B Computational Details

B.1 Network Simulation

The algorithm used to simulate the network (**ALGORITHM 1**) produces samples from the stationary equilibrium of the model.

1. The network simulation algorithm satisfies the detailed balance condition for the stationary distribution 5. Indeed for any given θ

$$\begin{aligned}
\Pr\left(g'|g,X,\theta\right)\pi\left(g,X,\theta\right) &= q_g\left(g'|g\right)\min\left\{1,\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{\exp\left[Q\left(g,X,\theta\right)\right]}\frac{q_g\left(g|g'\right)}{q_g\left(g'|g\right)}\right\}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)} \\
&= \min\left\{q_g\left(g'|g\right)\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)},\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)}q_g\left(g|g'\right)\right\} \\
&= q_g\left(g|g'\right)\min\left\{\frac{q_g\left(g'|g\right)}{q_g\left(g|g'\right)}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)},\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)}\right\} \\
&= q_g\left(g|g'\right)\min\left\{\frac{q_g\left(g'|g\right)}{q_g\left(g|g'\right)}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{\exp\left[Q\left(g',X,\theta\right)\right]},1\right\}\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)} \\
&= \Pr\left(g|g',X,\theta\right)\pi\left(g',X,\theta\right)
\end{aligned}$$

   This concludes the proof.

2. The algorithm generates a Markov Chain of network with finite state space. The chain is irreducible and aperiodic and therefore it is uniformly ergodic (see Theorem 4.9, page 52 in Levin et al. (2008)).

3. The bound to the convergence rate used in the text was derived by Diaconis and Stroock (1991), for reversible finite chains.

The algorithm has a very useful property that can be exploited in the posterior simulation to reduce the computational burden. Adapting the suggestion in Liang (2010), define $\mathcal{P}_{\theta'}^{(R)}(g'|g)$ as the transition probability of a Markov chain that generates g' with R Metropolis-Hastings updates of the network simulation algorithm, starting at the observed network g and using the proposed parameter θ'. Then,

$$\mathcal{P}_{\theta'}^{(R)}(g'|g) = \mathcal{P}_{\theta'}(g^1|g)\mathcal{P}_{\theta'}(g^2|g^1)\cdots\mathcal{P}_{\theta'}(g'|g^{R-1}), \tag{22}$$

where $\mathcal{P}_{\theta'}(g^j|g^i) = q_g(g^j|g^i)\alpha_{mh}(g^i,g^j)$ is the transition probability of the network simulation algorithm above. Since the Metropolis-Hastings algorithm satisfies the detailed balance condition, we can prove the following

LEMMA 1 *Simulate a network g' from the stationary distribution $\pi(\cdot, X, \theta')$ using a Metropolis-Hastings algorithm starting at the network g observed in the data. Then*

$$\frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} = \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]} \tag{23}$$

for all R, $g, g' \in \mathcal{G}$ and for any $\theta' \in \Theta$.

Proof. Let $\mathcal{P}_{\theta'}^{(R)}(g'|g)$ be defined as in (22). This is the transition probability of the chain that generates g' with R Metropolis-Hastings updates, starting at the observed network g and using the proposed parameter θ'. Notice that the Metropolis-Hastings algorithm satisfies the detailed balance for $\pi(g, X, \theta')$, therefore we have

$$\begin{aligned}
\mathcal{P}_{\theta'}^{(R)}(g|g')\pi(g', X, \theta') &= \mathcal{P}_{\theta'}(g_{R-1}|g')\mathcal{P}_{\theta'}(g_{R-2}|g_{R-1})\cdots\mathcal{P}_{\theta'}(g|g_1)\pi(g', X, \theta') \\
&= \mathcal{P}_{\theta'}(g_1|g)\mathcal{P}_{\theta'}(g_2|g_1)\cdots\mathcal{P}_{\theta'}(g'|g_{R-1})\pi(g, X, \theta') \\
&= \mathcal{P}_{\theta'}^{(R)}(g'|g)\pi(g, X, \theta')
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} &= \frac{\pi(g, X, \theta')}{\pi(g', X, \theta')} \\
&= \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\frac{c(\mathcal{G}, X, \theta')}{c(\mathcal{G}, X, \theta')} \\
&= \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}.
\end{aligned}$$

This concludes the proof. ■

One should notice that as long as the algorithm is started from the network g observed in the data (which is assumed to be a draw from the stationary equilibrium of the model), the equality in (23) is satisfied for any R.

