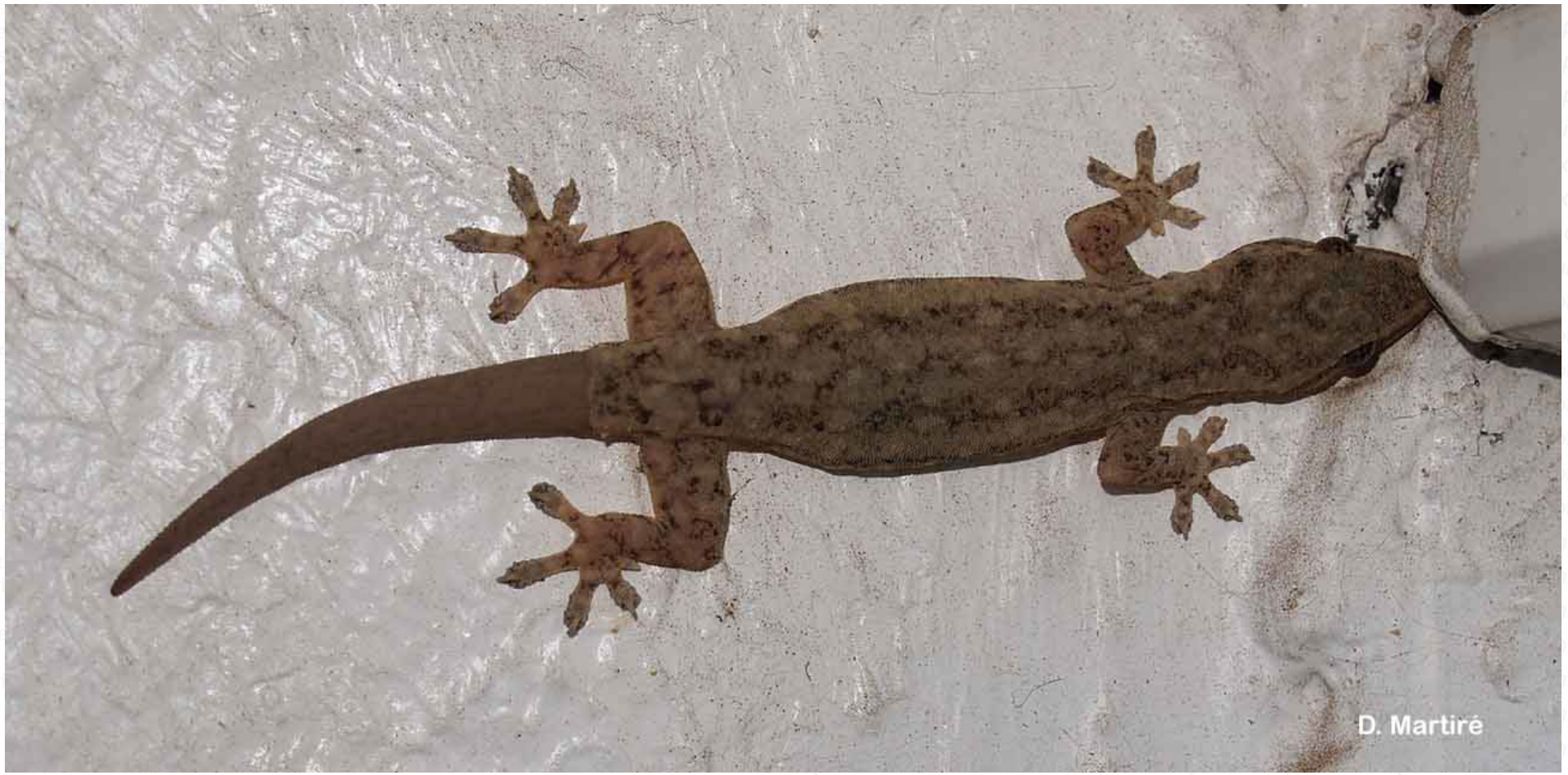
Modeling discrete character evolution on trees

Varanus salvator

Bitis gabonica



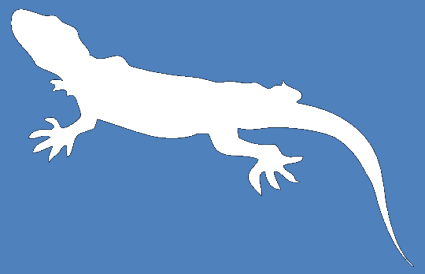
Gehyra mutilata

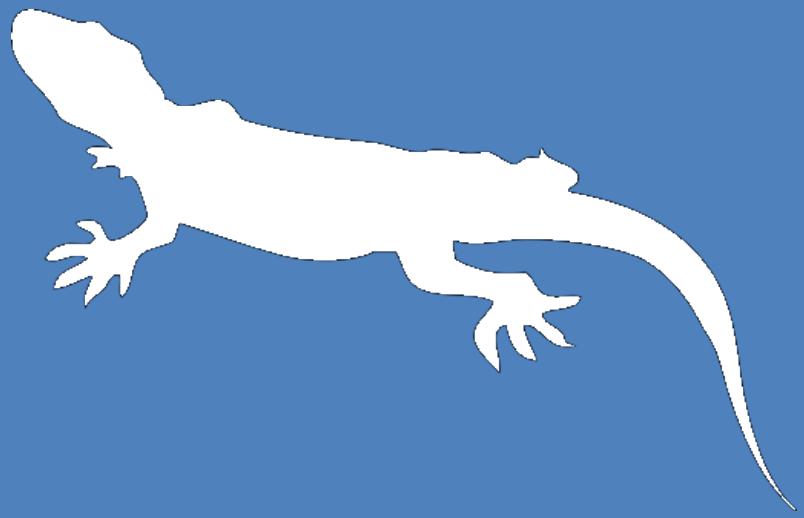
Lialis burtonis

Plestiodon fasciatus

Acontias meleagris

How do we model the process
by which squamates
(snakes and lizards)
lose their limbs?



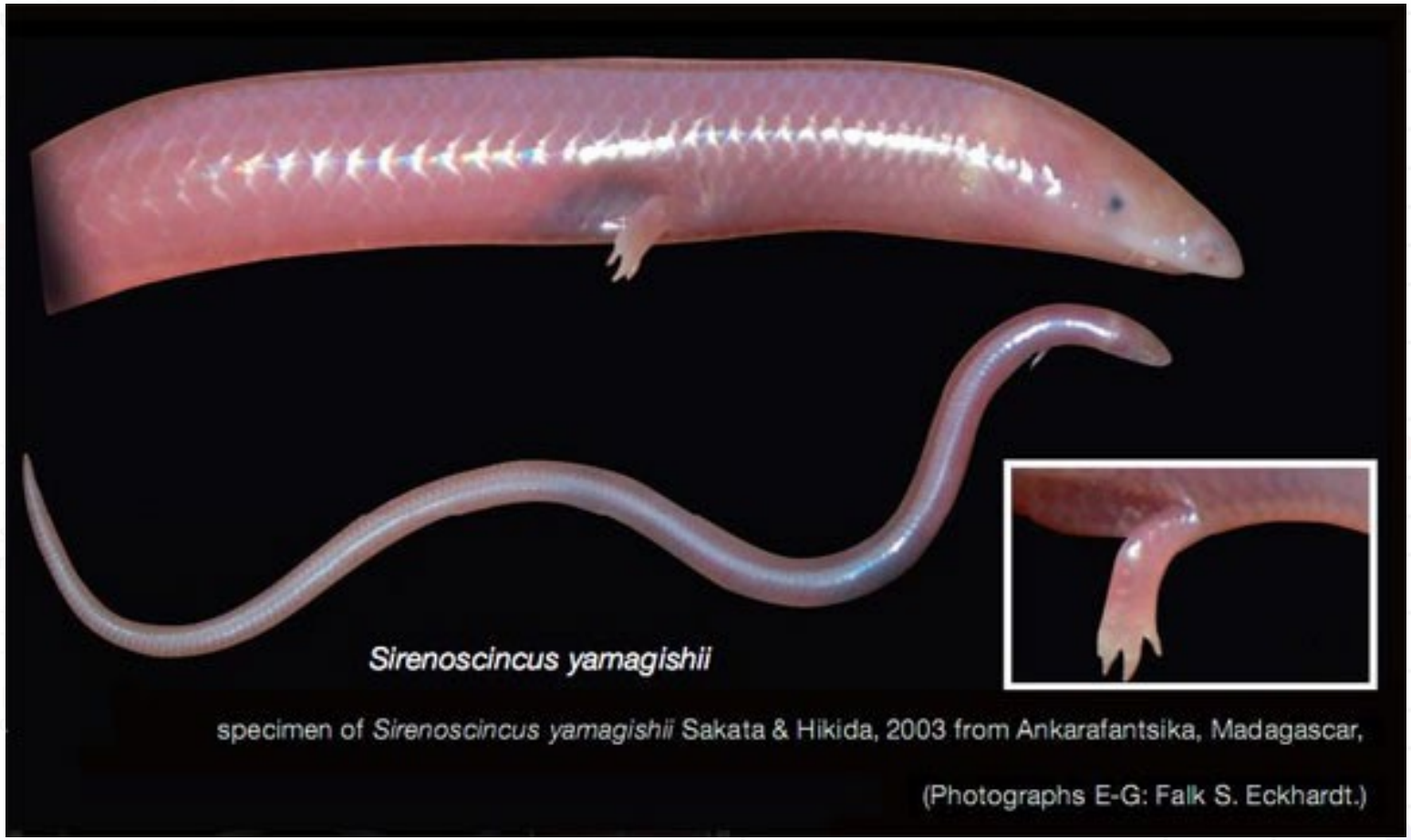


No limbs

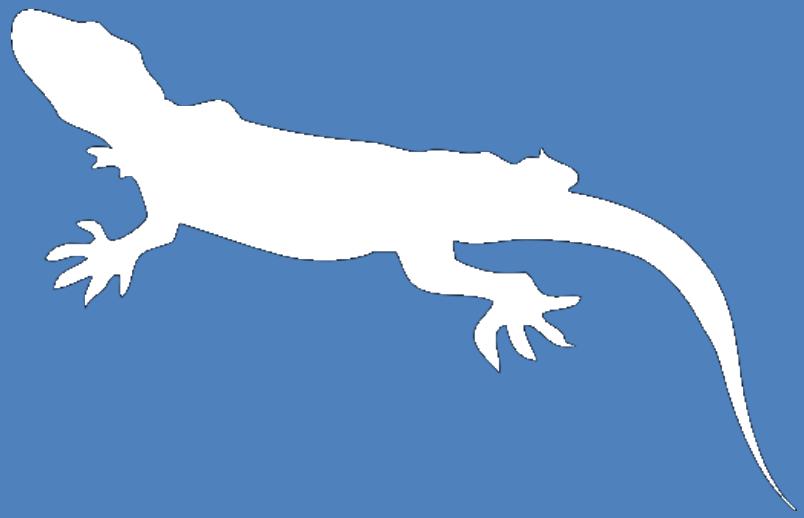
State 0

Limbs

State 1


Sirenoscincus yamagishii
specimen of Sirenoscincus yamagishii Sakata & Hikida, 2003 from Ankarafantsika, Madagascar,
(Photographs E-G: Falk S. Eckhardt.)

