

2 **Supplement A**

3 For all fits, we acquire 200,000 iterations on a single computing node,
 4 discard the first 100,000 draws as burnin, and thin by a factor of 5 to yield
 5 a sample of size 20,000.

6 We employ diffuse priors for most parameters in both the simulation and
 7 application.

8 **1 Priors**

9 Prior distributions are selected to be conjugate where possible.

$$\sigma^2 \sim \text{IG}(a_\sigma, b_\sigma) \quad (1)$$

$$[\alpha] \propto (1 + \alpha)^{\alpha_\alpha - 1} (1 - \alpha)^{\beta_\alpha - 1} \mathbf{1}_{(-1, 1)}(\alpha) \quad (2)$$

This is the kernel for the density of a random variable $X = 2Y - 1$ where $Y \sim \text{Beta}(\alpha_\alpha, \beta_\alpha)$. It is the same shape in a sense as the Beta, but spread over the support $[-1, 1]$, so that we allow for the possibility of anti-alignment.

10 See Section 1.1 for details.

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2) \quad (3)$$

$$c \sim \text{IG}(a_c, b_c) \quad (4)$$

$$\phi \sim \text{Beta}(\alpha_\phi, \beta_\phi) \quad (5)$$

$$p_1 \sim \text{Beta}(\alpha_{p_1}, \beta_{p_1}) \quad (6)$$

1.1 Prior for the strength of the alignment effect

The relevant support for α is the real interval $(-1, 1)$. Negative values for α correspond to the case when connected individuals tend to move in parallel, but opposite, directions. We generally expect α to be positive, but to allow for the possibility of anti-alignment behavior we specify a prior on the full support. To specify a flexible family of prior distributions with support $(-1, 1)$, we shift and scale a Beta distribution to align with $(-1, 1)$. In fitting our model with an MCMC algorithm, α is sampled with a Metropolis-Hastings step, therefore we only need the kernel of the full-conditional

$$[\alpha|\alpha_\alpha, \beta_\alpha] \propto (1 + \alpha)^{\alpha_\alpha - 1} (1 - \alpha)^{\beta_\alpha - 1} I_{(-1, 1)}(\alpha). \quad (7)$$

2 Full conditionals

Define the following variables as:

$$n \equiv \text{number of individuals in study population} \quad (8)$$

$$T \equiv \text{number of discrete time steps in mean position process} \quad (9)$$

It is convenient to also define

$$\mathbf{K}(t) \equiv (\mathbf{W}_+^c(t) - \alpha \mathbf{W}(t)) \otimes \mathbf{I}_2 \quad (10)$$

so that

$$\mathbf{Q}(t) = \sigma^{-2} \mathbf{K}(t). \quad (11)$$

²⁴ **2.1 σ : variability around mean step process**

²⁵ Gibbs step.

$$\sigma^2 | \cdot \sim \text{IG}(a^*, b^*) \tag{12}$$

$$a^* \equiv a_\sigma + nT \tag{13}$$

$$b^* \equiv b_\sigma + \frac{1}{2} \sum_{t=2}^T (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1) - \beta \tilde{\boldsymbol{\mu}}(t-1))' \mathbf{K}(t) (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1) - \beta \tilde{\boldsymbol{\mu}}(t-1)) \tag{14}$$

26 2.2 W: network

27 Gibbs step.

$$\mathbb{P}(w_{ij}(1) = 1|\cdot) \propto [\boldsymbol{\mu}_i(2)|\dots, w_{ij}(1) = 1] [\boldsymbol{\mu}_j(2)|\dots, w_{ij}(1) = 1] \quad (15)$$

$$\times p_{1|1}^{w_{ij}(2)} p_{0|1}^{1-w_{ij}(2)} p_1 \quad (16)$$

$$\mathbb{P}(w_{ij}(1) = 0|\cdot) \propto [\boldsymbol{\mu}_i(2)|\dots, w_{ij}(1) = 0] [\boldsymbol{\mu}_j(2)|\dots, w_{ij}(1) = 0] \quad (17)$$

$$\times p_{1|0}^{w_{ij}(2)} p_{0|0}^{1-w_{ij}(2)} (1 - p_1) \quad (18)$$

$$(19)$$

$$\mathbb{P}(w_{ij}(t) = 1|\cdot) \propto [\boldsymbol{\mu}_i(t)|\dots, w_{ij}(t) = 1] [\boldsymbol{\mu}_j(t)|\dots, w_{ij}(t) = 1] \quad (20)$$