The approximate exchange algorithm presented in this paper removes the requirement of exact sampling by exploiting the property of the stationary equilibrium characterization, described in Lemma 1.

Local simulations and Large steps

The theoretical results using graph limits and large deviations suggest that the local sampler has systematic convergence problems, even when in principle the simulation is trivial because links are asymptotically independent ($\beta > 0$).

To attenuate these convergence issue, we propose a modification that allows the sampler to make *larger steps*, in particular steps that are not $o(n)$. The local chain selects a link g_{ij} with probability $1/(n(n-1))$, proposing to swap the value to $1-g_{ij}$. We add the following large steps. First, with probability p_r, the sampler selects a player i at random (with probability $1/n$) and proposes to swap all his links, i.e. $g_{ij} = 1-g_{ij}$ for each $j = 1, ..., n$. Second, with probability p_c, the sampler selects a player i at random (with probability $1/n$) and proposes to swap all the links pointing at i, i.e. $g_{ji} = 1-g_{ji}$ for each $j = 1, ..., n$. Third, with probability p_f, the sampler selects uniformly at random $\lceil \lambda n \rceil$ links, where $\lambda \in (0,1)$, and proposes to swap all of them. Notice that this step size is a function of n, and in particular is not $o(n)$. The crucial ingredient is to make the length of the step a function of n. The parameter λ is under control of the researcher: higher values allow larger steps and increase the computational cost of sampling. Lastly, with probability p_{inv} the sampler proposes to invert the adjacency matrix. The goal of this large step is to provide a way to jump across modes of the stationary distribution, when it is bimodal. [46]

Using this sampler, we reproduce the simulation in Figure 1. We know that the local chain can get trapped in local maxima of the variational problem. If we simulate model (10) with parameters $(\alpha, \beta) = (-3, 3)$, we obtain Figure 8(A). While Theorem 2 states that the simulations should converge to the sparse network density $\mu_1 \approx 0.07$, we observe that the local sampler converges to a dense network with $\mu_2 \approx 0.93$, if started at dense networks. In other words, when started at a dense network (say the full network), the sampler gets trapped in a local maximum of the variational problem, with density $\mu_2 \approx 0.93$. Figure 8(B) shows that our modified sampler does not have this problem, and also the chains started at dense network converge to the correct (sparse) network density. This simple modification gets rid of the exponentially slow convergence of the local algorithm. More generally, these larger steps allow the sampler to escape local maxima of the potential function.

In general this modification should help the sampler when the likelihood has multiple modes. However, the improvement comes with the increase cost of sampling and additional computational time. In some models the cost may be substantial. For example, it is intuitive that in regions of the parameters space (say, $\beta < 0$) where the likelihood is unimodal, the gains from this modified sampler are minimal.

B.2 Posterior Simulation

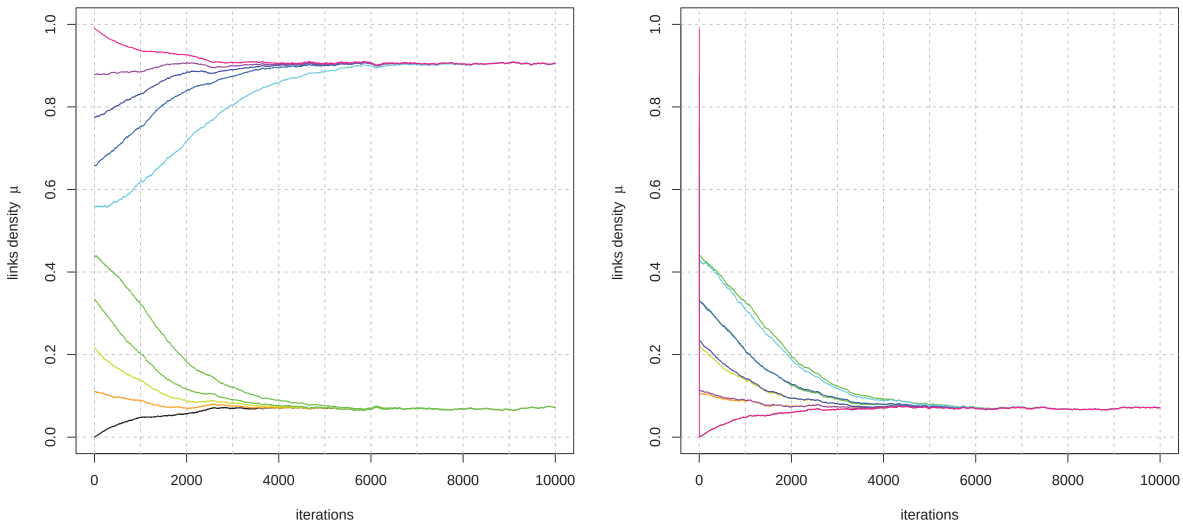
In this section I provide the technical details for the algorithm proposed in the empirical part of the paper. The first set of results show that the exchange algorithm generate (approximate) samples from the posterior distribution (7).