Ambloplites cavifrons
Ambloplites rupestris
Ambloplites ariommus
Archolites interruptus
Pomoxis nigromaculatus
Pomoxis annularis
Enneacantus obesus
Centrarchus macropterus
Micropterus dolomieu
Micropterus punctulatus
Micropterus floridanus
Micropterus salmoides
Micropterus treculi
Micropterus notius
Micropterus coosae
Lepomis cyanellus
Lepomis symmetricus
Lepomis gulosus
Lepomis macrochirus
Lepomis humilis
Lepomis megalotis
Lepomis marginatus
Lepomis auritus
Lepomis miniatus
Lepomis punctatus
Lepomis microlophus
Lepomis gibbosus
Acantharchus pomotis



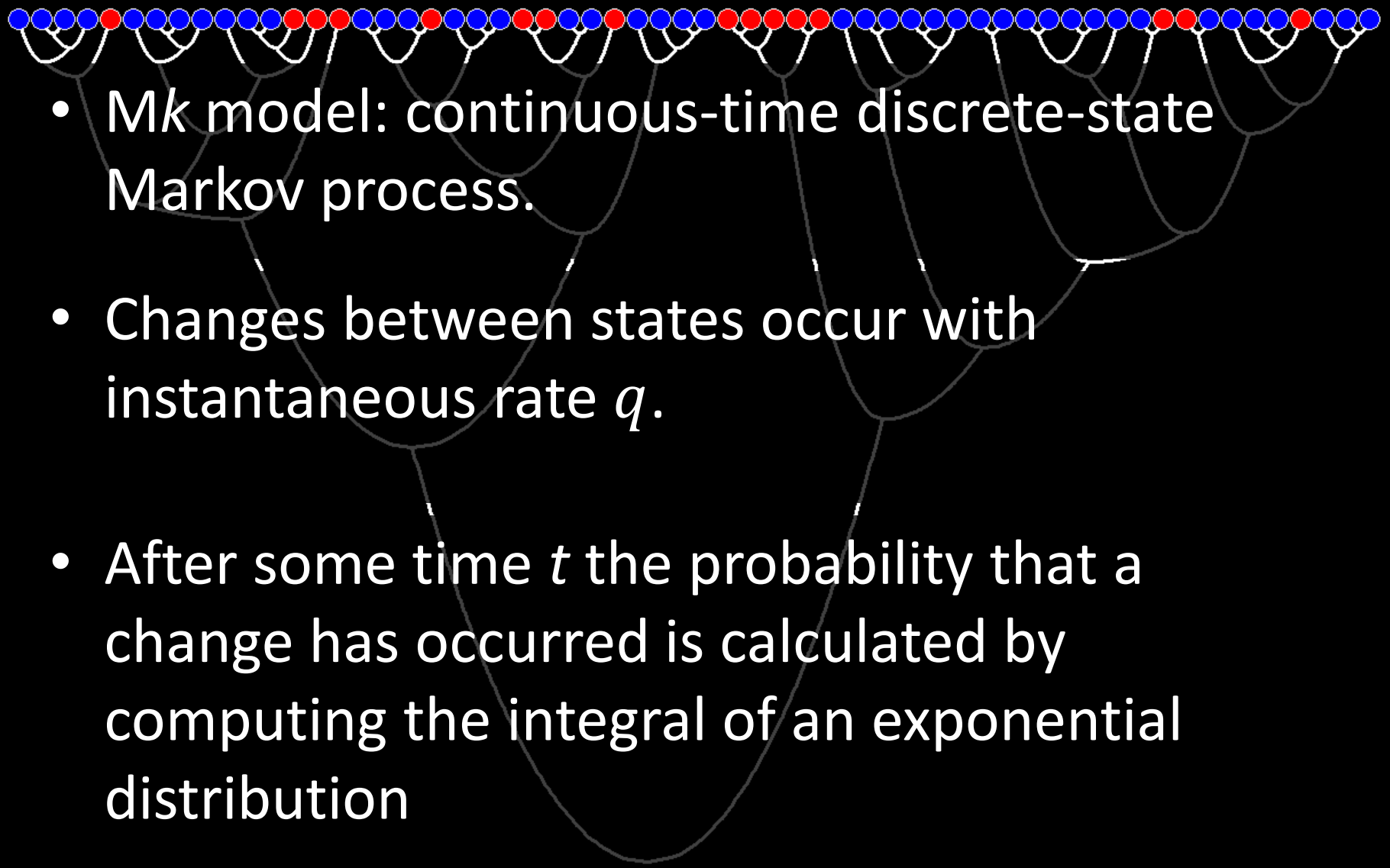
No limbs

State 0

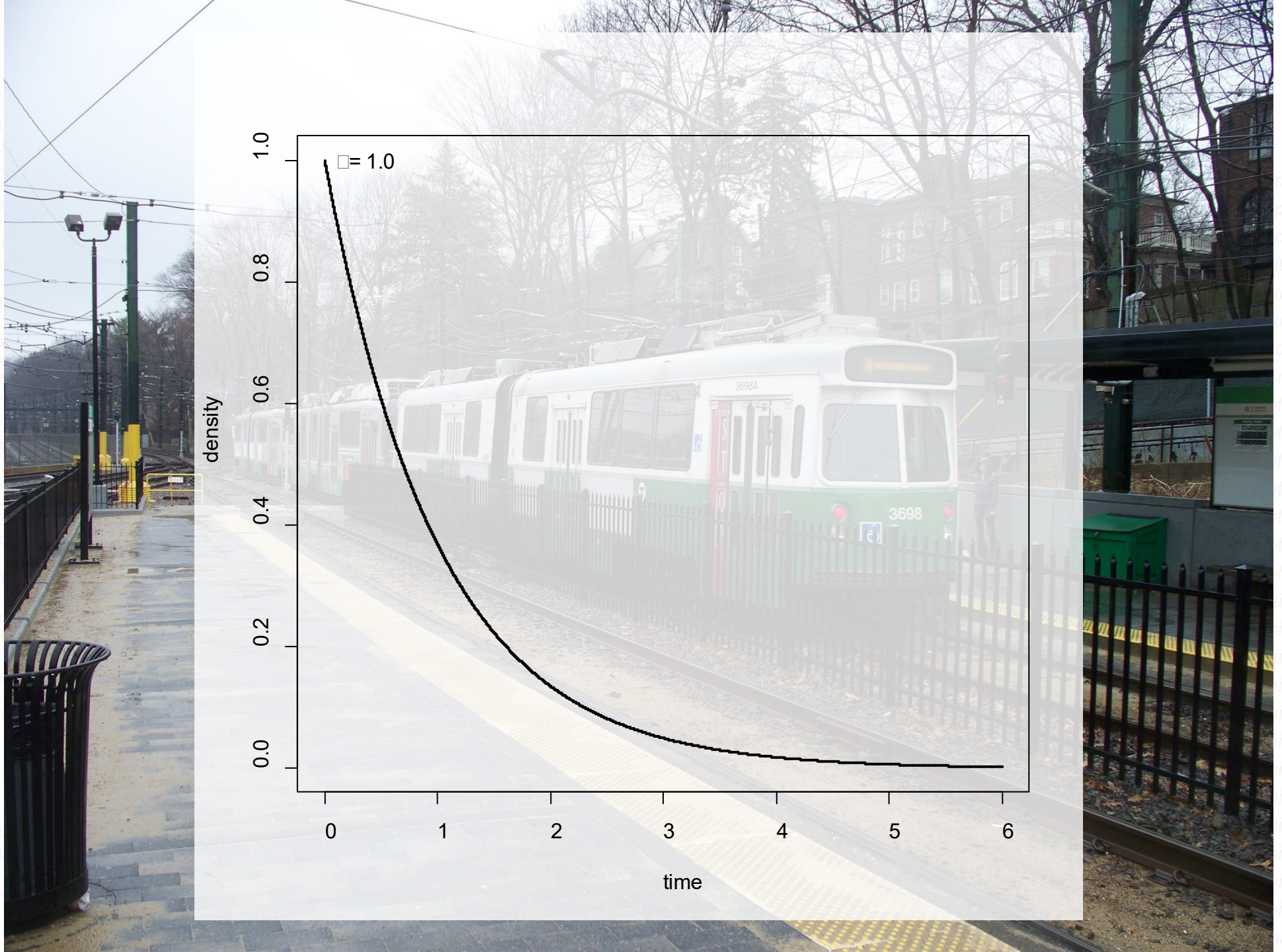
Limbs

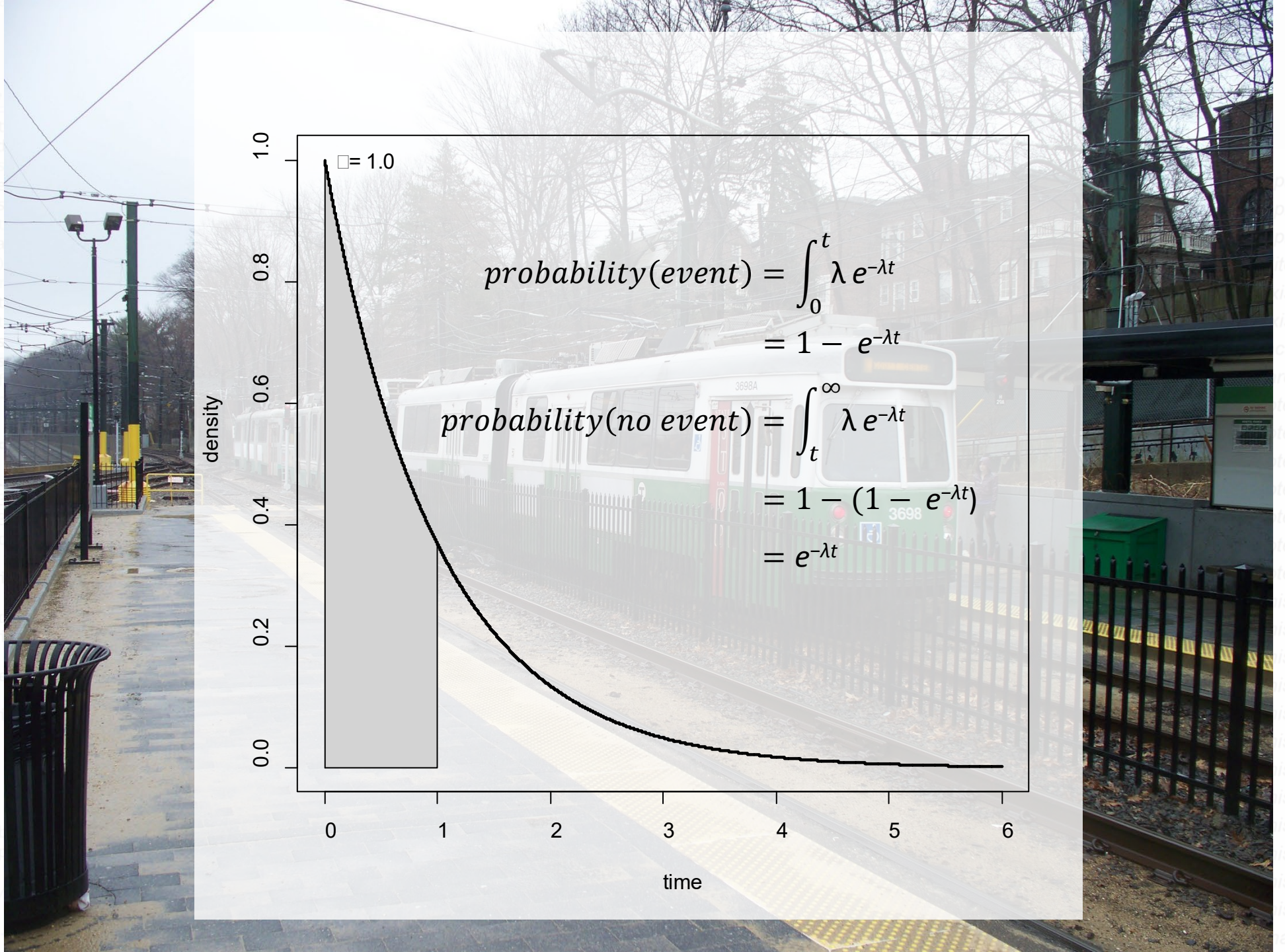
State 1

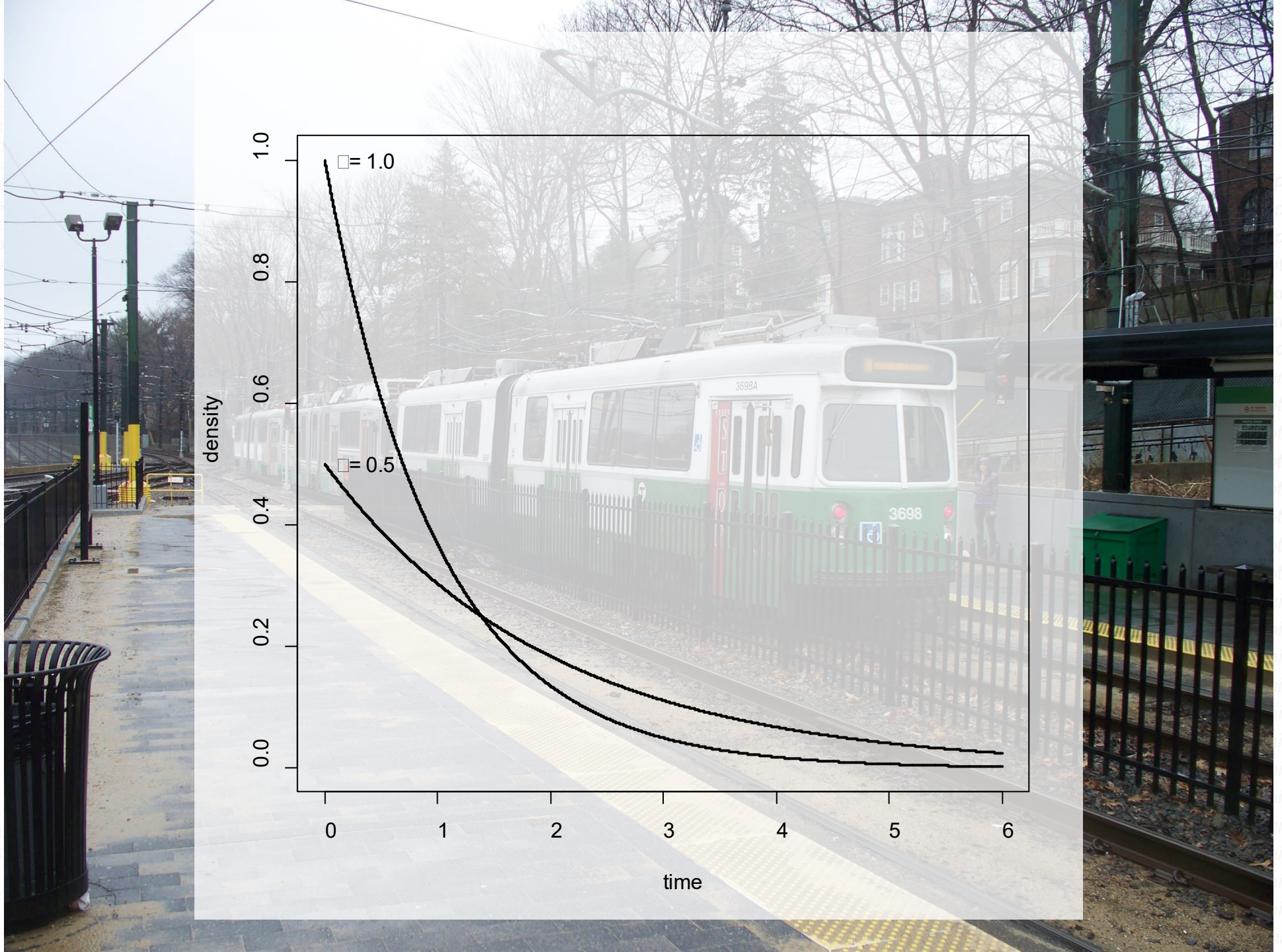
Discrete character evolution

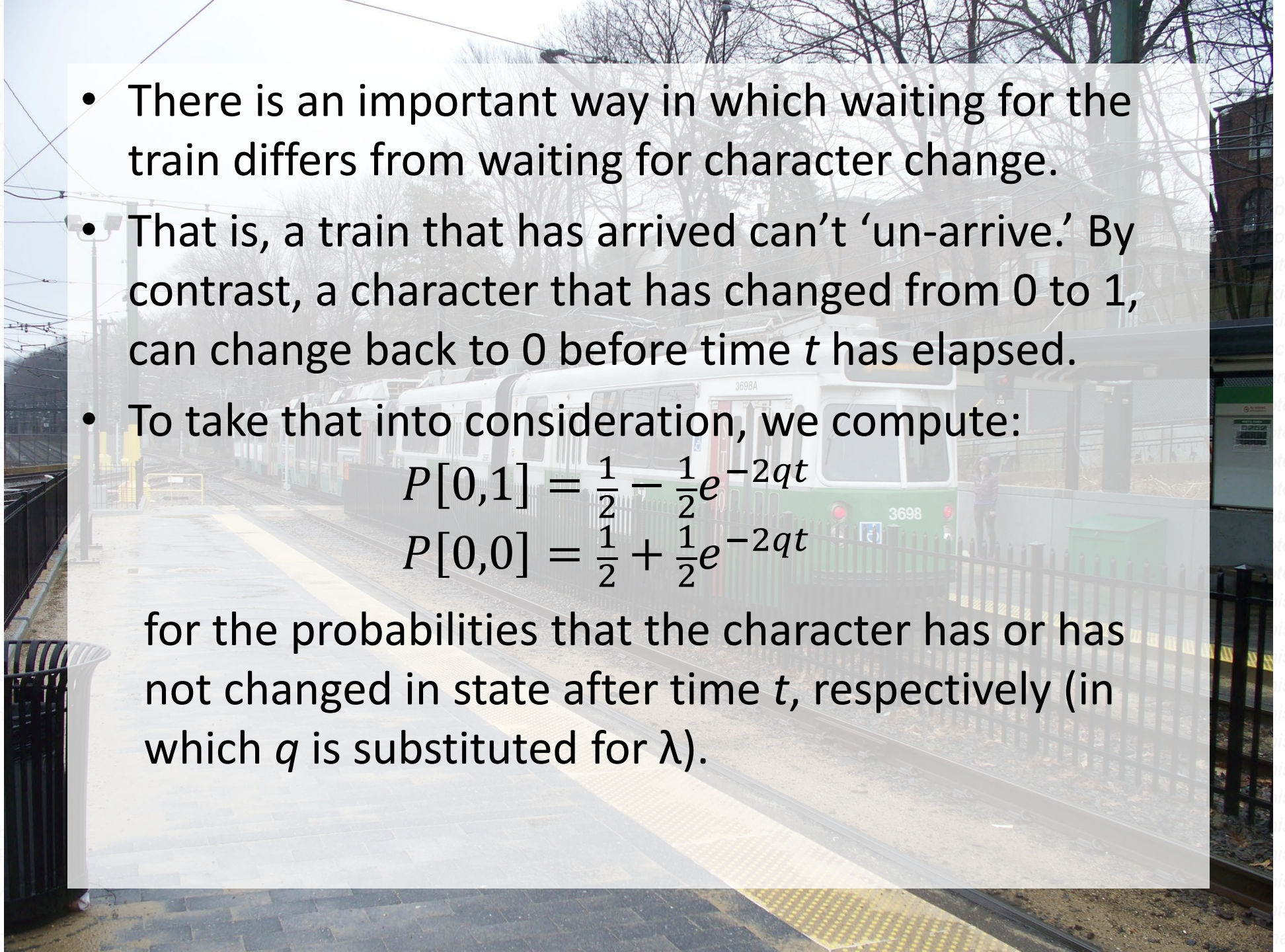


- M*k* model: continuous-time discrete-state Markov process.
- Changes between states occur with instantaneous rate q.
- After some time t the probability that a change has occurred is calculated by computing the integral of an exponential distribution


