How to Reduce Unnecessary Noise in Targeted Networks

Giacomo Aletti and Diane Saada

Abstract This work is a review of previous works on the stopping laws in networks. Among other results, we show a non combinatorial method to compute the stopping law, the existence of a minimal Markov chain without oversized information, the existence of a polynomial algorithm which projects the Markov chain onto the minimal Markov chain. Several applied examples are presented.

1 Introduction

Consider a finite state Markov chain, with state space E. The process is stopped when it reaches a sub-class T of E. It turns out that one does not need the whole information carried by its transition matrix in order to compute the law of reaching this class. The following paper, which is a compilation of several papers, deals with how to reduce this unnecessary information, first in real time and then for large times. An extension is also given for R-networks in [3]. The problem of finding general closed-forms for different kinds of waiting problems is widely studied, following various approaches. See, for example, [9] in the case of Bernoulli trials, [6], [10], [1] and [12] for its extensions to Markov-dependent trials, and [13] for another methodology. A new approach was given in [2–5], where it was proved that there exists an optimal projection for any given Markov problem which leaves the probability of reaching the target set unchanged. A simple ε-approximation of this projection exists, provided the system has evolved for a sufficient amount of time

G. Aletti
ADAMSS and Dipartimento di Matematica, Università di Milano, 20133 Milan, Italy
giacomo.aletti@unimi.it

D. Saada
Department of Statistics, Hebrew University, Jerusalem, Israel
dianes@mscc.huji.ac.il

A.K. Naimzada et al. (eds.), *Networks, Topology and Dynamics*,
Lecture Notes in Economics and Mathematical Systems 613,

and some conditions on the Markov chain are satisfied. In the framework of [2–4], a compatible projection is an equivalence relationship S on the indexing set E s.t.:

- $\forall e_i \in T, e_i S e_j \iff e_j \in T$.
- For any $\{e_i, e_j, e_k\} \subseteq E$: $e_i S e_j$, we have

$$\sum_{e_l S e_k} P(e_i, e_l) = \sum_{e_l S e_k} P(e_j, e_l),$$

  where T (the absorbing target class) and $P : E \times E \to R$ are given (P is the Markov matrix of the network and R is a semiring, see [3]).

In [2, 3] it was proved the existence of a polynomial-time algorithm which reaches the minimum Markov network. In [5], the question of a further reduction is posed, when time tends to infinity. An asymptotic conditional law of exit will exist, according to the shape of the transition submatrix which corresponds to the states leading to the target class. The methodology is based on spectral theory for non-negative matrices and in particular on the Perron–Frobenius theorem. The framework in [2–5] regards a huge class of problems which occur in many real situations. We recall here how this class of problem may appear:

1. In *finance* the filter rule for trading is a special case of the Markov chain stopping rule suggested in [4] (see, e.g., [11]).
2. "When enough is enough!" for example, an insured has an accident only occasionally in a while. How many accidents in a specified number of years should be used as a stopping time for the insured (in other words, when it should be discontinued the insurance contract).
3. *State dependent Markov chains.* Namely, the transition probabilities are given in terms of the history. In many situations, the matrix of the embedded problem may be reduced.
4. *Medical sciences*. Given that the length of a menstrual cycle has a known distribution, what is the probability that the length of a woman's menstrual cycle is the same three consecutive times?
5. *Small-world Networks*. Given one of the networks as in Fig. 1 (either as Markov network or as a graph), is it possible to reduce it in polynomial time and to preserve the law of reaching a given absorbing state?

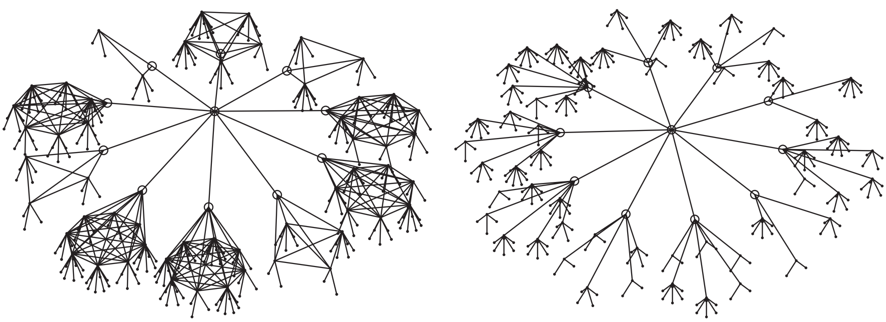


Fig. 1 Networks that may be compressed (see [4])

There are of course many other such examples (e.g., records: Arnold et al. [7] and optimization: Cairoli and Dalang [8]).

1.1 A Combinatorial Problem

Let $X = \{X\tau, \tau \in \mathbb{N}\}$ be a Markov chain on a finite state space $E = \{e_1, \ldots, e_n\}$:

$$\left.\begin{array}{c|ccc} & e_1 & \ldots & e_n \\ \hline e_1 & p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ e_1 & p_{n,1} & \ldots & p_{n,n} \end{array}\right\} =: P\,.$$

The process is stopped when it reaches one of some given states $T := (e_{n_i})_{i=1}^{k} \subseteq E$. For sake of clarity, we suppose $F = \{e_1, \ldots, e_k\}$ and $E \setminus T = \{e_{k+1}, \ldots, e_n\}$. To compute the law of stopping, we may consider a new Markov chain $X' = \{X'\tau, \tau \in \mathbb{N}\}$ on $T \cup (E \setminus T)$:

$$\left.\begin{array}{cc|cccccc} & & & T & & & E\setminus T & \\ & & e_1 & \ldots & e_k & e_{k+1} & \ldots & e_n \\ \hline & e_1 & \mathbf{1} & \ldots & \mathbf{0} & 0 & \ldots & 0 \\ T & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & e_k & \mathbf{0} & \ldots & \mathbf{1} & 0 & \ldots & 0 \\ & e_{k+1} & p_{k+1,1} & \cdots & p_{k+1,k} & p_{k+1,k+1} & \cdots & p_{k+1,n} \\ E \setminus T & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & e_n & p_{n,1} & \cdots & p_{n,k} & p_{n,k+1} & \cdots & p_{n,n} \end{array}\right\} =: \widehat{P}\,.$$