The original exchange algorithm developed in Murray et al. (2006) is slightly different from the one used here. The main modification is in Step 2: the original algorithm requires an *exact* sample from the stationary equilibrium of the model.

ALGORITHM 3 ***(EXACT EXCHANGE ALGORITHM)***

[46] We have seen that this is the case in the homogeneous player case, for many parameter values.

Figure 8: Local sampler versus Modified sampler



(A) (B)

Comparison of network samplers for model (10), with parameters $(\alpha, \beta) = (-3, 3)$. Panel (A) shows the simulation using the local-chain sampler, which converges to two different link densities ($\mu_1 \approx 0.07$ and $\mu_2 \approx 0.93$). However, we know from Theorem 2, that the correct simulation should converge to the sparse network density. So the local chain fails to sample correctly if we start it at a dense network, because it gets trapped at a local maximum of the stationary distribution. Panel (B) shows the simulation using the modified algorithm. We use $p_r = p_f = p_{inv} = 0.01$. The simulations converge to the correct link density for any starting value, therefore our modified algorithm provides a better sampler for the model.

Start at current parameter $\theta_t = \theta$ and network data g.

1. *Propose a new parameter vector θ'*

$$\theta' \sim q_\theta(\cdot|\theta) \tag{24}$$

2. *Draw an exact sample network g' from the likelihood*

$$g' \sim \pi\left(\cdot|X, \theta'\right) \tag{25}$$

3. *Compute the acceptance ratio*

$$\begin{aligned} \alpha_{ex}\left(\theta, \theta', g', g\right) &= \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]} \frac{c(\theta)c(\theta')}{c(\theta)c(\theta')}\right\} \\ &= \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\right\} \end{aligned} \tag{26}$$

4. Update the parameter according to

$$\theta_{t+1} = \begin{cases} \theta' & \textit{with prob. } \alpha_{ex}(\theta, \theta', g', g) \\ \theta & \textit{with prob. } 1 - \alpha_{ex}(\theta, \theta', g', g) \end{cases} \tag{27}$$

The difference between this algorithm and the approximate one is in step 2. The exact and approximate algorithms use the same acceptance ratio $\alpha_{ex}(\theta, \theta', g', g)$, a consequence of LEMMA 1. Indeed the acceptance ratio for the approximate algorithm is

$$\begin{aligned}
\widetilde{\alpha}_{ex}(\theta, \theta', g', g) &= \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\mathcal{P}^{(R)}_{\theta'}(g|g')}{\mathcal{P}^{(R)}_{\theta'}(g'|g)}\right\} && (28) \\
&= \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g', X, \theta')]}\right\} && (29) \\
&= \alpha_{ex}(\theta, \theta', g', g) && (30)
\end{aligned}$$

This result implies that to prove the convergence of the approximate algorithm to the exact algorithm, there is no need to prove convergence of $\widetilde{\alpha}_{ex}(\theta, \theta', g', g)$ to $\alpha_{ex}(\theta, \theta', g', g)$. The convergence of step 2 of the algorithm is sufficient.

B.2.1 Preliminary Lemmas for THEOREM 6

The convergence of the approximate exchange algorithm to the correct posterior distribution is proven in 4 steps.

1. First we prove that the exact exchange algorithm converges to the correct posterior (LEMMA 2)
2. Second, we prove that the approximate algorithm has a stationary distribution and it is ergodic (LEMMA 3, similar to the one in Liang 2010)
3. Third, we prove that the transition kernel of the approximate and exact algorithms are arbitrarily close for a large enough number of network simulations (LEMMA 4)
4. Fourth, we combine previous results to prove that the approximate algorithm converges to the correct posterior

A similar proof strategy is contained in Liang et al. (2010) and Andrieu and Roberts (2009).

