A REPORT ON MARKOV CHAINS

Siddharth Shah

May 28, 2020

1 Introduction

Modern probability theory studies chance processes for which, in a sequence of experiments the knowledge of previous outcomes influences predictions for future experiments. In principle, when we observe such a sequence of chance experiments, all of the past outcomes could influence our predictions for the next experiment. For example, this should be the case in predicting a students grades on a sequence of exams in a course, or predicting the result of the next match in a cricket series between India and Australia, etc. But to allow this much generality would make it very difficult to prove general results.

This report will begin with a brief introduction, followed by the analysis. The analysis will introduce the concepts of Markov chains, explain different types of Markov Chains and present examples of its applications.

1.1 Background

Andrei Markov was a Russian mathematician who lived between1856 and 1922. He was a poorly performing student and the only subject he didnt have difficulties in was mathematics. He later studied mathematics at the university of Petersburg and was lectured by Pafnuty Chebyshev, known for his work in probability theory. Markovs first scientific areas were in number theory, convergent series and approximation theory. In 1907, A. A. Markov began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment. This type of process is called a Markov chain. He was also very interested in poetry and the first application he found of Markovchains was in fact in a linguistic analysis of Pusjkins work EugeneOnegin.

2 What is a Markov chain?

A Markov chain is a mathematical system that consists of a set of *states*, $S = \{s_1, s_2, ..., s_r\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a *step*. The system experiences these transitions according to certain probabilistic rules. The defining characteristic of a Markov chain is that no matter how the process arrived at its present state, the possible future states (and the probabilities of transitioning into them in one step) are fixed. In other words, the probability of transitioning to any particular state is dependent solely on the current state and time elapsed.

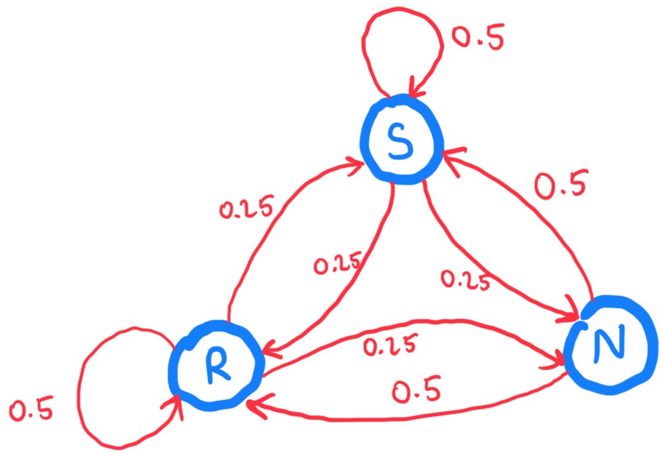
Consider a Markov chain with the set of states, $S = \{s_1, s_2, ..., s_r\}$. If the chain is currently in state s_i , then it moves to state s_j at the next step with a probability denoted by p_{ij} , and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state. That is, the starting state can also be *not fixed*, and rather depend on a probability distribution over the set of states, S. But usually a particular state is chosen as the starting state.

R. A. Howard, an American botanist, provides us with a picturesque description of a Markov chain as a frog jumping on a set of lily pads. The frog starts on one of the pads and then jumps from lily pad to lily pad with the appropriate transition probabilities.

2.1 Modelling a Markov chain

As one can expect, Markov chains may be modeled by finite state machines.

Let us look at an **example (2.1)** here.
According to Kemeny, Snell, and Thompson, the Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. With this information we form a Markov chain as follows. We take as states the kinds of weather R, N, and S. From the above information we determine the transition probabilities and draw the FSM.



3 Transition Matrix

In the above example of the land of Oz, we could have represented the transition probabilities in the form of a matrix, instead of the transition arrows of the FSM. This kind of representation is much more compact and readable. Such a matrix **P**, in which the ij^{th} entry p_{ij} is the probability of transitioning from the i^{th} state to the j^{th} state in one step is called the *matrix of transition probabilities,* or the *transition matrix.*

$$P_{3,3} = \begin{array}{cccc} & R & N & S \\ R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}$$

Here, the entries in the first, second and third rows represent the probabilities for the various kinds of weather following rainy, nice and snowy days, respectively.

Let us look at another **example (3.1)** now with a transition matrix.
The President of the United States tells person A his or her intention to run or not to run in the next election. Then A relays the news to B, who in turn relays the message to C, and so forth, always to some new person. We assume that there is a probability a that a person will change the answer from yes to no when transmitting it to the next person and a probability b that he or she will change it from no to yes. We choose as states the message, either yes or no. The transition matrix is then

$$
\begin{array}{ccc}
 & yes & no \\
P_{2,2} = yes & 1-a & a \\
no & b & 1-b
\end{array}
$$

The initial state represents the Presidents choice.

Next, we consider the question of determining the probability that, given the chain is in state i at some instant, it will be in state j after 2 steps. We denote this probability by ${p^{(2)}}_{ij}$. For exploring this question, let's go back to the example of the land of Oz. In this example, we see that if it is snowy today then the event that it is rainy two days from now is the disjoint union of the following three events: 1) it is rainy tomorrow and rainy two days from now, 2) it is nice tomorrow and rainy two days from now, and 3) it is snowy tomorrow and rainy two days from now. The probability of the first of these events is the product of the conditional probability that it is rainy tomorrow, given that it is snowy today, and the conditional probability that it is rainy two days from now, given that it is rainy tomorrow. Using the transition matrix P, we can write this product as $p_{31}p_{11}$. The other two events also have probabilities that can be written as products of entries of P. Thus we have,

$$ {p^{(2)}}_{ij} = p_{31}p_{11} + p_{32}p_{21} + p_{33}p_{31} $$

This equation must look quite familiar to the reader. It's the dot product of the third row of **P** with the first column of **P**. This is just what is done in obtaining the 3,1-entry of the product of **P** with itself. In general, if a Markov chain has r states, then

$$ {p^{(2)}}_{ij} = \sum_{k=1}^{r} p_{ik}p_{kj} $$

This brings us to a more general theorem, the hint for which can be observed from the above equation.

THEOREM 1. *Let* $\boldsymbol{P}$ *be the transition matrix of a Markov chain. The* ij^{th} *entry* p_{ij} *of the matrix* $\boldsymbol{P}^n$ *gives the probability that the Markov chain, starting in state* s_i *, will be in state* s_j *after* n *steps.*

We shall not be proving any theorem or corollary that we come across in this report, as the main aim is to illustrate application of the concept of Markov chains to some real world problems.