Thus, the probability of reaching T by time τ is reduced to the computation of the τ-th power of $\widehat{P}$:

$$\mathbb{P}(\cup_{i=1}\tau\{X_i \in T\}) = \mathbb{P}(\{X'\tau \in T\}) = \overbrace{(p_1^0, \ldots, p_n^0)}^{\mathbf{p_0}} \quad (\widehat{P})^{\tau} \quad (\overbrace{1, \ldots, 1}^{k \text{ terms}}, \overbrace{0, \ldots, 0}^{n-k \text{ terms}})'\,. \tag{1}$$

There exists a trivial reduction which preserves the above calculation for any $\tau \in \mathbb{N}$ and initial distribution $\mathbf{p_0}$:

$$\mathbb{P}(\cup_{i=1}\tau\{X_i \in T\}) = \left(\sum_{i=1}^{k} p_i^0, p_{k+1}^0, \ldots, p_n^0\right) \begin{pmatrix} \mathbf{1} & 0 & \ldots & 0 \\ \sum_{i=1}^{k} p_{k+1,i} & p_{k+1,k+1} & \cdots & p_{k+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{k} p_{n,i} & p_{n,k+1} & \cdots & p_{n,n} \end{pmatrix}^{\tau} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{2}$$

2 Target Problems

We begin the mathematical framework in this section with an example. Suppose (2) is written in the following way

$$\left(p_1^0, p_2^0, p_3^0, p_4^0, p_5^0\right) \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{12} & 0 & \frac{1}{6} \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & \frac{3}{8} & \frac{3}{8} & \frac{1}{4} & 0 \end{pmatrix}^{\tau} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} . \tag{3}$$

We define a problem of compression of redundant information, in terms of equivalence relationships on E (see [2–4]). First, we extend P to $\mathbb{P}_P$ in the classical way

$$\mathbb{P}_P : E \times \mathfrak{P}(E) \to \mathbb{R}_+ \cup \{0\} ,$$

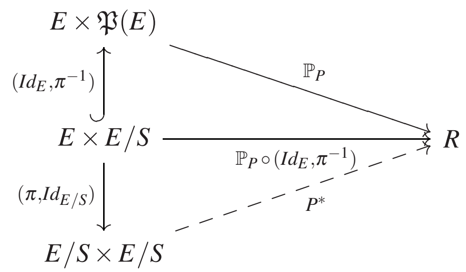
where $\mathfrak{P}(E)$ is the set of subsets of E and $\mathbb{P}_P(e,A) = \sum_{e_i \in A} P(e,e_i)$. Obviously, for each $e \in E$, $\mathbb{P}_P(e,\cdot)$ is a probability on $(E,\mathfrak{P}(E))$ that gives the conditional probability of reaching $\cdot$ given that we are in the state e.

As stated in the introduction, a compatible projection is an equivalence relationship S on the indexing set E s.t.:

1. $\forall e_i \in T,\, e_i S e_j \iff e_j \in T$.
2. For any $\{e_i, e_j, e_k\} \subseteq E$: $e_i S e_j$, we have

$$\sum_{e_l S e_k} P(e_i, e_l) = \sum_{e_l S e_k} P(e_j, e_l).$$

1 and 2 are satisfied if and only if the matrix $P^* : E/S \times E/S \to \mathbb{R}_+ \cup \{0\}$ such that the following diagram commutes, is well-defined:[1]



(4)

For what concerns (3), $\nexists j \in \{2,3,4,5\}$ such that $e_j S e_1$ by 1. Note that finding a nontrivial projection is not a local search. For example, we have $P(e_4,e_2) \neq P(e_5,e_2)$ but $P(e_4,e_2) + P(e_4,e_3) = P(e_5,e_2) + P(e_5,e_3)$, which means that $e_4 S e_5$ may be found if *we know* that $e_2 S e_3$.

[1] Here, $\pi : E \to E/S$ is the canonical projection.

Now, a nontrivial projection is given by

$$E/S = \{\underbrace{T = \{e_1\}}_{f_1}, \underbrace{\{e_2, e_3\}}_{f_2}, \underbrace{\{e_4, e_5\}}_{f_3}\}.$$

Accordingly, the new matrix, associated with the projected states $E/S = \{f_1, f_2, f_3\}$, is given by

$$\left.\begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline f_1 & 1 & 0 & 0 \\ f_2 & \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ f_2 & 0 & \frac{3}{4} & \frac{1}{4} \end{array}\right\} =: P^* .$$

The new Markov problem $(\{f_1, f_2, f_3\}, P^*)$ carries all the necessary information for the target problem. In fact, the states in f_i play all together with respect to the target, like if they were the same point. For example, problem (3) becomes

$$\left(p_1^0, p_2^0 + p_3^0, p_4^0 + p_5^0\right) \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}^{\tau} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} . \tag{5}$$

In the general case, we deal with a (at most) countable indexing set E (in (3), $E = \{e_1, \ldots, e_5\}$). We then take a suitable space $\mathcal{M}_E$ of matrices on E (again, for what concerns (3), $\mathcal{M}_E$ is the space of 5×5 stochastic matrices); in general it will be a monoid $(\mathcal{M}_E, \cdot)$ where $\cdot$ can be seen as the matrix multiplication. The function $\mathbb{P}$, as defined above, is assumed to, and plays the role of the conditional probability of reaching any family of states starting from a given state. A target set $T \subseteq E$ is fixed, and it is assumed to be an absorbing class. A target problem is therefore a triple (E, P, T) where:

- E is an, at most countable, indexing set.
- $P \in \mathcal{M}_E$ is a given matrix and $\mathbb{P}_P$ is well-defined.
- T is an absorbing class: $\mathbb{P}_P(t, T) = 1$ and $\mathbb{P}_P(t, E \setminus T) = 0$, for any $t \in T$.