□= 1.0
density
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
time


□= 1.0
density
time
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
$probability(event) = \int_0^t \lambda\, e^{-\lambda t}$
$= 1 - e^{-\lambda t}$
$probability(no\ event) = \int_t^\infty \lambda\, e^{-\lambda t}$
$= 1 - (1 - e^{-\lambda t})$
$= e^{-\lambda t}$

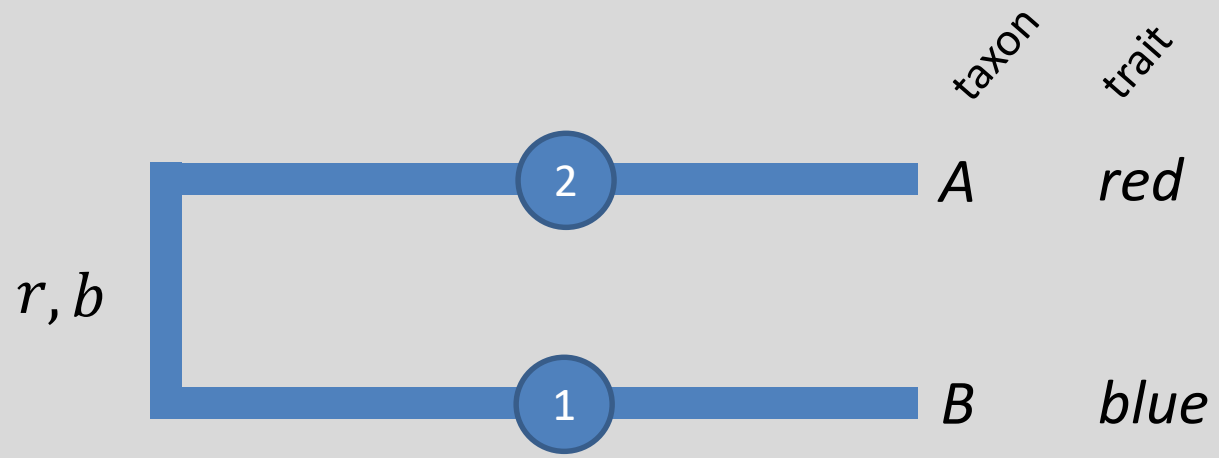

1.0
0.8
0.6
0.4
0.2
0.0
density
= 1.0
= 0.5
0
1
2
3
4
5
6
time



- There is an important way in which waiting for the train differs from waiting for character change.
- That is, a train that has arrived can't 'un-arrive.' By contrast, a character that has changed from 0 to 1, can change back to 0 before time *t* has elapsed.
- To take that into consideration, we compute:

$$P[0,1] = \frac{1}{2} - \frac{1}{2}e^{-2qt}$$
$$P[0,0] = \frac{1}{2} + \frac{1}{2}e^{-2qt}$$

for the probabilities that the character has or has not changed in state after time *t*, respectively (in which *q* is substituted for λ).