Let $Q(d\vartheta|\theta) = q_\theta(\vartheta|\theta)\,\nu(d\vartheta)$. The transition kernel of the exact exchange algorithm can be written as

$$
\begin{aligned}
P(\theta, d\vartheta) &= \left[\sum_{g'\in\mathcal{G}} \pi(g', \vartheta)\,\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta) \\
&+ \delta_\theta(d\vartheta)\left\{1 - \int_\Theta \left[\sum_{g'\in\mathcal{G}} \pi(g', \vartheta)\,\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\right\}
\end{aligned}
$$

and the transition kernel of the approximate exchange algorithm can be written as

$$
\begin{aligned}
\widetilde{P}_R(\theta, d\vartheta) &= \left[\sum_{g'\in\mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\,\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta) \\
&+ \delta_\theta(d\vartheta)\left\{1 - \int_\Theta \left[\sum_{g'\in\mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\,\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\right\}
\end{aligned}
$$

Let $\eta(\theta)$ be the average rejection probability for the approximate algorithm, i.e.

$$
\eta(\theta) := 1 - \int_\Theta \left[\sum_{g'\in\mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\,\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta) \tag{31}
$$

The next lemma proves that the transition kernel satisfies the detailed balance condition for the posterior distribution. For any pair of parameters $(\theta, \vartheta) \in \Theta$ we have

$$
P[\theta, \vartheta|g, X]\, p(\theta|g, X) = \Pr[\theta|\vartheta, g, X]\, p(\vartheta|g, X) \tag{32}
$$

The detailed balance condition is sufficient condition for the Markov chain generated by the algorithm to have stationary distribution the posterior (7) (for details see Robert and Casella (2005) or Gelman et al. (2003)).

LEMMA 2 *The exchange algorithm produces a Markov chain with invariant distribution (7).*

Proof. Define $\mathcal{Z} \equiv \int_\Theta \pi(g|X, \theta)\, p(\theta)\, d\theta$. In the algorithm the probability $\Pr[\vartheta|\theta, g, X]$ of transition to θ_j, given the current parameter θ and the observed data (g, X), can be computed as

$$
\Pr[\vartheta|\theta, g, X] = q_\theta(\vartheta|\theta)\, \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \alpha_{ex}(\theta, \vartheta, g', g). \tag{33}
$$

This is the probability $q_\theta(\vartheta|\theta)$ of proposing ϑ times the probability of generating the new network g' from the model's stationary distribution, $\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}$ and accepting the proposed parameter $\alpha_{ex}(\theta, \vartheta, g', g)$. Therefore the left-hand side of (32) can be written as

$$
\begin{aligned}
\Pr[\vartheta|\theta, g, X]\, p(\theta|g, X) &= q_\theta(\vartheta|\theta)\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\alpha_{ex}(\theta, \vartheta, g', g)\, p(\theta|g, X) \\
&= q_\theta(\vartheta|\theta)\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\alpha_{ex}(\theta, \vartheta, g', g)\frac{\frac{\exp[Q(g,X,\theta)]}{c(\mathcal{G},X,\theta)}p(\theta)}{\mathcal{Z}} \\
&= q_\theta(\vartheta|\theta)\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \\
&\times \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]}\frac{p(\vartheta)}{p(\theta)}\frac{q_\theta(\theta|\vartheta)}{q_\theta(\vartheta|\theta)}\frac{\exp[Q(g, X, \vartheta)]}{\exp[Q(g', X, \vartheta)]}\right\} \\
&\times \frac{\frac{\exp[Q(g,X,\theta)]}{c(\mathcal{G},X,\theta)}p(\theta)}{\mathcal{Z}}
\end{aligned}
$$

$$
= \min\left\{q_\theta(\vartheta|\theta)\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\frac{\exp[Q(g, X, \theta)]}{c(\mathcal{G}, X, \theta)}\frac{p(\theta)}{\mathcal{Z}}, q_\theta(\theta|\vartheta)\frac{\exp[Q(g', X, \theta]}{c(\mathcal{G}, X, \theta)}\frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\frac{p(\vartheta)}{\mathcal{Z}}\right\}
$$