$$\times [\boldsymbol{\mu}_i(t+1)|\dots, w_{ij}(t) = 1] [\boldsymbol{\mu}_j(t+1)|\dots, w_{ij}(t) = 1] \quad (21)$$

$$\times p_{1|1}^{w_{ij}(t+1)} p_{0|1}^{1-w_{ij}(t+1)} p_{1|0}^{1-w_{ij}(t-1)} p_{1|1}^{w_{ij}(t-1)} \quad (22)$$

$$\mathbb{P}(w_{ij}(t) = 0|\cdot) \propto [\boldsymbol{\mu}_i(t)|\dots, w_{ij}(t) = 0] [\boldsymbol{\mu}_j(t)|\dots, w_{ij}(t) = 0] \quad (23)$$

$$\times [\boldsymbol{\mu}_i(t+1)|\dots, w_{ij}(t) = 0] [\boldsymbol{\mu}_j(t+1)|\dots, w_{ij}(t) = 0] \quad (24)$$

$$\times p_{1|0}^{w_{ij}(t+1)} p_{0|0}^{1-w_{ij}(t+1)} p_{0|1}^{w_{ij}(t-1)} p_{0|0}^{1-w_{ij}(t-1)} \quad (25)$$

$$(26)$$

$$\mathbb{P}(w_{ij}(T) = 1|\cdot) \propto [\boldsymbol{\mu}_i(T)|\dots, w_{ij}(T) = 1] [\boldsymbol{\mu}_j(T)|\dots, w_{ij}(T) = 1] \quad (27)$$

$$\times p_{1|0}^{1-w_{ij}(T-1)} p_{1|1}^{w_{ij}(T-1)} \quad (28)$$

$$\mathbb{P}(w_{ij}(T) = 0|\cdot) \propto [\boldsymbol{\mu}_i(T)|\dots, w_{ij}(T) = 0] [\boldsymbol{\mu}_j(T)|\dots, w_{ij}(T) = 0] \quad (29)$$

$$\times p_{0|1}^{w_{ij}(T-1)} p_{0|0}^{1-w_{ij}(T-1)} \quad (30)$$

2.3 α : alignment/anti-alignment

M-H step. Kernel of full conditional density is

$$[\alpha|\cdot] \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=2}^T w_{i+}(t) (\mathbf{h}_i(t) - \alpha \bar{\mathbf{h}}_i(t))' (\mathbf{h}_i(t) - \alpha \bar{\mathbf{h}}_i(t)) \right\} \quad (31)$$

$$\times (1 + \alpha)^{\alpha_\alpha - 1} (1 - \alpha)^{\beta_\alpha - 1} \mathbf{1}_{(-1,1)}(\alpha) \quad (32)$$

where

$$\mathbf{h}_i(t) \equiv \boldsymbol{\mu}_j(t) - \boldsymbol{\mu}_j(t-1) - \beta \tilde{\boldsymbol{\mu}}_j(t-1) \quad (33)$$

$$\bar{\mathbf{h}}_i(t) \equiv \sum_{j \neq i} \frac{w_{ij}(t)}{w_{i+}^c(t)} \mathbf{h}_j(t) \quad (34)$$

and $\bar{\mathbf{h}}_i(t) = 0$ if animal i is completely unconnected at time t . Proposal distribution is normal:

$$\mathcal{N}(\alpha_{\text{iter}-1}, \sigma_{\alpha\text{-tune}}^2) \quad (35)$$

2.4 β : attraction/repulsion

Gibbs step.

$$\beta|\cdot \sim \mathcal{N}(\mu_\beta^*, \sigma_\beta^{*2}) \quad (36)$$

$$\sigma_\beta^{*2} \equiv \left(\sum_{t=2}^T \tilde{\boldsymbol{\mu}}(t-1)' \mathbf{Q}(t) \tilde{\boldsymbol{\mu}}(t-1) + \frac{1}{\sigma_\beta^2} \right)^{-1} \quad (37)$$

$$\mu_\beta^* \equiv \sigma_\beta^{*2} \left(\sum_{t=2}^T \tilde{\boldsymbol{\mu}}(t-1)' \mathbf{Q}(t) (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1)) + \mu_\beta \right) \quad (38)$$