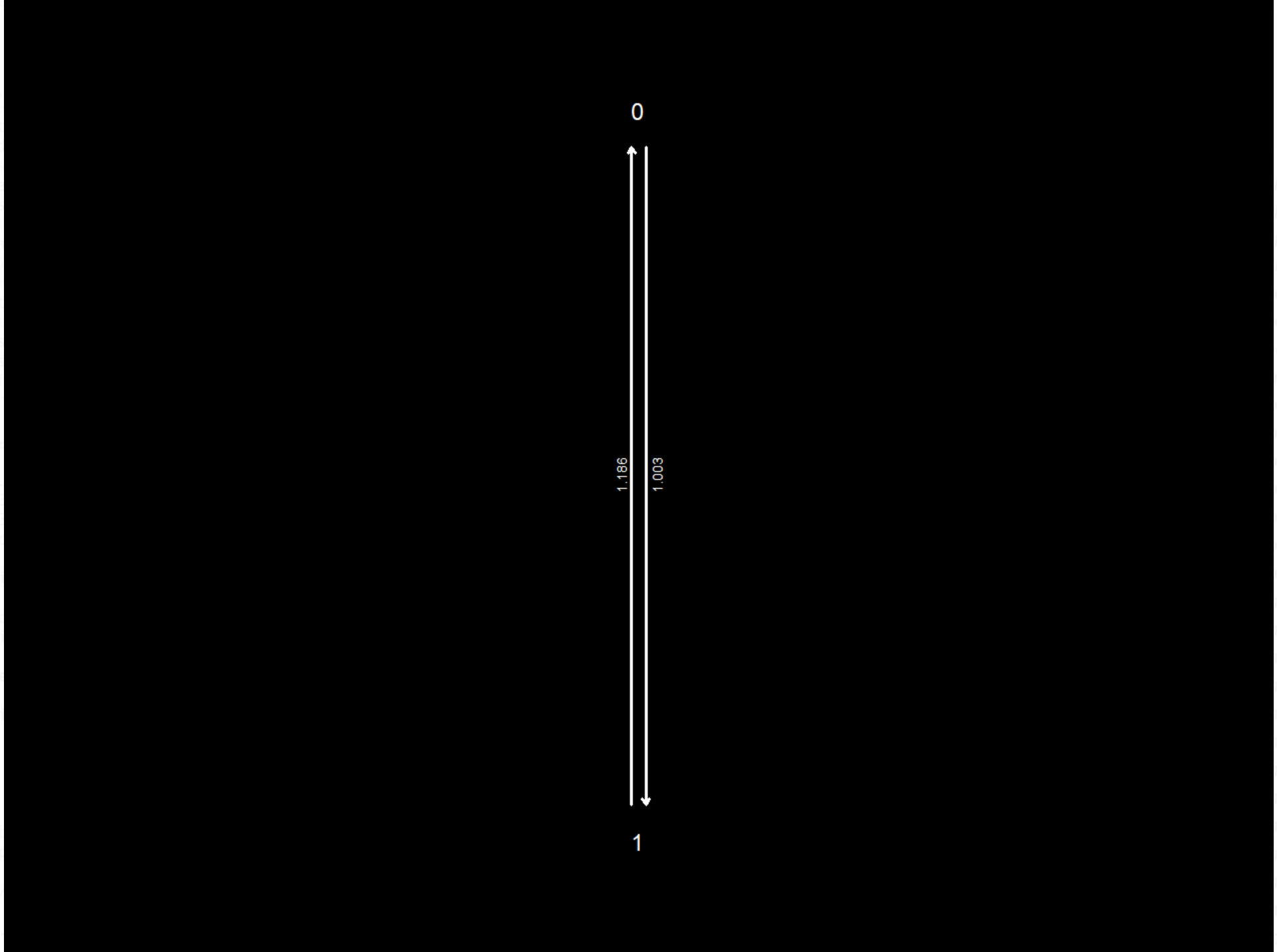


$$P(r,b) = P(b|r)_1 P(r|r)_2 + P(b|b)_1 P(r|b)_2$$

* assuming r & b are equiprobable at the root.

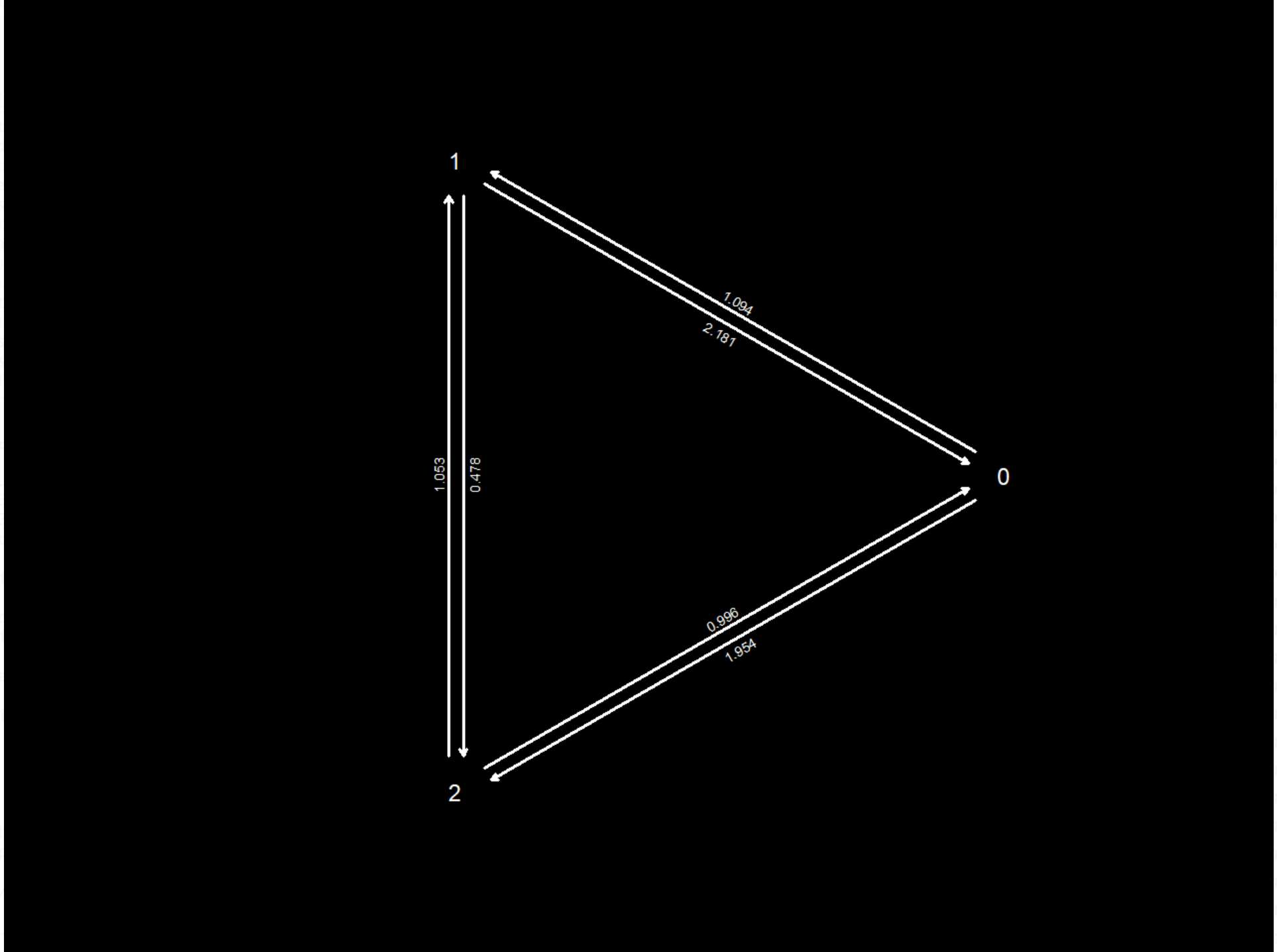
$$\mathbf{Q} = \begin{bmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{bmatrix}$$

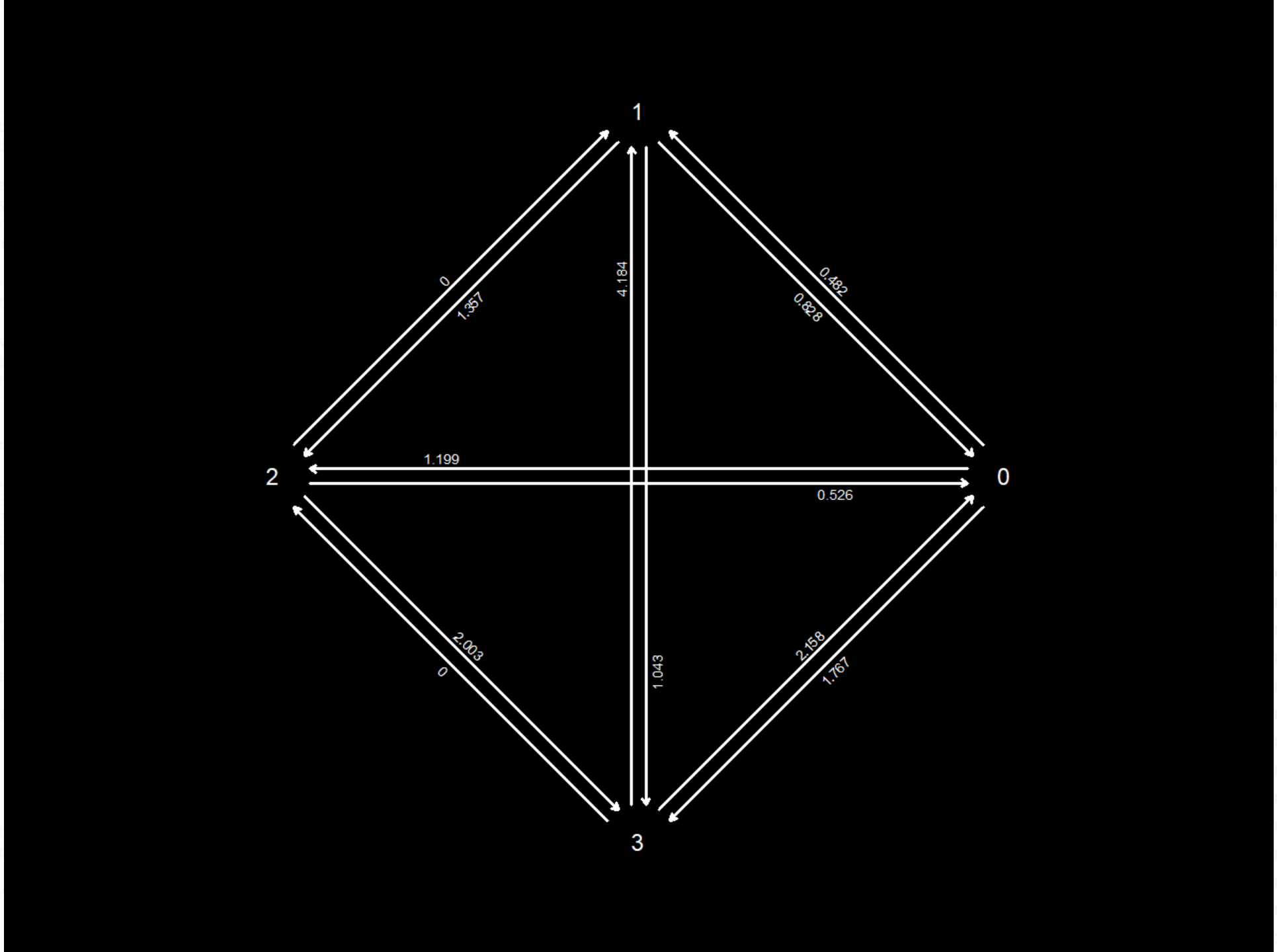
$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{Q}\boldsymbol{t})$$