$$
\begin{aligned}
&= q_\theta(\theta|\vartheta)\frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)}\frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\frac{p(\vartheta)}{\mathcal{Z}} \times \\
&\times \min\left\{1, \frac{\exp[Q(g', X, \vartheta)]}{\exp[Q(g, X, \vartheta)]}\frac{p(\theta)}{p(\vartheta)}\frac{q_\theta(\vartheta|\theta)}{q_\theta(\theta|\vartheta)}\frac{\exp[Q(g, X, \theta)]}{\exp[Q(g', X, \theta)]}\right\} \\
&= q_\theta(\theta|\vartheta)\frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)}\alpha(\vartheta, \theta, g', g)\frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}\frac{p(\vartheta)}{\mathcal{Z}} \\
&= q_\theta(\theta|\vartheta)\frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)}\alpha(\vartheta, \theta, g', g)p(\vartheta|g, X) \\
&= \Pr[\theta|\vartheta, g, X]\, p(\vartheta|g, X)
\end{aligned}
$$

The latter step proves the detailed balance for a generic network g'. Since the condition is satisfied for any network g', detailed balance follows from summing over all possible networks.
■

LEMMA 3 *(The approximate algorithm is ergodic)*
Assume the exact exchange algorithm is ergodic and that for any $\vartheta \in \Theta$

$$\frac{\mathcal{P}_{\vartheta}^{(R)}(g'|g)}{\pi(g', \vartheta)} > 0 \quad \textit{for any } g' \in \mathcal{G} \tag{34}$$

Then for any $R \in \mathbb{N}$ such that for any $\theta \in \Theta$, $\rho(\theta) > 0$, the transition kernel of the approximate algorithm $\widetilde{P}_R$ is also irreducible and aperiodic, and there exists a stationary distribution $\widetilde{p}(\theta)$ such that

$$\lim_{s\to\infty}\left\|\widetilde{P}_R^{(s)}(\theta_0, \cdot) - \widetilde{p}(\theta)\right\|_{TV} = 0 \tag{35}$$

Proof. The exact algorithm with transition kernel P is an irreducible and aperiodic Markov chain. To prove that the approximate algorithm with transition kernel $\widetilde{P}_R$ defines an ergodic Markov chain, it is sufficient to prove that the set of accessible states of P are also included in those of $\widetilde{P}_R$. The proof proceeds by induction.
Formally, we need to show that for any $s \in \mathbb{N}$, $\theta \in \Theta$ and $A \in \mathcal{B}(\Theta)$ such that $P^{(s)}(\theta, A) > 0$, implies $\widetilde{P}_R^{(s)}(\theta, A) > 0$.
Notice that for any $\theta \in \Theta$ and $A \in \mathcal{B}(\Theta)$,

$$\begin{aligned}
\widetilde{P}_R(\theta, A) &= \int_A \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\, \alpha_{ex}(\theta, \vartheta, g', g) \right] q_\theta(\vartheta|\theta)\, d\vartheta + \mathbb{I}(\theta \in A)\, \eta(\theta) \\
&\geq \int_A \left[\sum_{g' \in \mathcal{G}} \min\left\{ 1, \frac{\mathcal{P}_\vartheta^{(R)}(g'|g)}{\pi(g', \vartheta)} \right\} \pi(g', \vartheta)\, \alpha_{ex}(\theta, \vartheta, g', g) \right] q_\theta(\vartheta|\theta)\, d\vartheta + \mathbb{I}(\theta \in A)\, \eta(\theta) > 0
\end{aligned}$$

where the last inequality comes from $\frac{\mathcal{P}_\vartheta^{(R)}(g'|g)}{\pi(g',\vartheta)} > 0$ for any $g' \in \mathcal{G}$ and $\vartheta \in \Theta$.
This proves that the statement is true when $s = 1$. By induction we assume that it is true up to $s = n \geq 1$ and for some $\theta \in \Theta$ chose $A \in \mathcal{B}(\Theta)$ such that $P^{(n+1)}(\theta, A) > 0$ and assume that

$$\int_\Theta \widetilde{P}_R^{(n)}(\theta, d\vartheta)\, \widetilde{P}_R(\vartheta, A) = 0$$