We now consider the long-term behavior of a Markov chain when it starts in a state chosen by a probability distribution on the set of states, which we will call a *probability vector* . A *probability vector* with r components is a row vector whose entries are non-negative and sum to 1. If **u** is a probability vector which represents the initial state of a Markov chain, then we think of the ith component of **u** as representing the probability that the chain starts in state s_i . With this interpretation of random starting states, it is easy to prove the following theorem.

THEOREM 2. *Let $\boldsymbol{P}$ be the transition matrix of a Markov chain, and let* $\mathbf{u}$ *be the probability vector which represents the starting distribution. Then the probability that the chain is in state s_i after n steps is the i^{th} entry in the vector*

$$\boldsymbol{u}^{(n)} = \boldsymbol{u}\boldsymbol{P}^n.$$

We note that if we want to examine the behavior of the chain under the assumption that it starts in a certain state s_i , we simply choose $\mathbf{u}$ to be the probability vector with ith entry equal to 1 and all other entries equal to 0.

4 Types of Markov Chains

4.1 Ergodic Markov Chains

In this section, we shall briefly study ergodic Markov chains.

Definition 1: A Markov chain is called an ergodic chain if it is possible to go from every state to every state (not necessarily in one move). Such Markov chains are also called irreducible Markov chains.

Definition 2: The Markov chains, which are not irreducible (i.e., in which it isn't possible to go from each state to each other state in one or more steps) are called reducible Markov chains.

Definition 3: A Markov chain is called a regular chain if some power of the transition matrix has only positive elements.

In other words, for some n, it is possible to go from any state to any state in exactly n steps. It is clear from this definition that every regular chain is ergodic. On the other hand, an ergodic chain is not necessarily regular, as the following example shows.

Example 4.1:

Let the transition matrix of a Markov chain be defined by

$$\begin{array}{rcc} & State0 & State1 \\ P = State0 & 0 & 1 \\ State1 & 1 & 0 \end{array}$$

Then is clear that it is possible to move from any state to any state, so the chain is ergodic. However, if n is odd, then it is not possible to move from state 0 to state 0 in n steps, and if n is even, then it is not possible to move from state 0 to state 1 in n steps, so the chain is not regular.

4.1.1 Regular Markov chains

Any transition matrix that has no zeros determines a regular Markov chain. However, it is possible for a regular Markov chain to have a transition matrix that has zeros. The transition matrix of the Land of Oz example has $p_{NN} = 0$

but the second power P^2 has no zeros, so this is a regular Markov chain. An example of a nonregular Markov chain is an absorbing chain, which we will see in a short while.

THEOREM 3. *Let P be the transition matrix for a regular chain. Then, as $n \to \infty$, the powers P^n approach a limiting matrix $\boldsymbol{W}$ with all rows the same vector $\boldsymbol{w}$. The vector $\boldsymbol{w}$ is a strictly positive probability vector (i.e., the components are all positive and they sum to one).*

As is the theme of this report, we shall treat this theorem as a given, and move on to the next theorem, on our path to the interesting applications of Markov chains at the end of the report.

THEOREM 4. *Let P be a regular transition matrix, let*

$$W = \lim_{n\to\infty} P^n,$$

let w be the common row of W, and let c be the column vector all of whose components are 1. Then

1. *$wP = w$, and any row vector v such that $vP = v$ is a constant multiple of w.*

2. *$Pc = c$, and any column vector x such that $Px = x$ is a multiple of c.*

Proof. To prove part 1, we note that from Theorem 3,

$$P^n \to W.$$

Thus,

$$P^{n+1} = P^n \cdot P \to WP$$

But $P^{n+1} \to W$, and so $W = WP$ and hence, $w = wp$.

Let v be any vector with $vP = v$. Then $v = vP^n$, and passing to the limit, $v = vW$. Let r be the sum of the components of v. Then it is easily checked that $vW = rw$. So, $v = rw$.

To prove part 2, assume that $x = Px$. Then $x = P^n x$, and again passing to the limit, $x = Wx$. Since all rows of W are the same, the components of Wx are all equal, so x is a multiple of c.

Note that an immediate consequence of Theorem 4 is the fact that there is only one probability vector v such that vP = v.

□

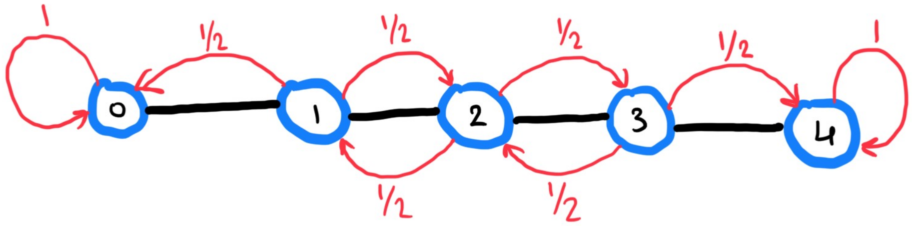
4.2 Absorbing Markov Chains

Definition 1: A state s_i of a Markov chain is called absorbing if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).

Definition 2: In an absorbing Markov chain, a state which is not absorbing is called *transient*.

As the definitions are quite clear, we shall quickly see an example related to absorbing Morkov chains. And then we shall move on to the next key concept.
Example 4.1

A man walks along a four-block stretch of Park Avenue, as shown in the figure. If he is at corner 1, 2, or 3, then he walks to the left or right with equal probability. He continues until he reaches corner 4, which is a bar, or corner 0, which is his home. If he reaches either home or the bar, he stays there. We form a Markov chain with states 0, 1, 2, 3, and 4. States 0 and 4 are absorbing states. The transition matrix is then



A Drunkard's walk

$$P = \begin{array}{c|ccccc} & State0 & State1 & State2 & State3 & State4 \\ \hline State0 & 1 & 0 & 0 & 0 & 0 \\ State1 & 1/2 & 0 & 1/2 & 0 & 0 \\ State2 & 0 & 1/2 & 0 & 1/2 & 0 \\ State3 & 0 & 0 & 1/2 & 0 & 1/2 \\ State4 & 0 & 0 & 0 & 0 & 1 \end{array}$$

The states 1, 2, and 3 are transient states, and from any of these it is possible to reach the absorbing states 0 and 4. Hence the chain is an absorbing chain. When a process reaches an absorbing state, we say that it is absorbed.