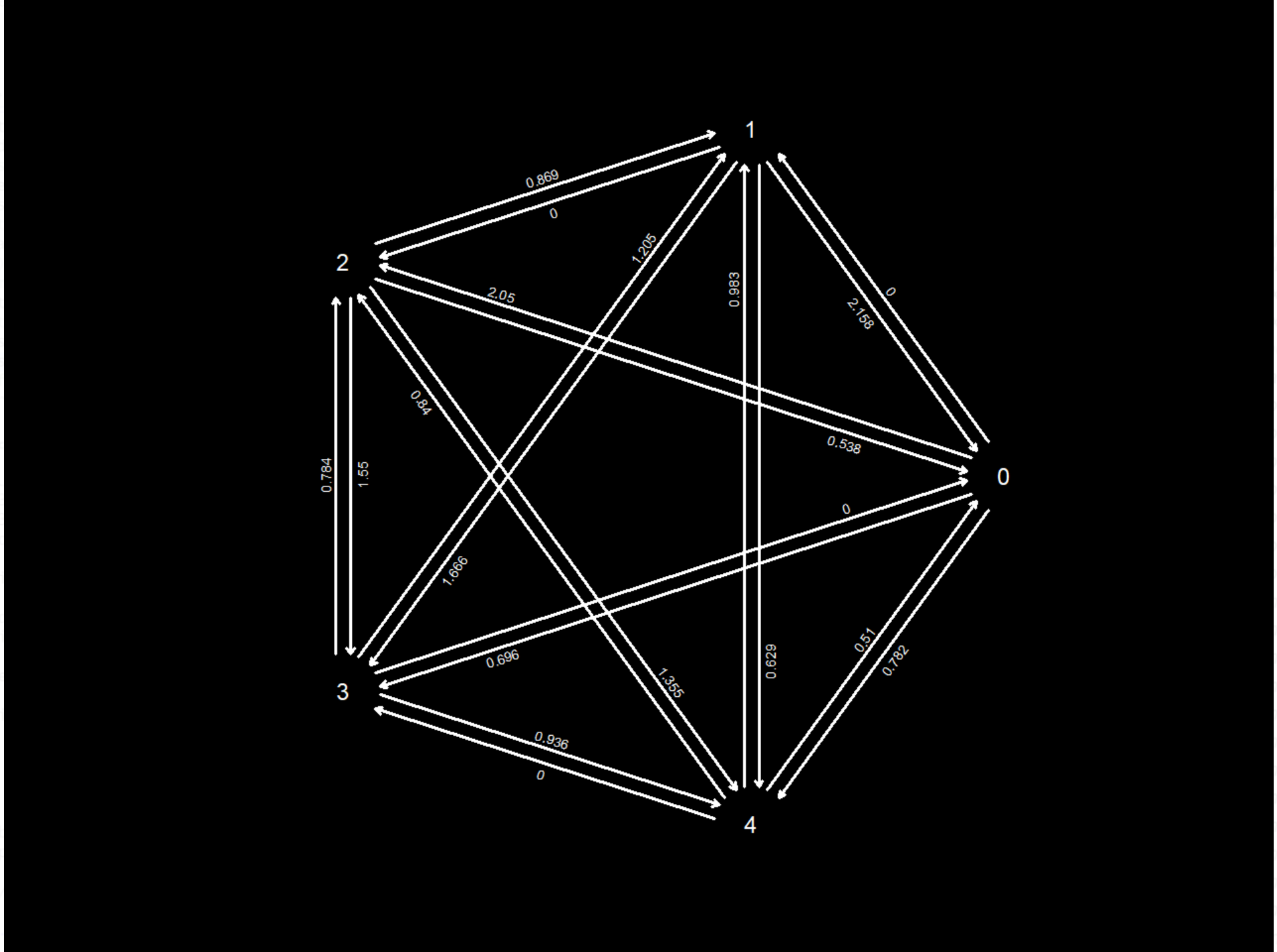

0
1.186
1.003
1

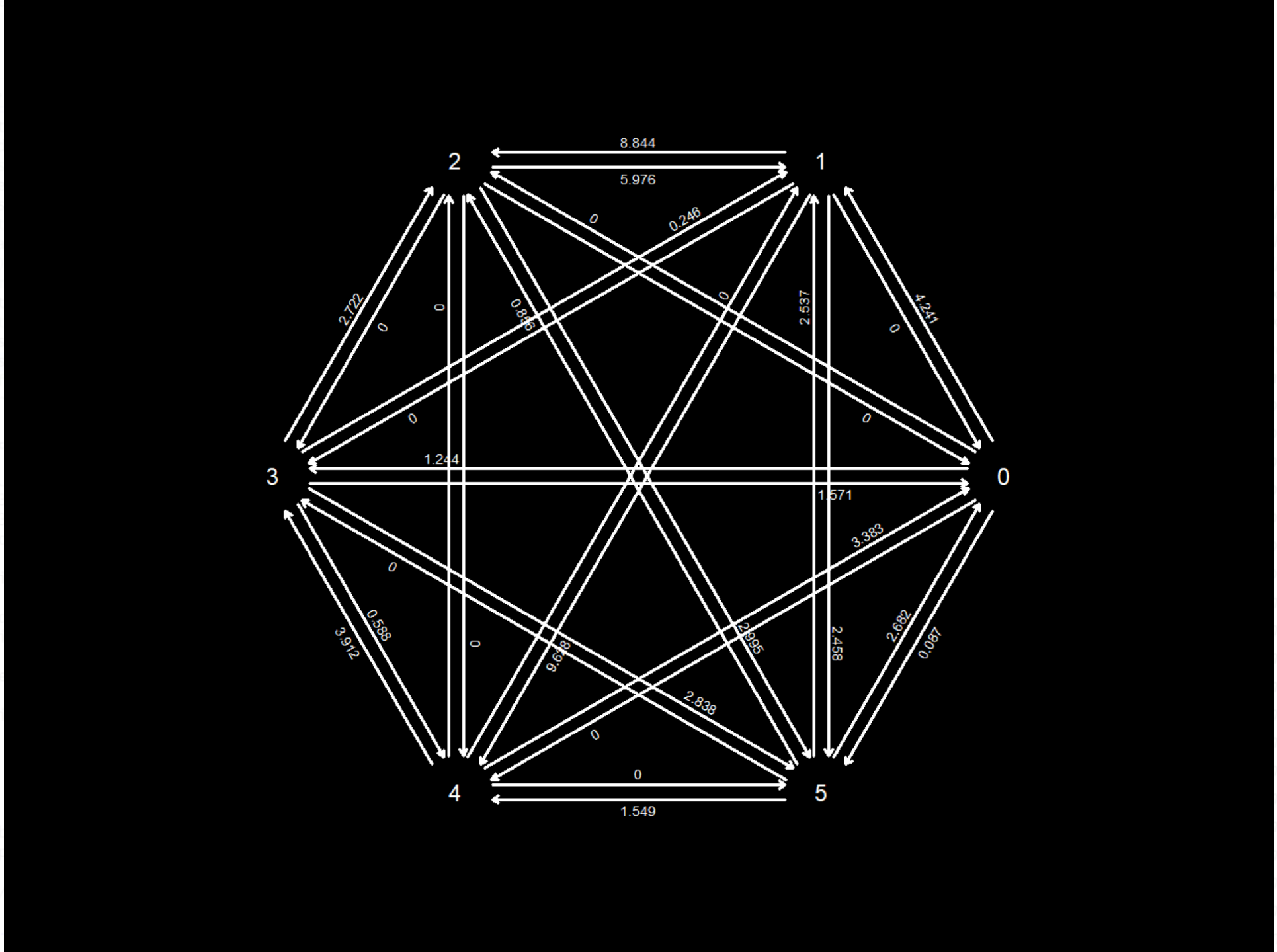
$$\mathbf{Q} = \begin{bmatrix} -(q_{01} + q_{02}) & q_{01} & q_{02} \\ q_{10} & -(q_{10} + q_{12}) & q_{12} \\ q_{20} & q_{21} & -(q_{20} + q_{21}) \end{bmatrix}$$