We are dealing here with Markov matrices only, and we leave more general extensions to [3]. In [3], an extended framework includes graph connection, as well. Note that one may always change the matrix P as in (2) so that T is an absorbing class, by defining

$$\widehat{P}(e_1, e_2) = \begin{cases} P(e_1, e_2), & \text{if } e_1 \notin T; \\ 1, & \text{if } e_1 \in T \text{ and } e_1 = e_2; \\ 0, & \text{if } e_1 \in T \text{ and } e_1 \neq e_2. \end{cases}$$

Let $\widetilde{E}$ be the set of all equivalence relations on E. Let $V, S \in \widetilde{E}$. We say that $V \preceq S$ if $a_1 V a_2$ implies $a_1 S a_2$ (if you think E as the set of all men and V is "belonging to the same state" while S is "belonging to the same continent", then $V \preceq S$). An equivalence relationship $S \in \widetilde{E}$ is called *compatible projection with respect to the target problem* (E, P, T) if:

1. $\forall e \in T,\, e S e_j \iff e_j \in T$ (i.e., the target set defines an equivalence class).
2. There exists $P^* \in \mathcal{M}_{E/S}$ such that (4) commutes.

We call $\mathcal{S} = \mathcal{S}(E,P,T)$ the set of all compatible projections.

The previous definition of compatible projection states when it is possible to project our target problem (E,P,T) into the smaller one $(E/S, P^*, t = \pi(T))$, without loosing necessary information (see [4]). In this framework, we can state the following general result.

Theorem 1 ([3, 4]). *For any target problem (E,P,T), there exists the optimal projection: $\exists S \in \mathcal{S}$ s.t. $V \preceq S$, $\forall V \in \mathcal{S}$.*

For example, note that the compatible projection $S \in \mathcal{S}$ which projects (3) into (5) is optimal. In fact, suppose there exists $S^* \in \widetilde{E}$ such that (a) $\{f_1\} \in E/S^*$, by 1 above; (b) $S \precnsim S^*$. (a) and (b) imply $E/S^* = \{\{e_1\}, \{e_2, e_3, e_4, e_5\}\}$, which is not a compatible projection, since $1/2 = \mathbb{P}_P(e_2, \{e_1\}) \neq \mathbb{P}_P(e_4, \{e_1\}) = 0$.

2.1 Target Algorithm

The proof for the existence of the optimal solution was based in [3, 4] on the fact that the set of compatible projections $\mathcal{S}$ has its $\preceq$-join in $\widetilde{S}$.

This proof is useless in practice when the Markov chain is so big that a search in $\widetilde{E}$ can be impracticable. In fact, as stated in the previous section, searching for a compressing map is not a local search and it appears as a non-polynomial search, in the sense that we have to look at the whole set of equivalent relations on E. In [2, 3] it was proved the existence of a polynomial algorithm which reaches the optimal projection. We give here the algorithm and we state this result in Theorem 2. The idea is to reach the optimal projection E/S – unknown – starting from a trivial and known relation $M_E \in \widetilde{E}$, given by the problem. $M_E \in \widetilde{E}$ is defined by the relation "being or not a member of T": for any $(e_i, e_j) \in E \times E$,

$$e_i M_E\, e_j \iff \{e_i, e_j\} \subseteq T \text{ or } \{e_i, e_j\} \subseteq (E \setminus T).$$

It is clear that M_E is not in general a compatible projection (see, for example, S^* at the end of the previous section). By definition, it is obvious that $S \preceq M_E$, $\forall S \in \mathcal{S}$ and hence, if M_E is compatible, then it is optimal.

We denote here by $F\pi$ the optimal equivalence unknown map, and we build a monotone operator $\mathcal{F}$ on $\widetilde{E}$ which will reach $E/F\pi$ starting from E/M_E. $\mathcal{F} : \widetilde{E} \to \widetilde{E}$ is defined as follows: for any $S \in \widetilde{E}$, let $s_1, s_2, \ldots$ be the classes of equivalence of E induced by S. For any $(e_l, e_k) \in E \times E$, define

$$e_l \mathcal{F}_{s_i} e_k \iff \mathbb{P}_P(e_l, s_i) = \mathbb{P}_P(e_k, s_i)$$
$$\mathcal{F}(S) = \bigcap_{i=1,2,\ldots} \mathcal{F}_{s_i} \cap S\,.$$

$\mathcal{F}(S)$ is a new equivalence relation, that defines, in consequence, new classes of equivalence. Two states belong to the same new class if they have the same behavior towards the classes of S. If this step does not define the compatible projection, then we go a step further applying $\mathcal{F}$ to the classes of $\mathcal{F}(S)$. For example, take $E/M_E = \{\{1\},\{2,3,4,5\}\}$ as in (3). For $s_1 = \{1\}$ we have

$$\mathbb{P}(1,\{1\}) = 1\,, \quad \mathbb{P}(2,\{1\}) = \mathbb{P}(3,\{1\}) = 1/2\,, \quad \mathbb{P}(4,\{1\}) = \mathbb{P}(5,\{1\}) = 0$$

and hence $E/\mathcal{F}(s_1) = \{\{1\},\{2,3\},\{4,5\}\}$. Note that $E/\mathcal{F}(s_1) = E/\mathcal{F}(s_2) \subseteq E/M_E$ and hence $E/\mathcal{F}(M_E) = \{\{1\},\{2,3\},\{4,5\}\}$. The new relationship $\mathcal{F}(M_E)$ is a fixed point for $\mathcal{F}$ and it is also the optimal relationship.