4.2.1 Canonical Form

Consider an arbitrary absorbing Markov chain. Renumber the states so that the transient states come first. If there are r absorbing states and t transient states, the transition matrix will have the following *canonical form*

$$\begin{array}{ccc} & TR. & ABS. \\ P = TR. & \mathbf{Q} & \mathbf{R} \\ ABS. & \mathbf{0} & \mathbf{I} \end{array}$$

Here $\mathbf{I}$ is an r-*by*-r indentity matrix, $\mathbf{0}$ is an r-*by*-t zero matrix, $\mathbf{R}$ is a nonzero t-*by*-r matrix, and $\mathbf{Q}$ is an t-*by*-t matrix. The first t states are transient and the last r states are absorbing.
A standard matrix algebra argument shows that $\mathbf{P}^n$ is of the form

$$\begin{array}{ccc} & TR. & ABS. \\ \mathbf{P}^n = TR. & \mathbf{Q}^n & (I + Q + Q^2 + \cdots + Q^{n-1})R \\ ABS. & \mathbf{0} & \mathbf{I} \end{array}$$

The form of $\mathbf{P}^n$ shows that the entries of $\mathbf{Q}^n$ give the probabilities for being in each of the transient states after n steps for each possible transient starting state.

4.2.2 Probability of Absorption

THEOREM 5. *In an absorbing Markov chain, the probability that the process will be absorbed is 1 (i.e.,* $\boldsymbol{Q}^n \to \boldsymbol{0}$ *as* $n \to \infty$*).*

As is the theme of this report, we shall treat this theorem too as a given, and move on to the next key concept, on our path to the interesting applications of Markov chains at the end of the report.

4.2.3 The Fundamental Matrix

THEOREM 6. *For an absorbing Markov chain the matrix* $\boldsymbol{I} - \boldsymbol{Q}$ *has an inverse* $\boldsymbol{N}$ *and* $\boldsymbol{N} = \boldsymbol{I} + \boldsymbol{Q} + \boldsymbol{Q}^2 + \cdots$*. The ij-entry* n_{ij} *of the matrix* $\boldsymbol{N}$ *is the expected number of times the chain is in state* s_j*, given that it starts in state* s_i *. The initial state is counted if* $i = j$*.*

Proof. Let $(I - Q)x = 0$, that is $x = Qx$. Then, iterating this we see that $x = Q^n x$. Since $Q^n \to 0$, we have $Q^n x \to 0$, so $x = 0$. Thus $(I - Q)^{-1} = N$ exists. The next thing to note is that

$$(1 - Q)(I + Q + Q^2 + \cdots + Q^n) = I - Q^{n+1}$$

Thus, multipying both sides by N gives

$$I + Q + Q^2 + \cdots + Q^n = N(I - Q^{n+1}).$$

Letting n tend to infinity, we have

$$N = I + Q + Q^2 + \cdots$$

Let s^i and s^j be two transient states, and assume throughout the remainder of the proof that i and j are fixed. Let $X^{(k)}$ be a random variable which equals 1 if the chain is in state s^j after k steps, and equals 0 otherwise. For each k, this random variable depends upon both i and j. We have

$$P(X^{(k)} = 1) = q_{ij}^{(k)},$$

and

$$P(X^{(k)} = 0) = 1 - q_{ij}^{(k)},$$

where $q_{ij}^{(k)}$ is the ij^{th} entry of Q^k. These equations hold for $k = 0$ since $Q^0 = I$. Therefore, since $X^{(k)}$ is a 0-1 random variable, $E(X^{(k)}) = q_{ij}^{(k)}$.

The expected number of times the chain is in state s_j in the first n steps, given that it starts in state s_i , is clearly

$$E(X^{(0)} + X^{(1)} + \cdots + X^{(n)}) = q_{ij}^{(0)} + q_{ij}^{(1)} + \cdots + q_{ij}^{(n)}.$$

Letting n tend to infinity we have

$$E(X^{(0)} + X^{(1)} + \cdots) = q_{ij}^{(0)} + q_{ij}^{(1)} + \cdots = n_{ij}.$$

□

Definition 3: For an absorbing Markov chain $\mathbf{P}$, the matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ is called the *fundamental matrix* for $\mathbf{P}$. The entry n_{ij} of $\mathbf{N}$ gives the expected number of times that the process is in the transient state s_j if it is started in the transient state s_i .

4.2.4 Time to Absorption

We now consider the question: Given that the chain starts in state s_i , what is the expected number of steps before the chain is absorbed? The answer is given in the next theorem.

THEOREM 7. *Let t_i be the expected number of steps before the chain is absorbed, given that the chain starts in state s_i , and let $\boldsymbol{t}$ be the column vector whose i^{th} entry is t_i. Then*

$$\boldsymbol{t} = \boldsymbol{N}\boldsymbol{c},$$

where $\boldsymbol{c}$ is a column vector all of whose entries are 1.

Proof. If we add all the entries in the i^{th} row of N, we will have the expected number of times in any of the transient states for a given starting state s_i , that is, the expected time required before being absorbed. Thus, t_i is the sum of the entries in the i^{th} row of N. If we write this statement in matrix form, we obtain the theorem. □

4.2.5 Absorption Probabilities

THEOREM 8. *Let b_{ij} be the probability that an absorbing chain will be absorbed in the absorbing state s_j if it starts in the transient state s_i . Let B be the matrix with entries b_{ij} . Then B is an t-by-r matrix, and*

$$\boldsymbol{B} = \boldsymbol{NR}$$

where N is the fundamental matrix and R is as in the canonical form.

Proof. We have

$$B_{ij} = \sum_n \sum_k q_{ik}^{(n)} r_{kj}$$

$$B_{ij} = \sum_k \sum_k q_{ik}^{(n)} r_{kj}$$

$$B_{ij} = \sum_k n_{ik} r_{kj}$$

$$B_{ij} = (NR)_{ij}$$

□

This completes the proof.

Having gained all of this knowledge, it's time to have a look at some of the applications of Markov chains.