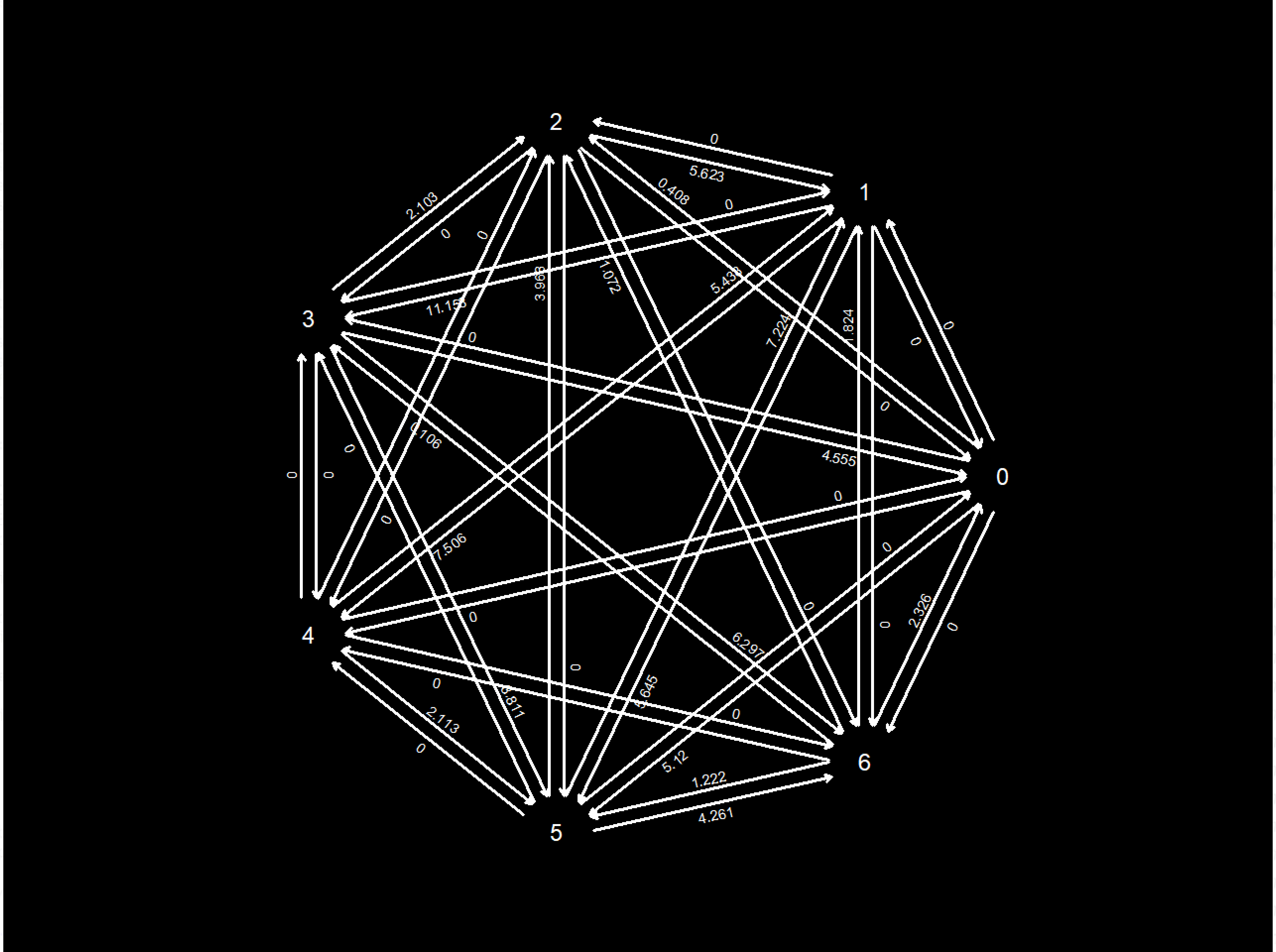
$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{Q}\boldsymbol{t})$$


1
0
2
1.094
2.181
1.053
0.478
0.996
1.954

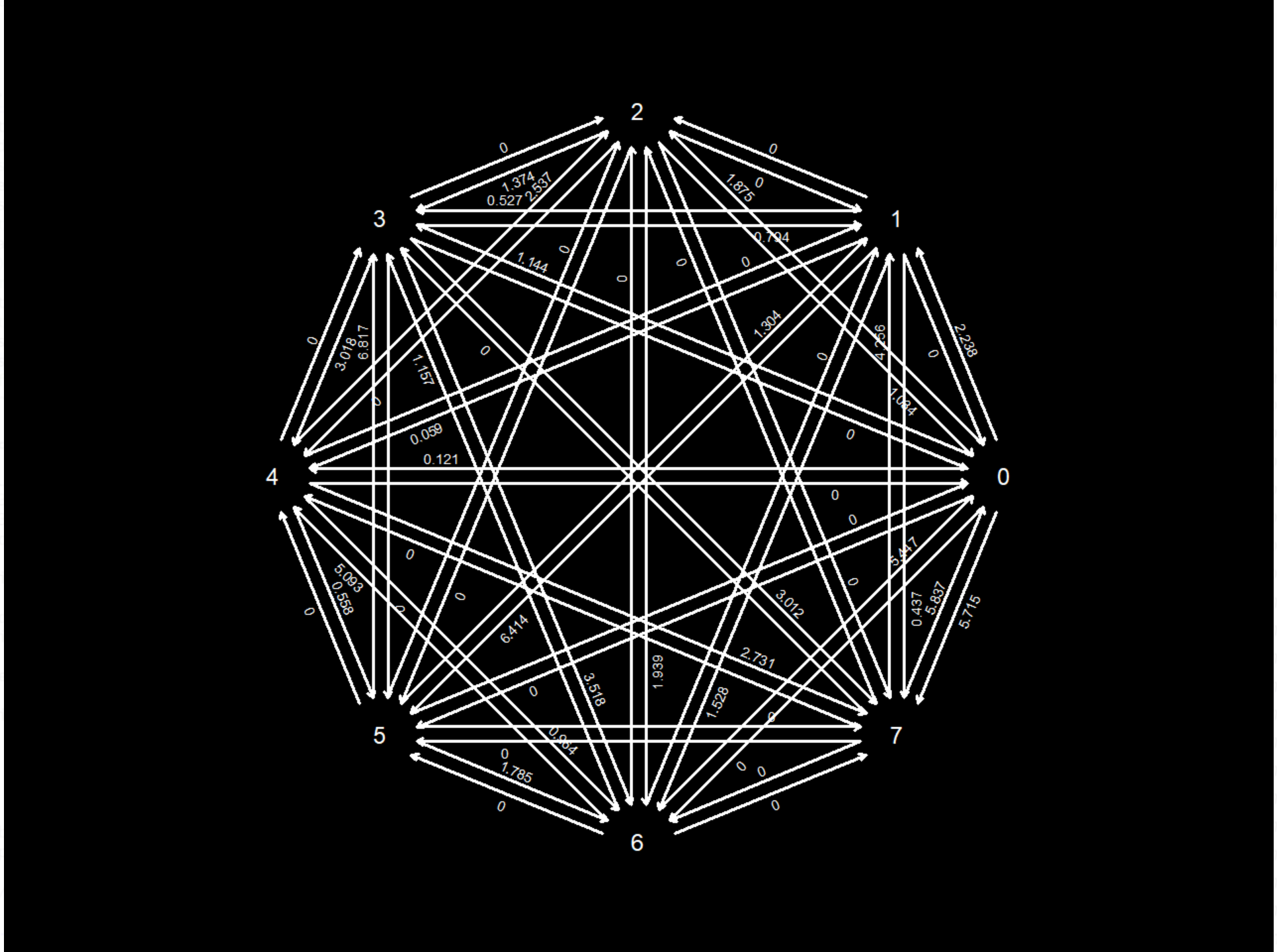

1
2
0
3
0
1.357
4.184
0.482
0.828
1.199
0.526
2.003
0
1.043
2.158
1.767


1
2
3
4
0
0.869
0
1.205
0.983
0
2.158
2.05
0.84
0.538
0.784
1.55
0
1.666
0.696
1.355
0.629
0.51
0.782
0.936
0

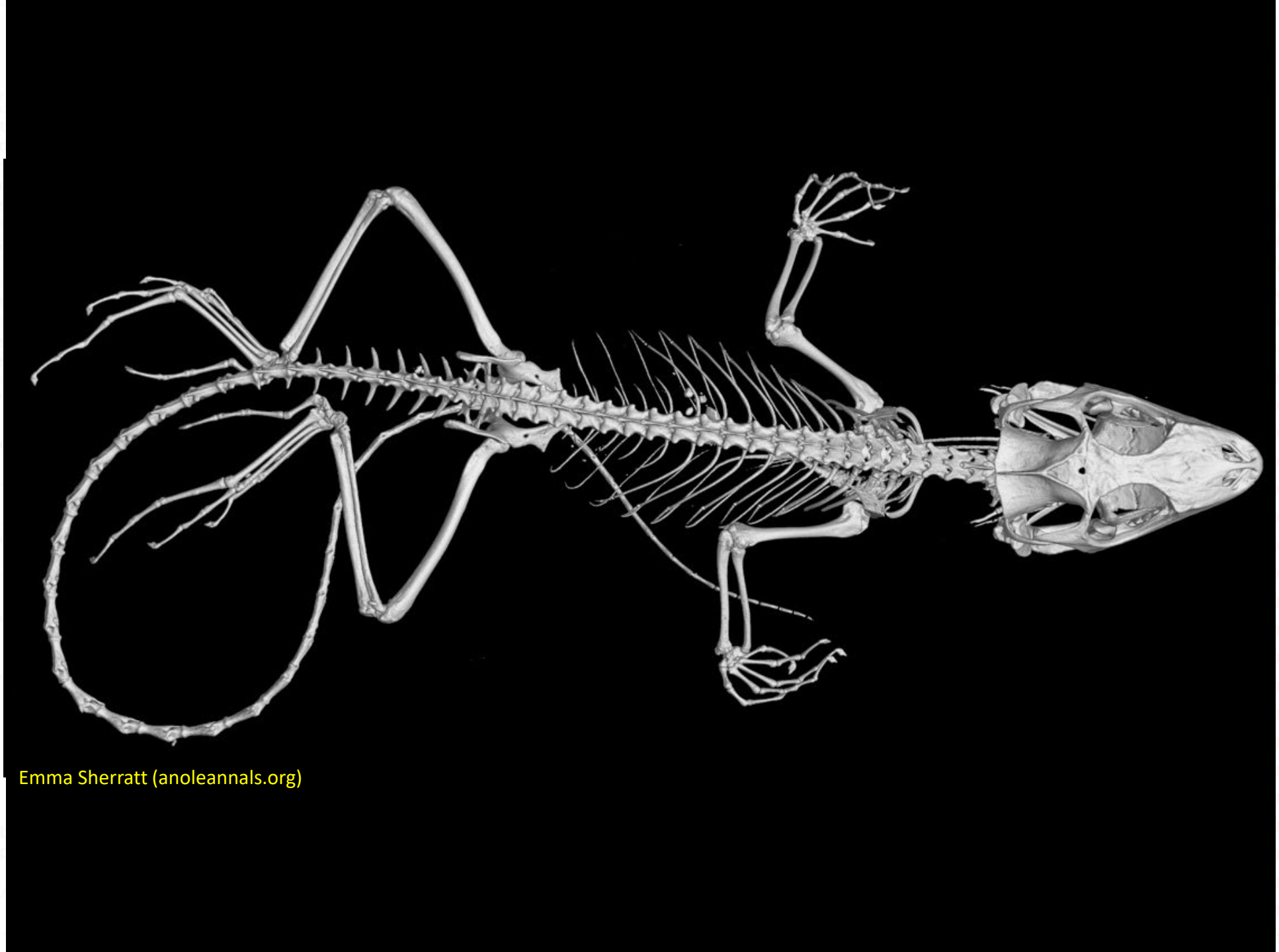

2
1
3
0
4
5
8.844
5.976
0
0.246
2.722
0
0
0.846
0
2.537
4.241
0
0
0
1.244
1.571
3.383
0
0.588
3.912
0
0.816
2.995
2.458
2.682
0.087
2.838
0
0
1.549