This leads to the following theorem.

Theorem 2 ([2, 3]). *For any target problem (E,T,P), there exists $m = m(E,T,P)$ s.t. $m \leq N-2$ and*

$$F\pi = \underbrace{\mathcal{F}\circ\mathcal{F}\circ\cdots\circ\mathcal{F}}_{m\ times}(M_E)\,,$$

where N is the cardinality of $E/F\pi$.

For what concerns the problem (3), we already knew that $|E/F\pi| = 3$ and hence $m \leq 1$ by Theorem 2. We have indeed noted that $\mathcal{F}(\mathcal{F}(M_E)) = \mathcal{F}(M_E)$. In fact, in the proof (see [2, 3]), it is also shown that $F\pi$ is the unique fixed point for a suitable restriction of $\mathcal{F}$ and this algorithm "works" on this restriction.

Remark 1. Note that the operator $\mathcal{F}$ may be computed in a $|E|$-polynomial time. Theorem 2 ensures that

$$\underbrace{\mathcal{F}\circ\mathcal{F}\circ\cdots\circ\mathcal{F}}_{\text{at most } |E/F\pi|-2 \text{ times } (\leq|E|)}$$

will reach F, given any triple (E,T,P). A MATLAB version of such an algorithm for multitarget $\mathbf{T}$ may be downloaded at http://www.mat.unimi.it/ aletti.

3 Large Time Projections

Suppose now that the Markov chain $\{X\tau, \tau \in \mathbb{N}\}$ is stationary on a finite set E and denote by P its transition matrix and by μ_0, the initial probability measure. Let A be the class of transient states which lead to T (the target class), and let $T_\infty = \{i \in E; P\tau_{ij} = 0, \forall j \in T, \forall \tau\}$ the class of the remaining states. As before, the transition matrix can be decomposed as follows:

$$P = \left(\begin{array}{cc|c} 1 & 0 & \\ 0 & 1 & 0 \\ \hline & & \\ v_\infty & v & A \end{array}\right),$$

v_∞ (resp. v) is the vector of probabilities of hitting T_∞ (resp. T) from A, and A is the sub-matrix of the states lying in A. We suppose that $A \geq 0$ -that is, if $A = (a_{ij})_{(i,j)\in\{1,\cdots,n\}^2}$ is a square matrix, then every a_{ij} is nonnegative- and that $A \neq 0$. This matrix also satisfies $\lim \tau A\tau = 0$. Here, $\underline{1}$ denotes the vector whose components are all equal to one. In order to know whether an asymptotic reduction can be done or not, we had to check whether the following limit exists

$$\lim_{\tau\to\infty} \mathbb{P}(X_{\tau+1} \in T | X\tau \in A) \quad \text{for any } \mu_0$$

and under which conditions this limit is independent of μ_0. It is easy to show that this limit is in fact equal to

$$\lim_{\tau\to\infty} \frac{\mu_{0|A}^T A\tau v}{\mu_{0|A}^T A\tau \underline{1}}.$$

$\mu_{0|A}$ is the trace on A of the initial probability μ_0. We suppose in the following that $\mu_{0|A} > 0$. We can decompose A into disjoint classes of communication, where i communicates with j if $i = j$ or if i leads to j and vice versa, and we obtain $A = \cup_{i=1}^N C_i^*$. We denote by $p(i)$ the period of state i. We recall that $p(i)$ is defined as the greatest common divisor of all integers $n \geq 1$ for which $A^n_{ii} > 0$, when it exists, otherwise we set $p(i) = \infty$. All the elements of a same class have the same period. The whole discussion in the following on the existence (and uniqueness) of a solution will depend upon the number of classes and their periodicity.

3.1 The Irreducible Case

This is the case when all the states communicate, and so $N = 1$. The existence of a limit depends on the three different cases for the common period of the states; either (1) $p(1) = 1$, or (2) $p(1) = k > 1$, or (3) $p(1) = \infty$, where 1 denotes the first state in the matrix A.The third case is not to be taken into account, see [5].

Theorem 3. *Suppose A is irreducible and aperiodic, then*

$$\lim_{\tau\to\infty} \frac{\mu_{0|A}^T A\tau v}{\mu_{0|A}^T A\tau \underline{1}} \tag{6}$$

exists for any initial probability, is independent of μ_0 *and is equal to* $\frac{(f_0)^T v}{(f_0)^T \underline{1}}$, *where* f_0 *is the first left eigenvector of A. If A is irreducible and periodic, then* (6) *exists if and only if the asymptotic probabilities of exit from each class of periodicity are equal.*

As a consequence, one obtains at infinity the behavior outlined in Fig. 2.