5 Applications of Markov chains

In this section, we'll see how Markov chains can be of great importance when it comes to real life experiments and applications. For that, we will be looking into 2 social/psycological experiments, which make use of absorbing Markov chains. Now, without further delay, we shall jump into the experiments.

5.1 Experiment 1: Conformity to group pressure

A famous psychology experiment examins the degree to which people conform to group pressure. In the experiment, some very simple tasks are given to the participants. A subject is led into a room and seated with some other participants. The experimenter presents them with a simple visual task - to determine which of three lines is the same length as a reference line. In sequence, each participant is asked to give their response, with the subject answering last. The task is designed so that the correct response is obvious. However, without knowledge to the subject, all of the other participants are confederates of the experimenter, and have been instructed to give the same incorrect answer. Thus, when it is finally the subjects turn to answer, the subject has to choose whether to give the correct answer (ignoring pressure to conform to the group) or the incorrect answer (giving into this pressure).

After recording the responses, the experimenter presents another similar task, and participants again have to answer sequentially with the subject going last. Let's say, 30 such tasks are given to the participants. Thus, for each subject, the data might consist of a sequence of responses such as

$$aabbbaabbabbbaaabbabbbbbbbbbbbbbbbbbbb$$

where a denotes the correct answer (against the group) and b denotes the incorrect answer (with the group). If this experiment is repeated with other subjects, one finds that the sequences generated by them illustrate a common pattern. The pattern being, most subjects initially waver between a's and b's, but eventually keep giving a single type of response (with some subjects choosing a and others choosing b). After a little thought, one can come to a conclusion that this stochastic process can be modelled as an Absorbing Markov chain (because the process seems to have been absorbed in certain states). Thus, we propose a Markov chain model with states $S = \{s_1, s_2, s_3, s_4\}$. The states are as follows:

1. temporarily against the group (responding a)
2. temporarily with the group (responding b)
3. permanently against the group (all further responses are a's)
4. permanently with the group (all further responses are b's)

In our model, states s_3 and s_4 are absorbing, while transitions can occur from state s_1 to s_3 or s_2, but not s_4, and from state s_2 to s_1 or s_4, but not s_3. More precisely, using the experimental data, we have the following transition matrix

$$P = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ \hline s_1 & 0.63 & 0.31 & 0.06 & 0 \\ s_2 & 0.46 & 0.49 & 0 & 0.05 \\ s_3 & 0 & 0 & 1 & 0 \\ s_4 & 0 & 0 & 0 & 1 \end{array}$$

This transition matrix P is certainly that of an absorbing Markov chain as from each of the non-absorbing states (s_1 and s_2) it is possible to transition into an absorbing state (s_3 or s_4).

Now, if we look at some of the powers P^n of P, we find the following results.

$$P^2 = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ \hline s_1 & 0.5395 & 0.3472 & 0.0978 & 0.0155 \\ s_2 & 0.5152 & 0.3827 & 0.0276 & 0.0745 \\ s_3 & 0 & 0 & 1 & 0 \\ s_4 & 0 & 0 & 0 & 1 \end{array}$$

$$P^{10} = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ s_1 & 0.3324 & 0.2269 & 0.3073 & 0.1334 \\ s_2 & 0.3368 & 0.2299 & 0.2375 & 0.1958 \\ s_3 & 0 & 0 & 1 & 0 \\ s_4 & 0 & 0 & 0 & 1 \end{array}$$

$$P^{1000} = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ s_1 & 0 & 0 & 0.6638 & 0.3362 \\ s_2 & 0 & 0 & 0.5987 & 0.4013 \\ s_3 & 0 & 0 & 1 & 0 \\ s_4 & 0 & 0 & 0 & 1 \end{array}$$

If we assume that the subjects start in one of the 2 temporary/non-absorbing states, then by looking at these matrices, we can infer that in the "short run" - after only 2 or 3 responses most subjects would still occupy one of the non-absorbing states. But in the "long run", after many responses, each subject will eventually become either a permanent non-conformist (s_3) or a permanent conformist (s_4). In particular, if a subject were initially a temporary non-conformist (s_1), the he/she has a 66.38% chance of ending up as a permanent non-conformist and a 33.62% chance of ending up a permanent conformist. If the subject were initially a temporary conformist (s_2), he/she has a 59.87% chance of ending up a permanent non-conformist and a 40.13% chance of ending up as a permanent conformist.

5.1.1 Analysis of expected time until absorption

As mentioned before, we can write P, P^t and P^∞ as

$$P = \begin{bmatrix} Q & R \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

$$P^t = \begin{bmatrix} Q^t & (I + Q + Q^2 + \cdots + Q^{t-1})R \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

$$P^\infty = \begin{bmatrix} \mathbf{0} & NR \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

where $N = I + Q + Q2 + \cdots = (I - Q)^{-1}$ is the fundamental matrix for the absorbing chain.

From here, we get R, Q and N as follows:

$$R = \begin{bmatrix} 0.06 & 0 \\ 0 & 0.05 \end{bmatrix}$$

$$Q = \begin{bmatrix} 0.63 & 0.31 \\ 0.46 & 0.49 \end{bmatrix}$$

$$N = \begin{bmatrix} 11.0629 & 6.7245 \\ 9.9783 & 8.026 \end{bmatrix}$$

Hence, from N, we can find $t = Nc$, as we have discussed before.

$$t = \begin{bmatrix} 17.7874 \\ 18.0043 \end{bmatrix}$$

Therefore, if the chain began in the 1st non-absorbing state (i.e., s_1), we can expect the chain to spend 11.06 steps in that state, 6.72 steps in the 2nd non-absorbing state (i.e., s_2), and hence 17.78 periods overall before absorption (into either state 1 or 2). Alternatively, if the chain began in the 2nd non-absorbing state (i.e., s_2), we can expect it to spend 9.98 steps in the 1st non-absorbing state, 8.03 steps in the 2nd non-absorbing state, and hence 18.00 steps before absorption. Of course, in the context of the example, each step corresponds to one response made by the subject.

5.2 Formation of dominance hierarchies

This is a study in which we are going to explore the structure of some social networks. In some types of networks, social ties between actors are not always reciprocated. For example, within an organization, we might observe worker i giving advice to worker j but not receiving advice in return. Similarly, in a large group of friends, i might like j, even though j might not like i very much. Such networks are sometimes called dominance hierarchies, and we will try to infer the "pecking order" within a group from the structure of the network. Google defines pecking order as "a hierarchy of status seen among members of a group of people or animals, originally as observed among hens". In the context of set theory, pecking order is similar to total order, which is a reflexive, transitive, anti-symmetric relation among the elements of a set, which necessarily applies to each pair of elements.