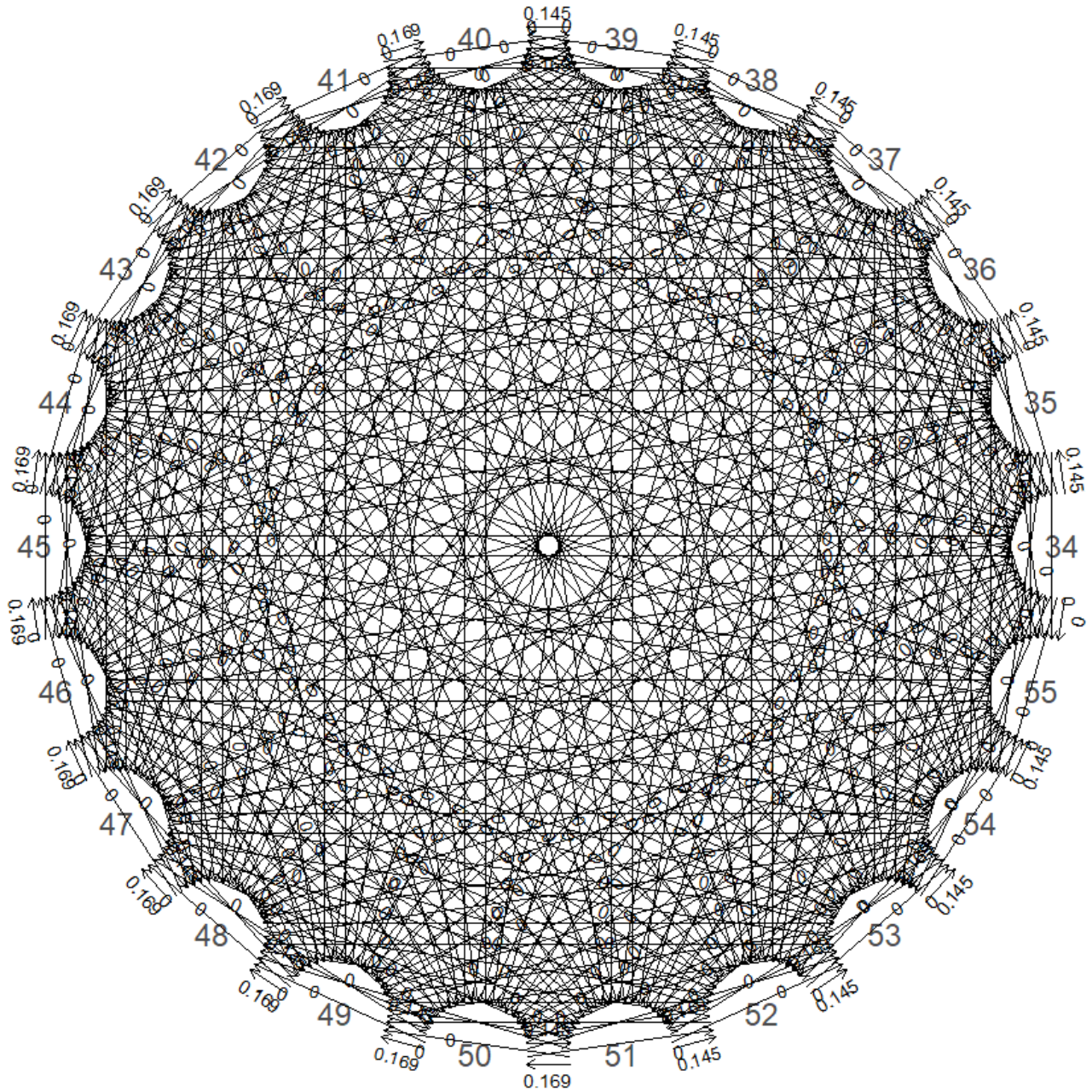
2
1
3
0
4
6
5
0
5.623
0.408
0
2.103
0
0
3.968
1.072
5.433
11.153
7.224
1.824
0
0
0
0
0.106
0
0
0
4.555
0
0
7.506
0
0
2.326
0
0
0
6.297
0
0
3.811
3.645
2.113
0
0
5.12
1.222
4.261


2
3
1
4
0
5
7
6
0
1.374
0.527
2.537
1.875
0.704
1.144
3.018
6.817
1.157
0.059
0.121
1.304
2.238
1.084
5.093
0.558
6.414
3.012
5.447
0.437
5.837
5.715
2.731
3.518
1.939
1.528
0.864
1.785


Phytools
Phylogenetic Tools for
Comparative Biology
(and Other Things)



Emma Sherratt (anoleannals.org)

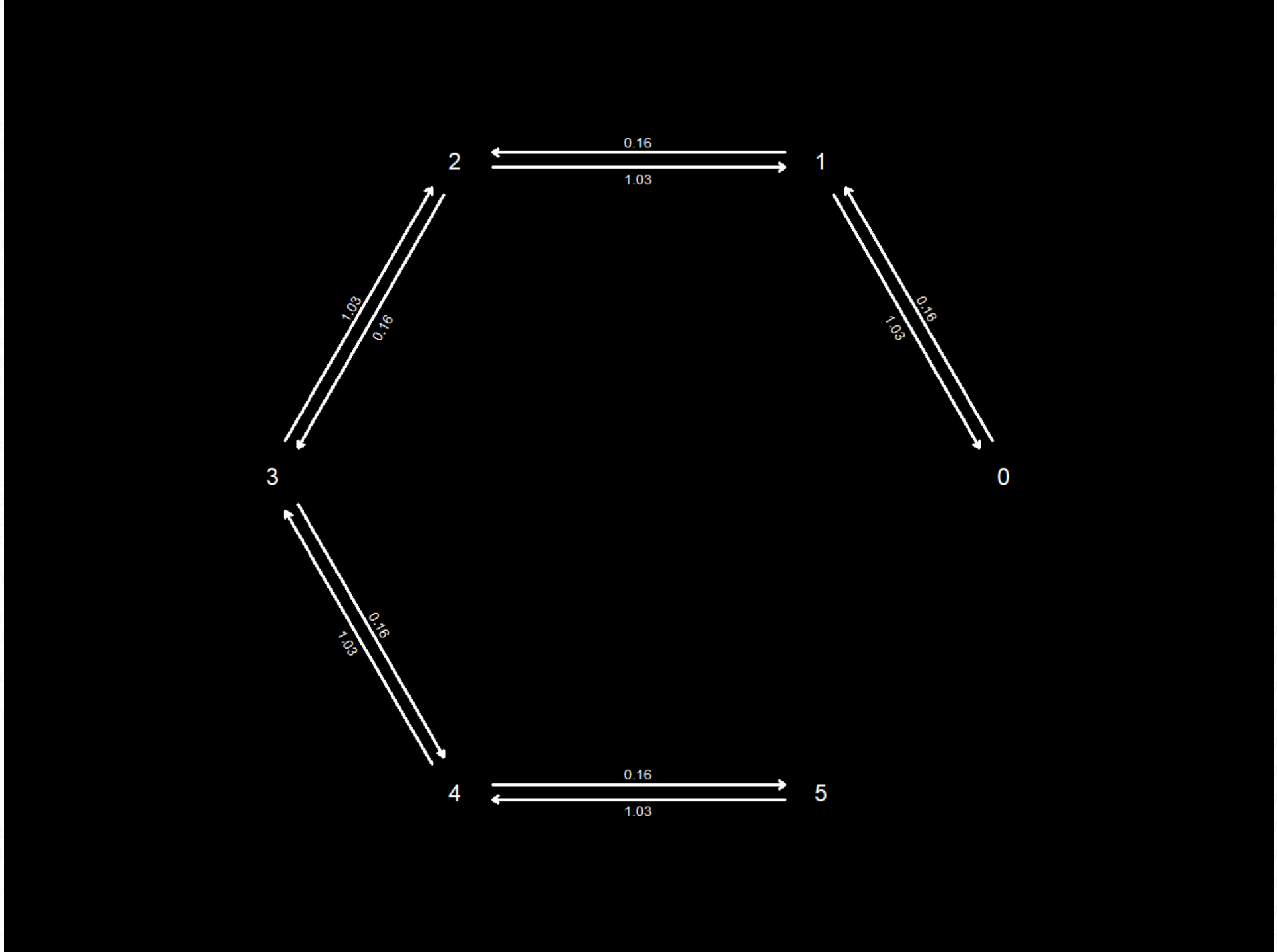

0.145
40
39
38
37
36
35
34
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
0.169

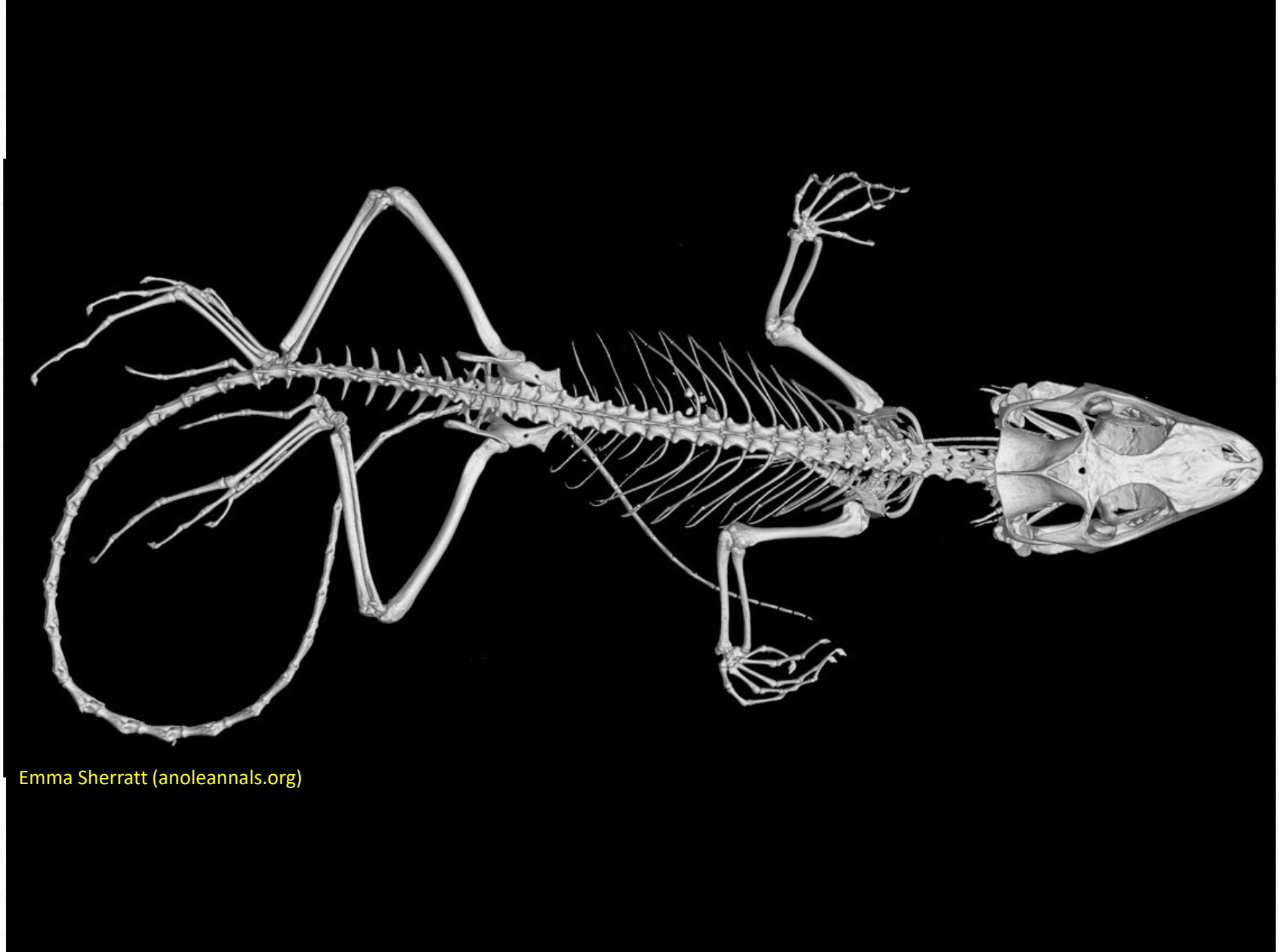
$$\mathbf{Q} = \begin{bmatrix} -(q_{01} + q_{02}) & q_{01} & q_{02} \\ q_{10} & -(q_{10} + q_{12}) & q_{12} \\ q_{20} & q_{21} & -(q_{20} + q_{21}) \end{bmatrix}$$