34 **2.5 ϕ : network stability**

35 M-H step. Kernel of full conditional density is

$$[\phi|\cdot] \propto (p_{0|0})^{g_{0|0}}(p_{1|0})^{g_{1|0}}(p_{0|1})^{g_{0|1}}(p_{1|1})^{g_{1|1}}(\phi)^{\alpha_\phi-1}(1-\phi)^{\beta_\phi-1} \quad (39)$$

$$g_{0|0} \equiv \sum_{i < j} \sum_{t=2}^T (1 - w_{ij}(t-1))(1 - w_{ij}(t)) \quad (40)$$

$$g_{1|0} \equiv \sum_{i < j} \sum_{t=2}^T (1 - w_{ij}(t-1))w_{ij}(t) \quad (41)$$

$$g_{0|1} \equiv \sum_{i < j} \sum_{t=2}^T w_{ij}(t-1)(1 - w_{ij}(t)) \quad (42)$$

$$g_{1|1} \equiv \sum_{i < j} \sum_{t=2}^T w_{ij}(t-1)w_{ij}(t) \quad (43)$$

36 with proposals coming from a Beta distribution centered on ϕ_{iter-1} :

$$\text{Beta}\left(\frac{\beta_{\phi\text{-tune}}\phi_{iter-1}}{1 - \phi_{iter-1}}, \beta_{\phi\text{-tune}}\right) \quad (44)$$

37 **2.6 p_1 : network density**

38 M-H step. Kernel of full conditional density is

$$[p_1|\cdot] \propto (p_1)^{\sum_{i < j} w_{ij}(1)}(1 - p_1)^{\sum_{i < j} (1 - w_{ij}(1))}(p_{0|0})^{g_{0|0}}(p_{1|0})^{g_{1|0}}(p_{0|1})^{g_{0|1}}(p_{1|1})^{g_{1|1}}(p_1)^{\alpha_{p_1}-1}(1 - p_1)^{\beta_{p_1}-1} \quad (45)$$

39 with proposals coming from a Beta distribution centered on $p_{1iter-1}$:

$$\text{Beta}\left(\frac{\beta_{p_1\text{-tune}}p_{1iter-1}}{1 - p_{1iter-1}}, \beta_{p_1\text{-tune}}\right) \quad (46)$$

40 **2.7 c : precision for unconnected animals**

41 Gibbs step.

$$c|\cdot \sim \text{IG}\left(a_c + \sum_{i=1}^n \sum_{t=2}^T \mathbf{1}_{\{w_{i+}(t)=0\}}, b_c + \frac{1}{2} \sum_{i=1}^n \sum_{t=2}^T \mathbf{1}_{\{w_{i+}(t)=0\}} \mathbf{h}_i(t)' \mathbf{h}_i(t)\right)$$

(47)

42 3 Hyper-/tuning-parameters

parameter	hyper parameters		tuning
α	$\alpha_\alpha = 1$	$\beta_\alpha = 1$	$\sigma_{\alpha-tune}^2 = 0.1^2$
β	$\mu_\beta = 0$	$\sigma_\beta^2 = 1000$	conjugate
p_1	$\alpha_{p_1} = 1$	$\beta_{p_1} = 1$	$\beta_{p_1-tune} = 24$
ϕ	$\alpha_\phi = 100$	$\beta_\phi = \frac{100}{9}$	$\beta_{\phi-tune} = 7$
c	$a_c = 3.5$	$b_c = 1.5$	conjugate
σ	$a_\sigma = 0.1$	$b_\sigma = 0.001$	conjugate

Table 1: Simulation

parameter	hyper parameters		tuning
α	$\alpha_\alpha = 1$	$\beta_\alpha = 1$	$\sigma_{\alpha-tune}^2 = 0.01^2$
β	$\mu_\beta = 0$	$\sigma_\beta^2 = 1000$	conjugate
p_1	$\alpha_{p_1} = 1$	$\beta_{p_1} = 1$	$\beta_{p_1-tune} = 10$
ϕ	$\alpha_\phi = 100$	$\beta_\phi = \frac{100}{9}$	$\beta_{\phi-tune} = 5$
c	$a_c = 3.5$	$b_c = 1.5$	conjugate
σ	$a_\sigma = 0.1$	$b_\sigma = 0.001$	conjugate

Table 2: Killer whales