But it's not that a pecking order will always exist for a group of actors. For instance, in a cricket league, let iBj indicate that team i beat team j this year. If, at the end of the league, final rankings of all the teams are to be declared, it would not create controversy if (i) every pair of teams played each other once and (ii) there were never any upsets. More formally, the who-beats-whom relation B on the set of teams S would need to be a total order, which requires iBj or jBi, but not both (except iBi - this is included for the sake of completeness) for all i, $j \in S$ and (iBj and jBk) implies iBk for all $i, j, k \in S$.

Of course, in reality, in leagues many teams do not play each other, and we sometimes see upsets corresponding to cycles where iBj and jBk but kBi.

Now, we will go to our experiment. In this experiment, previously unacquainted chickens are placed together in a pen. After being placed, the chickens will eventually establish a dominance hierarchy that is complete. Further, while these hierarchies sometimes contain cycles, transitivity is more common than the case that dominance is random among the chickens (i.e., if dominance between each pair of chickens were determined by flipping a coin). Based on this observation, we can propose that this result is due to the "bystander effect". Dominance can be established directly when chicken i attacks some other chicken k. However, if this attack is witnessed by bystander j, the attack may also induce dominance of i over j or dominance of j over k (or both). Thus, a single attack may generate a transitive subnetwork.

Next, we propose a model for the formation of this network. The process begins at time $t = 0$ with no pre-existing dominance relation. At each step, the following series of events occurs:

1. Some actor i attacks some actor k, and this attack is witnessed by some bystander j. The triplet (i, j, k) is drawn randomly from the set of all distinct triplets such that k does not already dominate i.

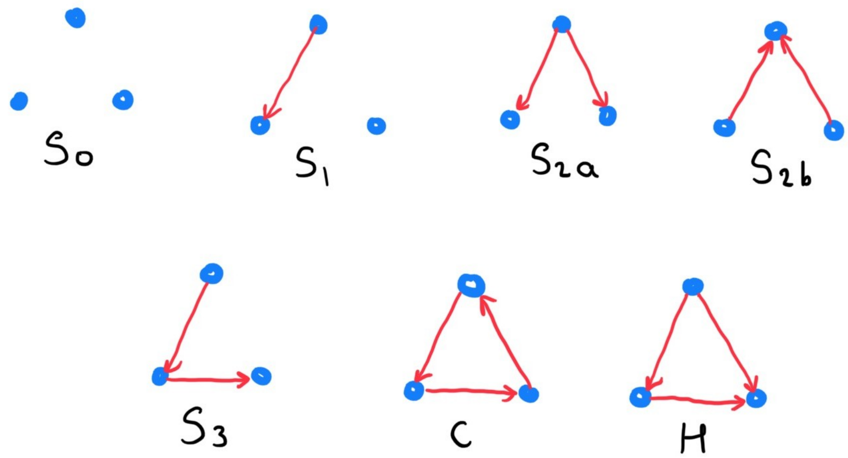
2. If attacker i does not already dominate attackee k, then i dominates k with probability π.

3. If dominance has not yet been established (in either direction) between attacker i and bystander j, then i dominates j with probability θ.

4. If dominance has not yet been established (in either direction) between bystander j and attackee k, then j dominates k with probability θ.

The probability in the events 3 and 4 of the series is assumed to be same to make our model a bit simpler. Also, it is an important observation that according to our model, once an actor i dominates j, this dominance stays forever. This preserves the asymmetry of the dominance relation (i.e., i dominates j only if j does not dominate i), and eventually results in a dominance relation that is complete.

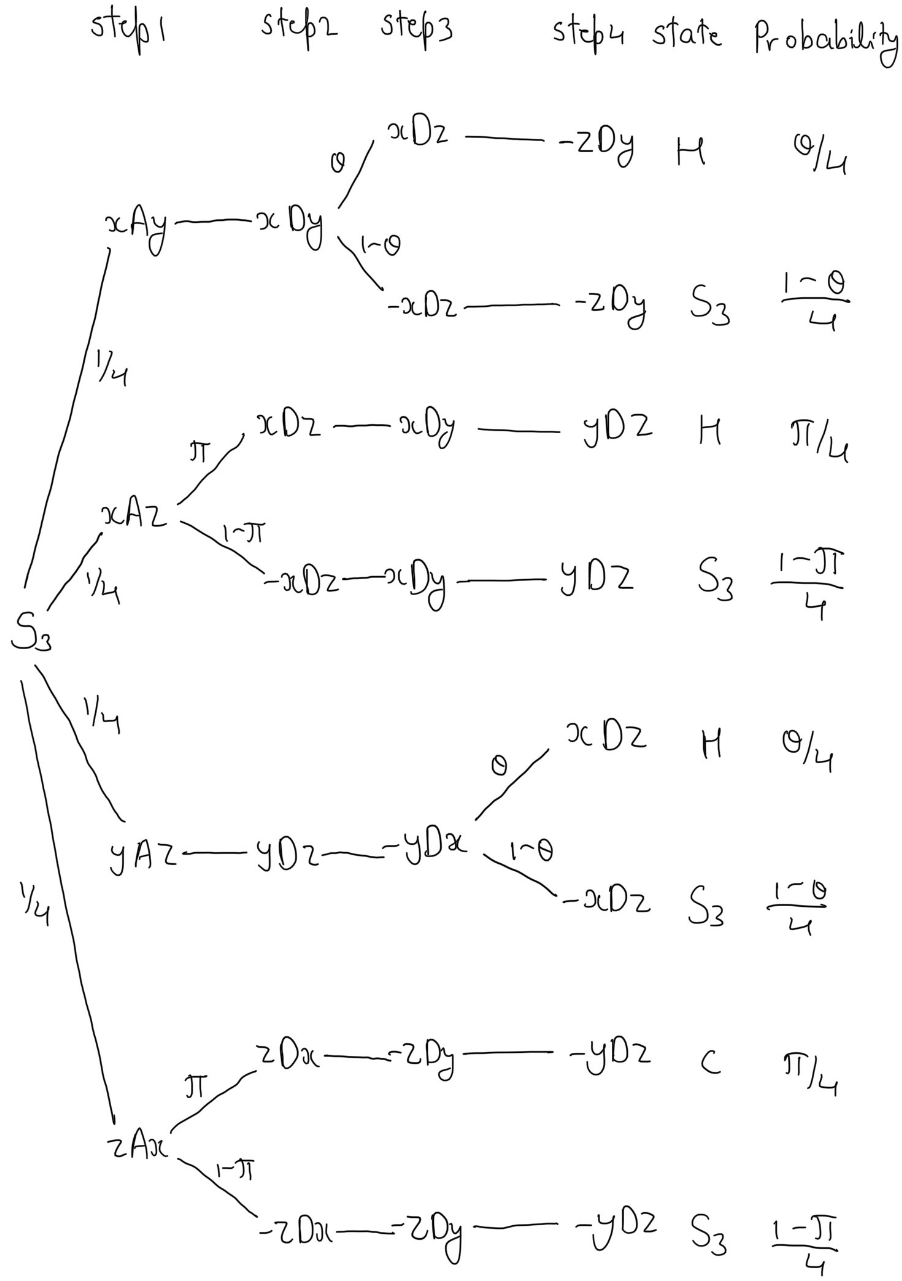
This model can be specified as an Absorbing Markov chain, with the states of the chain given by the possible configurations of the network. For simplicity, we will deal with an example of a group composed of 3 actors. Because we are concerned only with the structure of the network (rather than the identities of the actors), the 7 possible network configurations are depicted below. A directed edge from node i to node j indicates that i dominates j.