$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{Q}\boldsymbol{t})$$

$$\mathbf{Q} = \begin{bmatrix} -(q_{01} + q_{02}) & q_{01} & 0 \\ q_{10} & -(q_{10} + q_{12}) & q_{12} \\ 0 & q_{21} & -(q_{20} + q_{21}) \end{bmatrix}$$

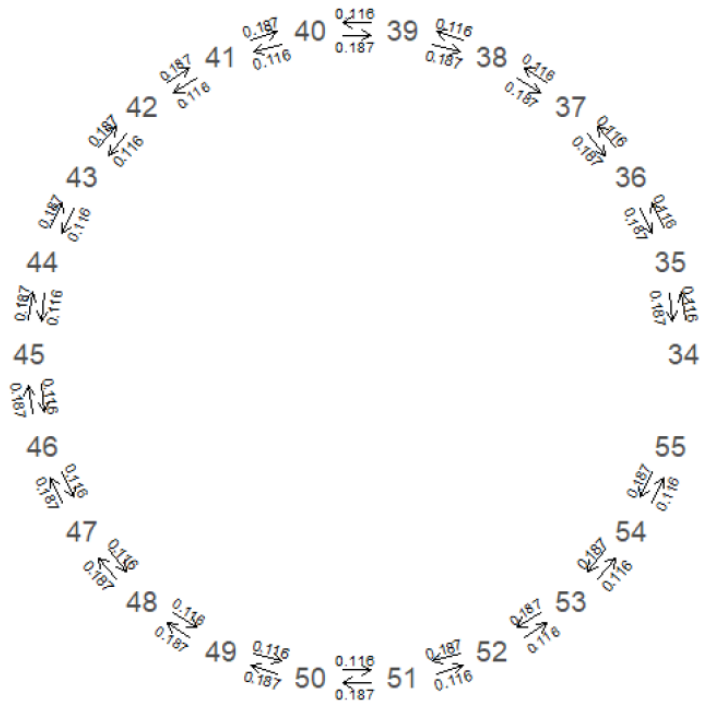
$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{Q}\boldsymbol{t})$$


2
1
3
0
4
5
0.16
1.03
1.03
0.16
0.16
1.03
0.16
1.03
0.16
1.03



Emma Sherratt (anoleannals.org)

a) Mainland caudal vertebrae number



b) Caribbean caudal vertebrae number

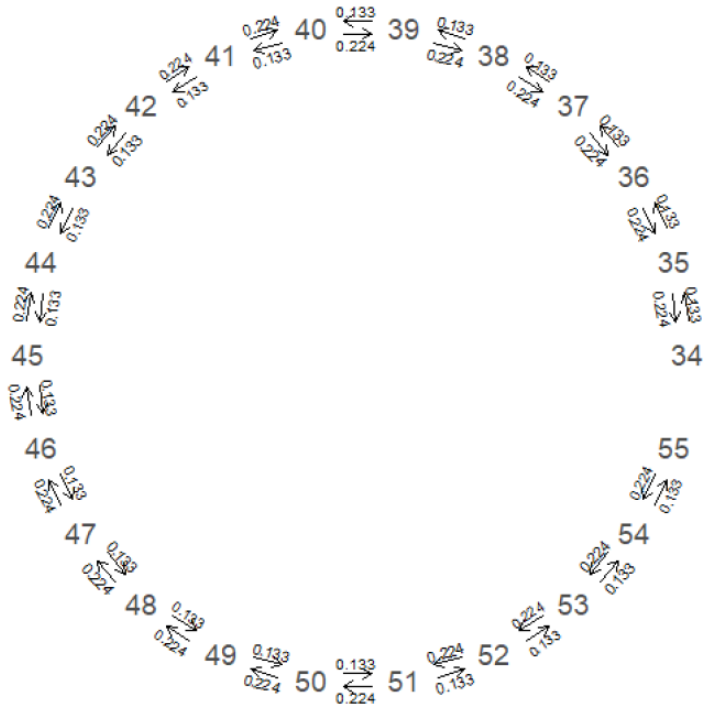


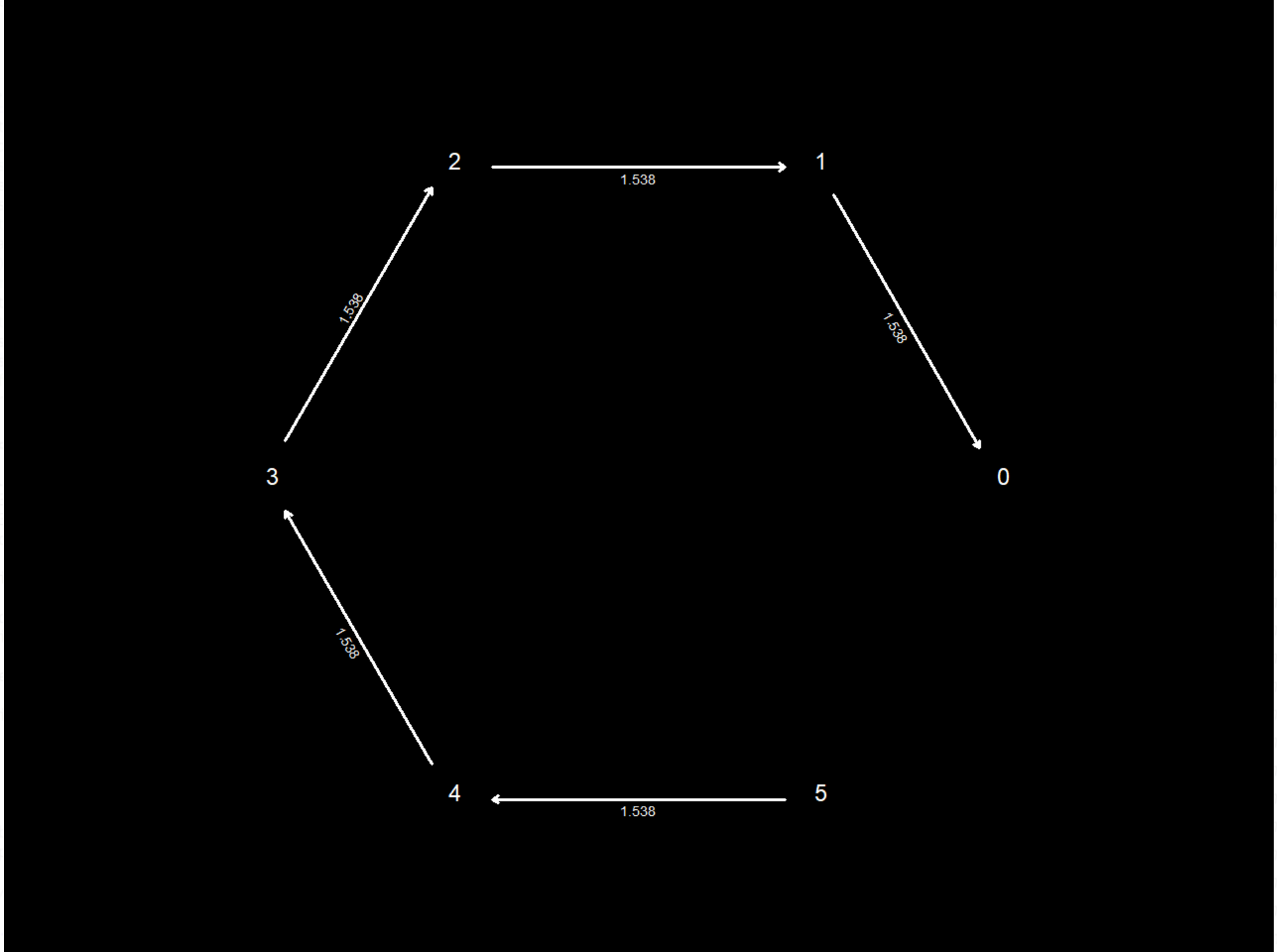
Figure 3. Fitted multi-rate symmetric ordered transition model for the evolution of caudal vertebrae number in mainland and Caribbean anole lineages.

$$\mathbf{Q} = \begin{bmatrix} -(q_{01} + q_{02}) & q_{01} & q_{02} \\ q_{10} & -(q_{10} + q_{12}) & q_{12} \\ q_{20} & q_{21} & -(q_{20} + q_{21}) \end{bmatrix}$$

$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{Q}\boldsymbol{t})$$

$$\mathbf{Q} = \begin{bmatrix} -(q_{01} + q_{02}) & 0 & 0 \\ q_{10} & -(q_{10} + q_{12}) & 0 \\ 0 & q_{21} & -(q_{20} + q_{21}) \end{bmatrix}$$

$$\mathbf{P(t)} = \exp(\mathbf{Q}t)$$


2
1
1.538
1.538
1.538
3
0
1.538
4
5
1.538

Discrete character evolution

- The M*k* model is the most common model for discrete character evolution on the tree – but it is not the only one.
- (Much like all tractable models) the model contains some fairly unrealistic assumptions.
- Some alternatives exist. Possibly the most important of these is a model called the *threshold model*.

```
data(sunfish.tree)
data(sunfish.data)
## extract discrete character (feeding mode)
fmode<-setNames(sunfish.data$feeding.mode,
    rownames(sunfish.data))
## do stochastic mapping
smap.trees<-make.simmap(sunfish.tree,fmode,model="ER",
    nsim=100)
## print a summary of the stochastic mapping
summary(smap.trees)
## plot a posterior probabilities of ancestral states
cols<-setNames(c("blue","red"),levels(fmode))
plot(summary(smap.trees),colors=cols,ftype="i")
legend("topleft",c("non-piscivorous","piscivorous"),
```

```
                x3 <- c(min(x2), x2, max(x2))
                y3 <- c(0, y2, 0)
                polygon(x3, y3, col = colors[x$trans[ii]],
                  border = FALSE)
                lines(p[[ii]]$mids - bw/2, p[[ii]]$density,
                  type = "s")
                dd <- 0.03 * diff(par()$usr[3:4])
                lines(hpd[[ii]], rep(max(p[[ii]]$density) +
                  dd, 2))
                text(mean(hpd[[ii]]), max(p[[ii]]$density) +
                  dd, "HPD", pos = 3)
                ii <- ii + 1
              }
            }
        }
    }
}
<bytecode: 0x000002b03146c8f0>
<environment: namespace:phytools>
> getwd()
[1] "C:/Users/liamj/Dropbox/PC/Documents"
>
> plot(summary(smap.trees),colors=cols,ftype="i")
>
```