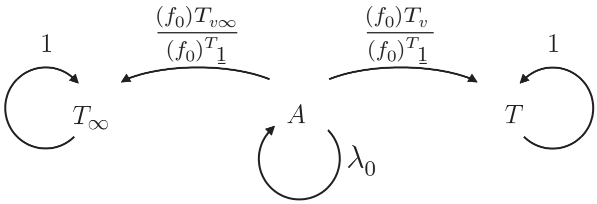


Fig. 2 Behavior when A is irreducible and aperiodic

3.2 The Reducible Case

In the reducible case, A is decomposed into $N > 1$ disjoint classes C_i^*, $i = 1,\ldots,N$. As the classes are disjoint, when the chain exits one class, it does not go back to it. It follows that we can reorder the matrix A so that it will be equal to an upper triangular (non-negative) matrix with block square irreducible matrices on its diagonal, each corresponding to a class. The existence of the asymptotic probability of reaching T from A will therefore depend on the percent of mass that is distributed on each class. If the more important mass is associated with the final classes, this means that the mass in each of the remaining classes will decrease with time and we will only have to take into account this family of final classes. In this case, a limit will exist. A class is called basic if the sub-matrix of A associated with it, admits as spectral radius the spectral radius of A.

Theorem 4. *Suppose the matrix A is reducible with spectral radius $\lambda_0 > 0$, and suppose the final classes are the only basic classes. Then there exists a unique asymptotic probability (depending on μ_0) of exit from A to T if and only if the asymptotic probability of exit from each final class also exists. Moreover, if the final classes have dominant spectral radius and if some of them are the only basic classes, then there exists a unique asymptotic probability of exit from A to T if μ_0 charges these classes.*

Finally, a particular case,

Theorem 5. *Suppose A admits a Jordan decomposition of the form $D+N$, where D is a diagonal block primitive sub-matrices with the same spectral radius λ_0 and N is an upper triangular non-negative matrix. Then the limiting conditional distribution exists.*

4 Extension to Multiple Targets and Examples of Markov Networks

The previous results may be extended to multiple targets problems. More precisely, let $\mathbf{T} = \{T_1, T_2, \ldots\}$ be target disjoint sets on the same R-network (E,P) over $\mathcal{M}_E(R)$. We are interested in the optimal $\{T_1, T_2, \ldots\}$-compatible relationship S such that (4) holds.

The answer is trivial, since each target class T_i defines its equivalence relationship S_i. It is not difficult to show that the required set S is just $S = \cap S_i$, see [2] and [5]. We start here by showing some classical Network problems that cannot be projected in smaller ones.

Example 1 (Negative Binomial Distribution). Repeat independently a game with probability p of winning until you win n games.

Let $S\tau = \sum_{i=1} \tau Y_i$, where $\{Y_i, i \in \mathbb{N}\}$ is a sequence of i.i.d. bernoulli random variable with $Prob(\{Y_i = 1\}) = 1 - Prob(\{Y_i = 0\}) = p$. Our interest is engaged by the computation of the probability of reaching n starting from 0. Let $E = \{0, 1, \ldots, n\}$ be the set of levels we have reached. We have

$$
\left.\begin{array}{c|cccccc}
 & 0 & 1 & 2 & \ldots & n-1 & n=T \\
\hline
0 & (1-p) & p & 0 & \ldots & 0 & 0 \\
1 & 0 & (1-p) & p & \ddots & 0 & 0 \\
2 & 0 & 0 & (1-p) & \ddots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
n-1 & 0 & 0 & 0 & \ldots & (1-p) & p \\
n=T & 0 & 0 & 0 & \ldots & 0 & 1
\end{array}\right\} =: P
$$

Since the length of the minimum path for reaching the target state n from different states is different, the problem cannot be projected on a smaller one by [4, Proposition 31].

Example 2 (Consecutive winning). Repeat independently a game with probability p of winning until you win n consecutive games. The problem is similar to the previous one, where

$$
\left.\begin{array}{c|cccccc}
 & 0 & 1 & 2 & \ldots & n-1 & n=T \\
\hline
0 & (1-p) & p & 0 & \ldots & 0 & 0 \\
1 & (1-p) & 0 & p & \ddots & 0 & 0 \\
2 & (1-p) & 0 & 0 & \ddots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
n-1 & (1-p) & 0 & 0 & \ldots & 0 & p \\
n=T & 0 & 0 & 0 & \ldots & 0 & 1
\end{array}\right\} =: P
$$

The problem is again not projectable by [4, Proposition 31].

Example 3 (Gambler's ruin). Let two players each have a finite number of pennies (say, n_1 for player one and n_2 for player two). Now, flip one of the pennies (from either player), with the first player having p probability of winning, and transfer a penny from the loser to the winner. Now repeat the process until one player has all the pennies.

Let $S\tau = \sum_{i=1} \tau(2Y_i - 1)$, where $\{Y_i, i \in \mathbb{N}\}$ is a sequence of i.i.d. bernoulli random variable with $Prob(\{Y_i = 1\}) = 1 - Prob(\{Y_i = 0\}) = p$. Our interest is engaged by the computation of the probability of reaching $T_1 = n_2$ or $T_2 = -n_1$ (multiple target) starting from 0. Let $E = \{-n_2, \dots, -1, 0, 1, \dots, n_1\}$ be the set of levels we have reached. We have

	$-n_1 = T_2$	$-n_1+1$	$\dots$	-1	0	1	$\dots$	n_2-1	$n_2 = T_1$
$-n_1 = T_2$	1	0	$\dots$	0	0	0	$\dots$	0	0
$-n_1+1$	$(1-p)$	0	$\ddots$	0	0	0	$\dots$	0	0
$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
-1	0	0	$\ddots$	0	p	0	$\dots$	0	0
0	0	0	$\vdots$	$(1-p)$	0	p	$\dots$	0	0
1	0	0	$\vdots$	0	$(1-p)$	0	$\ddots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$
n_2-1	0	0	$\vdots$	0	0	0	$\dots$	0	p
$n_2 = T_1$	0	0	$\dots$	0	0	0	$\dots$	0	1

This problem is clearly not projectable on a smaller one, since it is for T_1 (for example). The problem may be reduced if and only if we are interested in the time of stopping (without knowing who wins, i.e. $\mathbf{T} = T_1 \cup T_2$) and $p = 1/2$. In this case, the relevant information is the distance from the nearest border and hence the problem may be half-reduced.