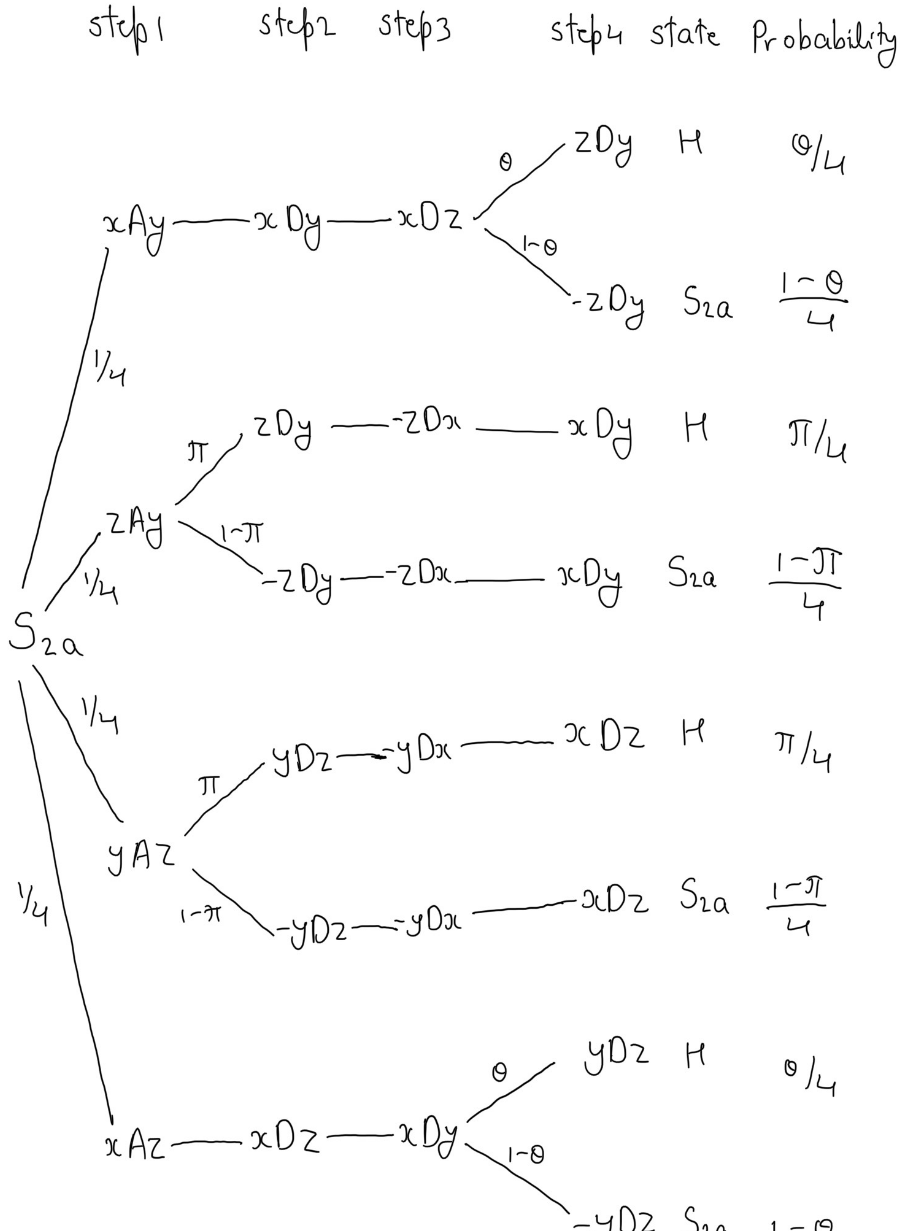


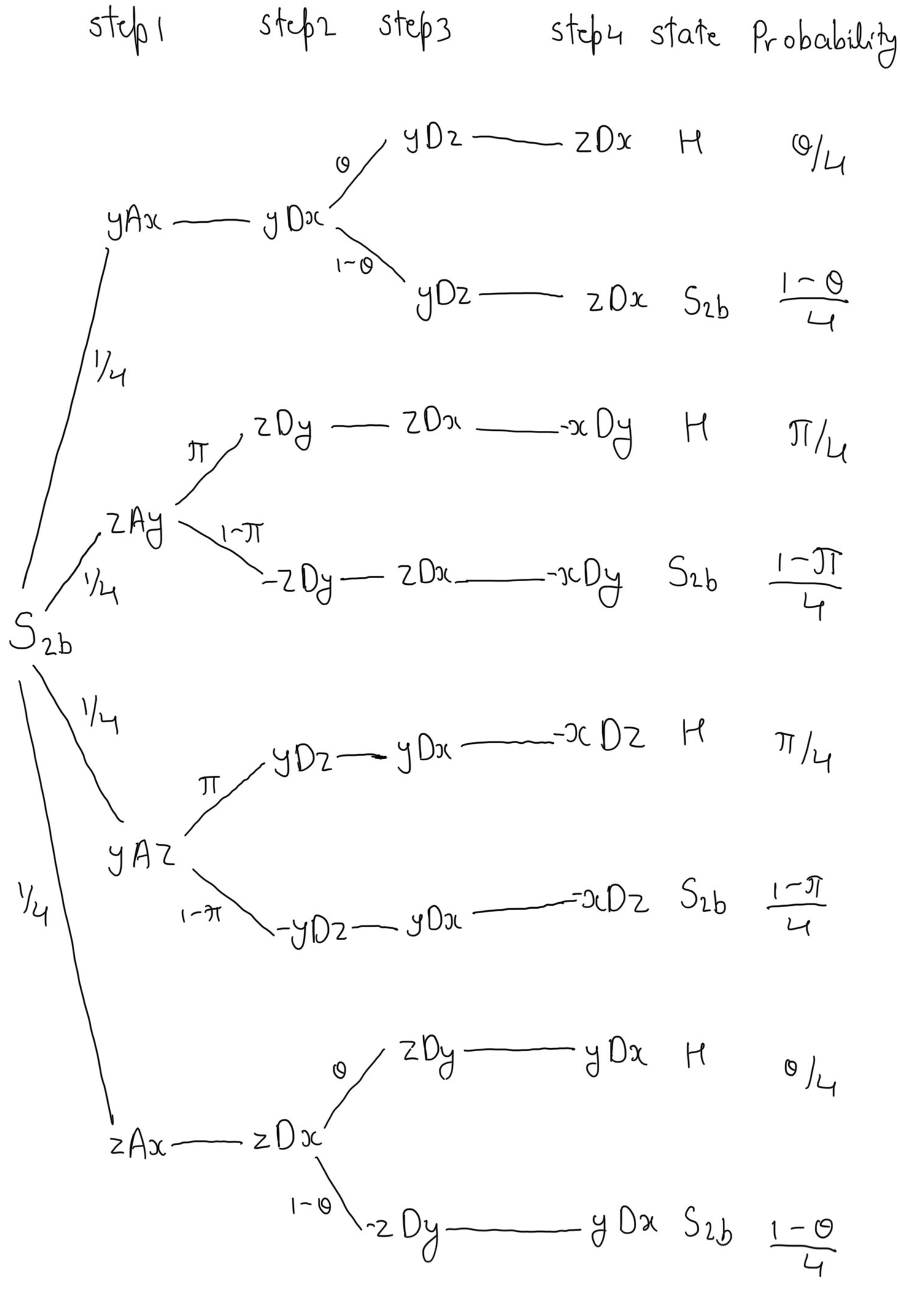
Given the assumptions of the model, the chain is intially in state s_0 (the empty graph), and will eventually reach one of the two absorbing states, C (for "cycle") or H (for "hierarchy").

Having considered the states of the chain, we will now look at the transition matrix. One approach to build the transition matrix is to construct, for each of the non-absorbing states, a probability tree diagram. Using this diagram, we can then determine the probability of moving from that non-absorbing state to each possible state in the next step. The following images depict the probability tree diagrams for s_3, s_{2a}, s_{2b}. The rest of the probabilities have been calculated analytically (i.e., for states s_0, s_1) as their tree diagrams were quite enormous.

(In these diagrams, iAj indicates that i attacks j, iDj indicates that i dominates j, and -iDj indicates that i does not dominate j. x, y, z are the nodes of the graphs, such that in state s_1 the only directed edge is from x to y. The left out node is z (i.e., the labelling of the nodes is the same in all images). This convention is followed in all the images. Unlabelled edges indicate certain (probability 1) outcomes.)


step1 step2 step3 step4 state Probability
S3
1/4
xAy
xDy
θ
1-θ
xDz
-zDy
H
θ/4
-xDz
-zDy
S3
(1-θ)/4
xAz
π
1-π
xDz
xDy
yDz
H
π/4
-xDz
xDy
yDz
S3
(1-π)/4
yAz
yDz
-yDx
θ
1-θ
xDz
H
θ/4
-xDz
S3
(1-θ)/4
zAx
π
1-π
zDx
-zDy
-yDz
C
π/4
-zDx
-zDy
-yDz
S3
(1-π)/4