This implies that $\widetilde{P}_R(\vartheta, A) = 0$, $\widetilde{P}_R^{(n)}(\theta, \cdot)$-a.s.; by the induction assumption at $s = 1$ it follows that $P(\vartheta, A) = 0$, $\widetilde{P}_R^{(n)}(\theta, \cdot)$-a.s.
From this and the induction assumption at $s = n$, $P(\vartheta, A) = 0$, $P^{(n)}(\theta, \cdot)$-a.s. (assume not, then $P(\vartheta, A) > 0$, $P^{(n)}(\theta, \cdot)$-a.s. which by induction would imply $\widetilde{P}_R(\vartheta, A) > 0$, which is a contradiction). The latter step contradicts $P^{(n+1)}(\theta, A) > 0$ and the result follows. ■

The next step consists of proving that the transition kernel of the approximate algorithm $\widetilde{P}_R(\theta, \vartheta)$ and the exact algorithm $P(\theta, \vartheta)$ are arbitrarily close for a large enough number of network simulations R. Formally we prove a statement which is equivalent to proving convergence in total variation norm.[47]

LEMMA 4 *(Convergence of the exact and approximate transition kernels)*
Let $\epsilon \in (0, 1]$. There exists a number of simulations $R_0 \in \mathbb{N}$ such that for any function $\phi : \Theta \to [-1, 1]$ and any $R > R_0$,

$$\left| \widetilde{P}_R \phi(\theta) - P\phi(\theta) \right| < 2\epsilon \tag{36}$$

[47] See Levin et al. (2008), proposition 4.5, page 49.

Proof. The transition of the exchange algorithm is

$$\begin{aligned} P\left(\phi(\theta),\phi(\vartheta)\right) &= \int_\Theta \phi(\vartheta)\left[\sum_{g'\in\mathcal{G}} \pi\left(g',\vartheta\right)\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \\ &+ \phi(\theta)\left[1-\int_\Theta\left[\sum_{g'\in\mathcal{G}} \pi\left(g',\vartheta\right)\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta\right] \end{aligned}$$

while the transition kernel for the approximate algorithm is

$$\begin{aligned} \widetilde{P}_R\left(\phi(\theta),\phi(\vartheta)\right) &= \int_\Theta \phi(\vartheta)\left[\sum_{g'\in\mathcal{G}} \mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \\ &+ \phi(\theta)\left[1-\int_\Theta\left[\sum_{g'\in\mathcal{G}} \mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta\right] \end{aligned}$$

and therefore the difference is

$$\begin{aligned} S &= P\left(\phi(\theta),\phi(\vartheta)\right)-\widetilde{P}_R\left(\phi(\theta),\phi(\vartheta)\right) \\ &= \int_\Theta \phi(\vartheta)\left[\sum_{g'\in\mathcal{G}}\left[\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right]\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \\ &- \phi(\theta)\int_\Theta\left[\sum_{g'\in\mathcal{G}}\left[\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right]\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \end{aligned}$$

Consider the quantity

$$\begin{aligned} S_0 &= \int_\Theta\left[\sum_{g'\in\mathcal{G}}\left[\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right]\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \\ &\leq \int_\Theta\left[\sum_{g'\in\mathcal{G}}\left|\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right|\alpha_{ex}\left(\theta,\vartheta,g',g\right)\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \end{aligned}$$

and since $\alpha_{ex}\left(\theta,\vartheta,g',g\right)\leq 1$ for any $(\theta,\vartheta)\in\Theta\times\Theta$ and $(g',g)\in\mathcal{G}\times\mathcal{G}$, we have

$$\begin{aligned} S_0 &\leq \int_\Theta\left[\sum_{g'\in\mathcal{G}}\left|\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right|\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \\ &= \int_\Theta\left[2\sup_{g'\in\mathcal{G}}\left|\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right|\right] q_\theta\left(\vartheta|\theta\right)d\vartheta \end{aligned}$$

The convergence of the network simulation algorithm implies that for any $\varepsilon>0$, there exists an $R_0\left(\vartheta,\varepsilon\right)\in\mathbb{N}$ such that for any $R>R_0\left(\vartheta,\varepsilon\right)$ and for any $g\in\mathcal{G}$

$$2\sup_{g'\in\mathcal{G}}\left|\pi\left(g',\vartheta\right)-\mathcal{P}_\vartheta^{(R)}\left(g'|g\right)\right|\leq\varepsilon$$

