mobility and its evolution with locations updates: we represent a location as a state of a Markov chain. A sequence of visited locations is therefore a Markovian trajectory, and location-updates amount to conditioning this trajectory on a set of intermediate states. In this setting, we can quantify the evolution of the user's mobility predictability as she/he discloses some of the locations she/he visited by computing the entropy of conditional Markov trajectories. Consequently, users are empowered with an objective technique to protect their privacy: they are able to anticipate the evolution of their mobility predictability as they reveal a subset of the locations they visited.

Acknowledgment

The authors would like to thank O. Lévêque and E. Telatar for their feedback about this paper.

References

[1] L. Yen, M. Saerens, A. Mantrach, and M. Shimbo, "A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances," in *Proc. 14th SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 785–793.

[2] S. Lloyd and H. Pagels, "Complexity as thermodynamic depth," *Ann. Phys.*, vol. 188, no. 1, pp. 186–213, 1988.

[3] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, "Localization of the maximal entropy random walk," *Phys. Rev. Lett.*, vol. 102, p. 160602, 2009.

[4] M. Saerens, Y. Achbany, F. Fouss, and L. Yen, "Randomized shortest-path problems: Two related models," *Neural Comput.*, vol. 21, pp. 2363–2404, 2009.

[5] L. Ekroot and T. Cover, "The entropy of Markov trajectories," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1418–1421, Jul. 1993.

[6] R. B. Dial, "A probabilistic multipath traffic assignment model which obviates path enumeration," *Transp. Res.*, vol. 5, pp. 83–111, 1971.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[8] W. J. Stewart, *Probability, Markov Chains, Queues, and Simulation*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[9] J. Kemeny and J. Snell, *Finite Markov Chains*. New York, NY, USA: Van Nostrand, 1969.

[10] G. H. Golub and C. F. van Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in Mathematical Sciences, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

Mohamed Kafsi (S'12) is a Ph.D. student in the School of Computer and Communication Sciences at EPFL. His research interests lie at the intersection of graph theory, information theory, and data mining. He received the M.S. and B.S. degrees in communication systems from EPFL. During his Bachelor's studies, he spent one year at the Electrical and Computer Engineering Department of Carnegie Mellon University (CMU). He is a member of the winning team of the 2012 Nokia Mobile Data Challenge.

Matthias Grossglauser (M'95) is an Associate Professor in the School of Computer and Communication Sciences at EPFL. He received his Diplôme d'Ingénieur en Systèmes de Communication degree from EPFL in 1994, the M.Sc. degree from the Georgia Institute of Technology in 1994, and the Ph.D. from the University Pierre et Marie Curie (Paris 6) in 1998. His research interests are in social and information networks, privacy, mobile and wireless networking, and network traffic measurement and modeling. He received the 1998 Cor Baayen Award from the European Research Consortium for Informatics and Mathematics (ERCIM), the IEEE INFOCOM 2001 Best Paper Award, and the 2006 CoNEXT/SIGCOMM Rising Star Award. He served on the editorial board of IEEE/ACM Transactions on Networking, and on numerous Technical Program Committees. From 2007–2010, he was with the Nokia Research Center (NRC) in Helsinki, Finland, holding the positions of Laboratory Director, then of Head of a tech-transfer program focused on data mining, analytics, and machine learning. In addition, he served on Nokia's CEO Technology Council, a technology advisory group reporting to the CEO. From 2003–2007, he was an Assistant Professor at EPFL. From 1998 to 2002, he was a Senior, then Principal Member of Research Staff in the Networking and Distributed Systems Laboratory at AT&T Research in New Jersey. From 1995 to 1998, he was a Ph.D. student at INRIA Sophia Antipolis, France.

Patrick Thiran (S'89–M'96–SM'12) received the electrical engineering degree from the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1989, the M.S. degree in electrical engineering from the University of California at Berkeley, USA, in 1990, and the Ph.D. degree from EPFL, in 1996. He is a Full Professor at EPFL. He became an Adjunct Professor in 1998, an Assistant Professor in 2002, an Associate Professor in 2006, and a Full Professor in 2011. From 2000 to 2001, he was with Sprint Advanced Technology Labs, Burlingame, CA, USA. His research interests include communication networks, performance analysis, dynamical systems, and stochastic models. He is currently active in the analysis and design of wireless and PLC networks, in network monitoring, and in data-driven network science. Dr. Thiran served as an Associate Editor for the IEEE Transactions on Circuits and Systems in 1997–1999, and as an Associate Editor for the IEEE/ACM Transactions on Networking in 2006–2010. He was the recipient of the 1996 EPFL Ph.D. award and of the 2008 Crédit Suisse Teaching Award.