The following classical problem may be reduced.

Example 4 (Random walk on a cube). A particle performs a symmetric random walk on the vertices of a unit cube, i.e., the eight possible positions of the particle are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(1,1,0), \dots, (1,1,1)$, and from its current position, the particle has a probability of $1/3$ of moving to each of the three neighboring vertices. This process ends when the particle reaches $(0,0,0)$ or $(1,1,1)$.

Let $T_1 = (0,0,0)$, $T_2 = (1,1,1)$. The following transition matrix

	$(0,0,0)$	$(1,0,0)$	$(0,1,0)$	$(0,0,1)$	$(1,1,0)$	$(1,0,1)$	$(0,1,1)$	$(1,1,1)$
$(0,0,0)$	1	0	0	0	0	0	0	0
$(1,0,0)$	1/3	0	0	0	1/3	1/3	0	0
$(0,1,0)$	1/3	0	0	0	1/3	0	1/3	0
$(0,0,1)$	1/3	0	0	0	0	1/3	1/3	0
$(1,1,0)$	0	1/3	1/3	0	0	0	0	1/3
$(1,0,1)$	0	1/3	0	1/3	0	0	0	1/3
$(0,1,1)$	0	0	1/3	1/3	0	0	0	1/3
$(1,1,1)$	0	0	0	0	0	0	0	1

can be easily reduced on

	t_1	f_1	f_2	t_2
t_1	1	0	0	0
f_1	1/3	0	2/3	0
f_2	0	2/3	0	1/3
t_2	0	0	0	1

where $t_i = T_i$ and $f_i = \{e = (e_1, e_2, e_3) : \sum e_j = i\}$. If we are only interested in the time of stopping (i.e. $\mathbf{T} = T_1 \cup T_2$), the previous problem may be reduced to a geometrical one. Clearly, this results hold also for random walk on a d-dimensional cube.

We give, in the following, several examples to the different results we obtained before.

Example 5 (Medical science). We intend to find the probability that the length of a woman's menstrual cycle can be the same three consecutive times. If the length of a menstrual cycle is uniformly distributed between 26 and 35 days (and the length of menstrual cycles being independent from one another), then the process may be seen as a Markov chain on $E = \{26, \dots, 35\}$, where

$$P = \begin{pmatrix} 1/10 & \dots & 1/10 \\ \vdots & \ddots & \vdots \\ 1/10 & \dots & 1/10 \end{pmatrix}.$$

The problem can be solved by introducing the stopping time defined by

$$S = \inf\{\tau \in \mathbb{N} \colon X_{\tau-2} = X_{\tau-1} = X\tau\}.$$

This problem can naturally be embedded in a 21 states Markov problem whose transition matrix is defined in (7), see [4].

	T	26.26	27.27	…	35.35	26	27	…	35
T	**1**	0	0	…	0	0	0	…	0
26.26	1/10	0	0	…	0	0	1/10	…	1/10
27.27	1/10	0	0	…	0	1/10	0	…	1/10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35.35	1/10	0	0	…	0	1/10	1/10	…	0
26	0	1/10	0	…	0	0	1/10	…	1/10
27	0	0	1/10	…	0	1/10	0	…	1/10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	0	0	0	…	1/10	1/10	1/10	…	0

(7)

It is simplified by considering the process $Z\tau$ with the following three states:

$$\widehat{E} = \begin{cases} & \text{there is no } N \le \tau \colon X_{N-2} = X_{N-1} = X_N : \\ 1, & \text{if } X_{\tau-1} \neq X\tau \\ 2, & \text{if } X_{\tau-2} \neq X_{\tau-1} = X\tau \\ 3 = T, & \text{if there exists } N \le \tau \colon X_{N-2} = X_{N-1} = X_N \end{cases}$$

with initial distribution $\mu_0^T = (1,0,0)$ and matrix

$$\left.\begin{array}{c|ccc} & 1 & 2 & 3=T \\ \hline 1 & p & (1-p) & 0 \\ 2 & p & 0 & (1-p) \\ 3=T & 0 & 0 & 1 \end{array}\right\} =: \widehat{P}\,,$$

where $p = 9/10$. Simple calculations (see [4]) give the corresponding Hazard rate:

$$H_S(\tau) = 1 - \underbrace{\frac{1}{20}\frac{(9+\sqrt{117})^{\tau+1} - (9-\sqrt{117})^{\tau+1}}{(9+\sqrt{117})^{\tau} - (9-\sqrt{117})^{\tau}}}_{\xrightarrow[\tau\to\infty]{}\hat{\lambda}_0} = \mathbb{P}(X_{\tau+1} \in T | X\tau \in A)\,.$$

The matrix A of the transient states is

$$A = \begin{pmatrix} p & 1-p \\ p & 0 \end{pmatrix}.$$

Therefore, the first eigenvalue is

$$\lambda_0 = \frac{p + \sqrt{-3p^2+4p}}{2} \underset{(p=9/10)}{=} \frac{9+\sqrt{117}}{20},$$

while the first left eigenvector f_0 associated to λ_0 is

$$(f_0)^T = \left(\tfrac{p+\sqrt{-3p^2+4p}}{2(1-p)},\, 1\right),$$

and the limit conditional distribution is

$$\frac{(f^0)^T}{(f^0)^T\underline{1}} = \left(\tfrac{p+\sqrt{-3p^2+4p}}{2-p+\sqrt{-3p^2+4p}},\, \tfrac{2(1-p)}{2-p+\sqrt{-3p^2+4p}}\right).$$

We have

$$\frac{(f^0)^T v}{(f^0)^T\underline{1}} = \frac{(f^0)^T}{(f^0)^T\underline{1}} v = \left(\tfrac{p+\sqrt{-3p^2+4p}}{2-p+\sqrt{-3p^2+4p}},\, \tfrac{2(1-p)}{2-p+\sqrt{-3p^2+4p}}\right)\begin{pmatrix} 0 \\ 1-p \end{pmatrix} = 1 - \lambda_0.$$

As expected, we note that $\lambda_0 = \hat{\lambda}_0$.