step1 step2 step3 step4 state Probability
S2a
1/4
xAy — xDy — xDz
θ
zDy H θ/4
1−θ
−zDy S2a (1−θ)/4
zAy
π
zDy — −zDx — xDy H π/4
1−π
−zDy — −zDx — xDy S2a (1−π)/4
yAz
π
yDz — −yDx — xDz H π/4
1−π
−yDz — −yDx — xDz S2a (1−π)/4
xAz — xDz — xDy
θ
yDz H θ/4
1−θ
−yDz S2a (1−θ)/4



We find that all of the transition probabilities are identical for the states s_{2a} and s_{2b}. That is, the rows of the transition matrix corresponding to s_{2a} and s_{2b} are identical.This makes it possible to merge these two states into a single state. Labeling this state as s_2, the model thus has 6 states, and the transition matrix is given below.

	H	C	s_0	s_1	s_2	s_3
H	1	0	0	0	0	0
C	0	1	0	0	0	0
s_0	$\pi\theta^2$	0	$(1-\pi)(1-\theta)^2$	$\pi(1-\theta)^2 + 2(1-\pi)\theta(1-\theta)$	$2\pi\theta(1-\theta)$	$(1-\pi)\theta^2$
s_1	$\frac{\theta^2+4\pi\theta}{5}$	0	0	$\frac{(1-\theta)^2+4(1-\pi)(1-\theta)}{5}$	$\frac{2[(1-\pi)\theta+(\pi+\theta)(1-\theta)]}{5}$	$\frac{2[\pi(1-\theta)+(1-\pi)\theta]}{5}$
s_2	$\frac{\theta+\pi}{2}$	0	0	0	$\frac{2-\theta-\pi}{2}$	0
s_3	$\frac{2\theta+\pi}{4}$	$\frac{\pi}{4}$	0	0	0	$\frac{2-\theta-\pi}{2}$

Having developed the model, we will now assess whether the argument about thebystander effect is true or not. In particular, we will examine how the strength of the direct effect (given by the parameter π) and the bystander effect (given by the parameter θ) influence the probability that the chain is eventually absorbed into the transitive hierarchy rather than the non-transitive cycle.