Example 6 (The gambler ruin). A gambler A plays against a gambler B a sequence of heads or tails independent games. The total sum of their wealth is a\$. At each game, A wins one dollar or loses it with probability p and $q = 1-p$ respectively. The game stops when one of the gamblers is ruined. Denote by $X\tau$ the wealth of A at the end of the τ-th game. $X\tau$ is a Markov chain with set of states $E = \{0,\ldots,a\}$. Its transition matrix is given by

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}.$$

If we take the position of A, then $T = \{a\}$ and $T_\infty = \{0\}$.

Then the previous matrix can be rearranged:

$$P = \begin{pmatrix} 1 & 0 & & & \mathbf{0} & & \\ 0 & 1 & & & & & \\ q & 0 & 0 & p & \cdots & 0 & 0 \\ 0 & 0 & q & 0 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & p & 0 & \cdots & \cdots & q & 0 \end{pmatrix}.$$

In this case,

$$A = \begin{pmatrix} 0 & p & \cdots & 0 & 0 \\ q & 0 & p & \cdots & 0 \\ \vdots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & q & 0 & p \\ 0 & \cdots & \cdots & q & 0 \end{pmatrix}, \qquad v = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ p \end{pmatrix}, \qquad v_\infty = \begin{pmatrix} q \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

A is an irreducible matrix with period 2. Suppose first that $a = 2k+1$ ($2k$ transient states $A = \{1,\dots,2k\}$ plus $T = \{a\}$ and $T_\infty = \{0\}$). Then, if we permute the order of the states,

$$\hat{A} = \begin{pmatrix} \mathbf{0} & \begin{matrix} p & 0 & \cdots & \cdots & 0 \\ q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & \cdots & q & p \end{matrix} \\ \begin{matrix} q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & \cdots & q & p \\ 0 & \cdots & \cdots & 0 & q \end{matrix} & \mathbf{0} \end{pmatrix}, \qquad \hat{v} = v, \qquad \hat{v}_\infty = v_\infty.$$

Suppose now that $a = 2k$ ($2k-1$ transient states $A = \{1,\dots,2k-1\}$ plus $T = \{a\}$ and $T_\infty = \{0\}$). Then

$$\hat{A} = \begin{pmatrix} \mathbf{0} & \begin{matrix} p & 0 & \cdots & \cdots & 0 \\ q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & \cdots & q & p \\ 0 & \cdots & \cdots & 0 & q \end{matrix} \\ \begin{matrix} q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & \cdots & q & p \end{matrix} & \mathbf{0} \end{pmatrix}, \qquad \hat{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ p \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \hat{v}_\infty = v_\infty.$$

In both cases, $a = 2k+1$ and $a = 2k$, $\hat{A}^2$ takes the form

$$\hat{A}^2 = \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}$$

with $(C_i)^m \gg 0$ for all $m \geq k-1$. Moreover,

$$v_{2-\mathrm{mod}(a,2)} = (0,0,\ldots,0)^T, \qquad v_{1+\mathrm{mod}(a,2)} = (0,0,\ldots,p)^T$$

In this case it is obvious that the conditional limit does not exist (see [5]); however, we can compute the asymptotic conditional law given each class of periodicity. For example, if $a = 10$ and $p = 1/2$, we obtain

$$C_1 = \begin{pmatrix} 1/4 & 1/4 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/4 & 1/4 \end{pmatrix} \quad \text{and} \qquad C_2 = \begin{pmatrix} 1/2 & 1/4 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/4 & 1/2 \end{pmatrix}.$$

We have

$$\lambda_0^{(1)} = \lambda_0^{(2)} = \frac{5+\sqrt{5}}{8},$$

while the corresponding first eigenvectors are

$$\frac{(f^0)^{T^{(1)}}}{(f^0)^T \underline{1}} = \left(\tfrac{1}{2(3+\sqrt{5})},\ 1/4,\ \tfrac{\sqrt{5}+1}{2(3+\sqrt{5})},\ 1/4,\ \tfrac{1}{2(3+\sqrt{5})}\right)$$
$$\frac{(f^0)^{T^{(2)}}}{(f^0)^T \underline{1}} = \left(\tfrac{1}{3+\sqrt{5}},\ \tfrac{\sqrt{5}+1}{2(3+\sqrt{5})},\ \tfrac{\sqrt{5}+1}{2(3+\sqrt{5})},\ \tfrac{1}{3+\sqrt{5}}\right).$$

Example 7 (Random walk on a polygon). A particle can move on a regular polygon with r sides. Its vertices are numbered from 0 to $r-1$. If at some time the particle is on the vertex i ($0 \leq i \leq r-1$) then, right afterwards, it will be in state $i+1$ (mod r) with probability p and in the state $i-1$ (mod r) with probability q. We assume that $r = 2N$. We also suppose that the particle can exit the polygon to a target set from each vertex 0 to $N-1$ and to T_∞ from each other vertex, both with probability s. We denote by $X\tau$ the vertex visited by the particle at time τ, its transition matrix is given hereafter