To begin, lets suppose a moderately strong direct effect ($\pi = 0.5$) but no bystander effect ($\theta = 0$). In the usual way, we can determine the long-run outcome probability by raising the transition matrix to some high power.

$$
\begin{array}{cccccccc}
 & & H & C & s_0 & s_1 & s_2 & s_3 \\
 & H & 1 & 0 & 0 & 0 & 0 & 0 \\
 & C & 0 & 1 & 0 & 0 & 0 & 0 \\
P = & s_0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\
 & s_1 & 0 & 0 & 0 & 0.6 & 0.2 & 0.2 \\
 & s_2 & 0.25 & 0 & 0 & 0 & 0.75 & 0 \\
 & s_3 & 0.125 & 0.125 & 0 & 0 & 0 & 0.75
\end{array}
$$

$$
\begin{array}{cccccccc}
 & & H & C & s_0 & s_1 & s_2 & s_3 \\
 & H & 1 & 0 & 0 & 0 & 0 & 0 \\
 & C & 0 & 1 & 0 & 0 & 0 & 0 \\
P^{100} = & s_0 & 0.75 & 0.25 & 0 & 0 & 0 & 0 \\
 & s_1 & 0.75 & 0.25 & 0 & 0 & 0 & 0 \\
 & s_2 & 1 & 0 & 0 & 0 & 0 & 0 \\
 & s_3 & 0.5 & 0.5 & 0 & 0 & 0 & 0
\end{array}
$$

We are assuming that the chain always begins in state s_0. Thus, from the 3^{rd} row of the P^{100} matrix, we find there is a 75% chance that the chain is ultimately absorbed in the transitive hierarchy (state H), and hence a 25% chance that the chain is absorbed into the non-transitive cycle (state C). It is in this baseline case that we are essentially flipping a coin to determine the direction of dominance for each pair.

Now, keeping $\pi = 0.5$ let's consider what happens when we introduce a weak bystander effect ($\theta = 0.2$).

$$P = \begin{array}{c|cccccc} & H & C & s_0 & s_1 & s_2 & s_3 \\ \hline H & 1 & 0 & 0 & 0 & 0 & 0 \\ C & 0 & 1 & 0 & 0 & 0 & 0 \\ s_0 & 0.2 & 0 & 0.32 & 0.48 & 0.16 & 0.02 \\ s_1 & 0.088 & 0 & 0 & 0.448 & 0.264 & 0.2 \\ s_2 & 0.35 & 0 & 0 & 0 & 0.65 & 0 \\ s_3 & 0.225 & 0.125 & 0 & 0 & 0 & 0.65 \end{array}$$

$$P^{100} = \begin{array}{c|cccccc} & H & C & s_0 & s_1 & s_2 & s_3 \\ \hline H & 1 & 0 & 0 & 0 & 0 & 0 \\ C & 0 & 1 & 0 & 0 & 0 & 0 \\ s_0 & 0.8982 & 0.1018 & 0 & 0 & 0 & 0 \\ s_1 & 0.8706 & 0.1294 & 0 & 0 & 0 & 0 \\ s_2 & 1 & 0 & 0 & 0 & 0 & 0 \\ s_3 & 0.6429 & 0.3571 & 0 & 0 & 0 & 0 \end{array}$$

Again assuming that the chain begins in state s_0, there is now a greater chance (89.82%) that the network is ultimately transitive. It is consistent with our original argument; the bystander effect seems to promote transitivity.

The following table gives the probability that the Absorbing Markov chain is absorbed by the state H after the 1000^{th} step for different values of θ and π, given that the chain started in state s_0.

0	1	1	1	1	1	1	1	1	1	1
0.75	0.9067	0.9372	0.9493	0.9552	0.9585	0.9602	0.9609	0.9609	0.9603	0.9591
0.75	0.8774	0.9137	0.9285	0.9355	0.9388	0.94	0.9398	0.9385	0.9364	0.9333
0.75	0.8638	0.9037	0.9207	0.9284	0.9314	0.9318	0.9304	0.9278	0.9241	0.9192
0.75	0.8562	0.8996	0.9192	0.9281	0.9313	0.9313	0.9292	0.9255	0.9205	0.9143
0.75	0.8515	0.8982	0.9208	0.9315	0.9357	0.9360	0.9336	0.9295	0.9238	0.9167
0.75	0.8482	0.8981	0.9241	0.9372	0.9428	0.944	0.942	0.938	0.9322	0.925
0.75	0.8459	0.8987	0.9281	0.944	0.9517	0.9542	0.9533	0.9499	0.9448	0.9382
0.75	0.8442	0.8997	0.9325	0.9513	0.9615	0.9659	0.9665	0.9645	0.9607	0.9556
0.75	0.8429	0.9009	0.937	0.959	0.9718	0.9785	0.9811	0.9811	0.9792	0.9763
0.75	0.8418	0.9021	0.9415	0.9667	0.9825	0.9917	0.9968	0.9991	0.9999	1

In this table, columns correspond to $\theta \in 0, 0.1, \cdots, 1$; rows correspond to $\pi \in 0, 0.1, \cdots, 1$ and entries give the probability that the process is eventually absorbed in state H.

The two results already computed appear in row 6 (corresponding to $\pi = 0.5$), columns 1 and 3 (corresponding to $\theta = 0$ and $\theta = 0.2$). Intuition suggests that a stronger bystander effect should increase the probability of transitivity. However, we can see from the above table that an increase in θ beyond a certain value can actually reduce this probability. For instance, consider the 6th row of the table (so that $\pi = 0.5$). As θ begins to rise from 0, the probability of absorption in H also rises. However, once θ reaches 0.6, this probability reaches a maximum, and any further increase in the bystander effect causes the probability of absorption in H to fall. It is also interesting to note that, keeping the strength of the bystander effect constant, the probability of transitivity is lowest when the direct effect is moderate (not too high, not too low). For instance, consider the final column of the table (so that $\theta = 1$). If the direct effect is either absent ($\pi = 0$) or maximum ($\pi = 1$) then state H is reached with probability 1. But given intermediate values of π,there is some chance that the process will be absored in state C. Overall, analysis of the model suggests that the implications of the bystander effect need more refinement, and the model should have been a little more complex (taking into consideration some other possible factors).