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ s & 0 & 0 & p & 0 & \cdots & 0 & 0 & q \\ s & 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ s & 0 & 0 & q & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & s & 0 & 0 & 0 & \cdots & 0 & p & 0 \\ 0 & s & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & s & p & 0 & 0 & \cdots & 0 & q & 0 \end{pmatrix}.$$

Consider now the sub-matrix A of transient states leading to T. This matrix is irreducible, periodic with period 2 and $A/(p+q)$ bi-stochastic. The largest eigenvalue is $\lambda_0 = p+q$, see [5]. The left eigenvector associated with λ_0 is $v^0 = \underline{1}$. It follows that the limiting conditional probability law is equal to $\frac{(f^0)^T}{(f^0)^T\underline{1}} = (1/r,\dots,1/r)$, which is the result one would expect for a matrix that is bi-stochastic (up to a multiplicative constant), and irreducible. It, indeed, admits a unique stationary probability given by $u = (u_1,\dots,u_r)$ such that $u_i = 1/r$, for each i.

Then, if we permute the order of the states,

$$\hat{A} = \begin{pmatrix} \mathbf{0} & \begin{matrix} p & 0 & \cdots & \cdots & q \\ q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & \cdots & q & p \end{matrix} \\ \begin{matrix} q & p & \cdots & \cdots & 0 \\ 0 & q & p & \cdots & 0 \\ & \ddots & \ddots & & \\ 0 & \cdots & \cdots & q & p \\ p & \cdots & \cdots & 0 & q \end{matrix} & \mathbf{0} \end{pmatrix}, \qquad \hat{v} = v, \qquad \hat{v}_\infty = v_\infty.$$

Notice that the first sub-matrix A_1 is composed by the even states (we start from 0) and the second sub-matrix by the odd ones.

It follows that $\hat{v} = (s,s,\dots,0,0,s,s,\dots,0)^T \qquad \hat{v}_\infty = v_\infty$. In fact, if N is itself an even number, $\hat{v}_1 = \hat{v}_2 = \begin{pmatrix} s \\ s \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ and the conditional limit of exit is $\frac{f_0^{(1)}\hat{v}_1}{f_0^{(1)}\underline{1}} = \frac{f_0^{(2)}\hat{v}_2}{f_0^{(2)}\underline{1}} = \frac{sN/2}{N} = \frac{s}{2}$. The result will be the same if we consider a random walk on the line with reflecting barriers together with a jump to the other side.

Example 8 (The bonus and malus model). An insurance company orders the bonus–malus levels of its clients according to integers $0,1,2,\dots$. The level 0 is the most advantageous for the client. Let $0 \le i \le j$. If the bonus-malus level of an insured is i at time τ, it will be j at time $\tau+1$ if, between times τ and $\tau+1$, he had $j-i$ accidents. We denote by $(X\tau)$ the sequence of the bonus-malus levels for this insured. Time unit is a year and we suppose that the number of accidents during a year is a Poisson random variable with parameter λ. The probability that the insured moves from level i to level j, $j \ge i$ is equal to $\pi_{\lambda,j-i} = e^{-\lambda}\lambda^{j-i}/(j-i)!$. Furthermore, we suppose that the contract is canceled once the insured has had N accidents. $(X\tau)$ is a Markov chain with set of states $\{0,\dots,N\}$ and transition matrix

$$P = \begin{pmatrix} \pi_{\lambda,0} & \pi_{\lambda,1} & \pi_{\lambda,3} & \cdots & \pi_{\lambda,N} \\ 0 & \pi_{\lambda,0} & \pi_{\lambda,1} & \cdots & \pi_{\lambda,N-1} \\ 0 & 0 & \pi_{\lambda,0} & \cdots & \pi_{\lambda,N-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

N can be considered to be an absorbing state and to represent the target set T.

This is the Jordan case with $D+N$, where D is a diagonal matrix, and N is an upper triangular matrix. According to Theorem 5, the conditional limit probability of exit from the last class exists.

References

1. Aki S, Hirano K (1999) Sooner and later waiting time problems for runs in Markov dependent bivariate trials. Ann Inst Statist Math 51(1):17–29
2. Aletti G (2006) Compressing redundant information in Markov chains. http://arxiv.org/math.PR/0605099
3. Aletti G (2008) Compressing redundant information in networks (submitted)
4. Aletti G, Merzbach E (2006) Stopping markov processes and first path on graphs. J Eur Math Soc (JEMS) 8(1):49–75
5. Aletti G, Saada D (2008) Asymptotic reduction of Markov chains (submitted)
6. Antzoulakos DL, Philippou AN (1997) Probability distribution functions of succession quotas in the case of Markov dependent trials. Ann Inst Statist Math 49(3):531–539
7. Arnold BC, Balakrishnan N, Nagaraja HN (1998) Records. Wiley series in probability and statistics: probability and statistics. Wiley, New York
8. Cairoli R, Dalang RC (1996) Sequential stochastic optimization. Wiley series in probability and statistics: probability and statistics. Wiley, New York
9. Ebneshahrashoob M, Sobel M (1990) Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas. Statist Probab Lett 9(1):5–11
10. Koutras MV, Alexandrou VA (1997) Sooner waiting time problems in a sequence of trinary trials. J Appl Probab 34(3):593–609
11. Lam K, Yam HC (1997) Cusum techniques for technical trading in financial markets. Financial Engineering & the Japanese Markets 4(3):257–274
12. Stefanov VT (2000) On some waiting time problems. J Appl Probab 37(3):756–764
13. Stefanov VT, Pakes AG (1997) Explicit distributional results in pattern formation. Ann Appl Probab 7(3):666–678