Pick $R_0(\varepsilon) = \max_{\vartheta\in\Theta}\{R_0(\vartheta,\varepsilon)\}$. Then for any $\epsilon \in (0,1]$, there is an $R_0(\epsilon) \in \mathbb{N}$ such that for any $R > R_0(\epsilon)$ and for any $g \in \mathcal{G}$

$$S_0 \quad \leq \quad \int_\Theta \epsilon q_\theta(\vartheta|\theta)\, d\vartheta = \epsilon$$

This implies that

$$|S| \leq |2S_0| = 2\epsilon \tag{37}$$

■

The next theorem is the main result for the convergence. It states that the approximate exchange algorithm converges to the correct posterior distribution, provided that the number of network simulations and parameter samples are big enough.

B.2.2 Proof of THEOREM 6

. **Proof.** The main idea is to decompose the total variation in two components

$$\begin{aligned}\left\|\widetilde{P}_R^{(s)}(\theta_0,\cdot) - p(\cdot|g,X)\right\|_{TV} &= \left\|\widetilde{P}_R^{(s)}(\theta_0,\cdot) - P^{(s)}(\theta_0,\cdot) + P^{(s)}(\theta_0,\cdot) - p(\cdot|g,X)\right\|_{TV} \\ &\leq \left\|\widetilde{P}_R^{(s)}(\theta_0,\cdot) - P^{(s)}(\theta_0,\cdot)\right\|_{TV} + \left\|P^{(s)}(\theta_0,\cdot) - p(\cdot|g,X)\right\|_{TV}\end{aligned}$$

and prove that each component converges. We will use the same idea, but rewrite the total variation in a more convenient form.[48] For any function $\phi : \Theta \to [-1,1]$ we have

$$\begin{aligned}\left|\widetilde{P}_R^{(s)}\phi(\theta_0) - p(\phi)\right| &= \left|\widetilde{P}_R^{(s)}\phi(\theta_0) - P^{(s)}\phi(\theta_0) + P^{(s)}\phi(\theta_0) - p(\phi)\right| \\ &\leq \left|\widetilde{P}_R^{(s)}\phi(\theta_0) - P^{(s)}\phi(\theta_0)\right| + \left|P^{(s)}\phi(\theta_0) - p(\phi)\right|\end{aligned}$$

The second component converges because the exact exchange algorithm is ergodic, as stated in Lemma. For any $\varepsilon > 0$ there is number of simulation steps $s(\theta_0,\varepsilon)$, such that for any $s \geq s(\theta_0,\varepsilon)$

$$\left|P^{(s)}\phi(\theta_0) - p(\phi)\right| \leq \varepsilon \tag{38}$$

For the remaining of the proof, I will set $s_0 := s(\theta_0,\varepsilon)$. I use the telescoping sum decomposition in Andrieu and Roberts (2009) (page 15, adapted from last formula)

$$\begin{aligned}\left|\widetilde{P}_R^{(s_0)}\phi(\theta_0) - P^{(s_0)}\phi(\theta_0)\right| &= \left|\sum_{l=0}^{s_0-1}\left[P^{(l)}\widetilde{P}_R^{(s_0-l)}\phi(\theta_0) - P^{(l+1)}\widetilde{P}_R^{(s_0-(l+1))}\phi(\theta_0)\right]\right| \\ &= \left|\sum_{l=0}^{s_0-1}P^{(l)}\left(\widetilde{P}_R - P\right)\widetilde{P}_R^{(s_0-(l+1))}\phi(\theta_0)\right|\end{aligned}$$

[48] See Levin et al. (2008), proposition 4.5, page 49.

Now we can apply s_0 times the result of LEMMA 4 (as in Liang et al. (2010) and Andrieu and Roberts (2009)) to prove that there exists an $R_0(\theta_0, \varepsilon) \in \mathbb{N}$ such that for any $R > R_0(\theta_0, \varepsilon)$

$$\left| \widetilde{P}_R^{(s_0)} \phi(\theta_0) - P^{(s_0)} \phi(\theta_0) \right| \leq 2 s_0 \varepsilon \tag{39}$$

this implies

$$\left| \widetilde{P}_R^{(s)} \phi(\theta_0) - p(\phi) \right| \leq (2 s_0 + 1) \varepsilon \tag{40}$$

We conclude the proof by choosing $\varepsilon = \epsilon / (2 s_0 + 1)$.

This proves that the approximate exchange algorithm is ergodic, therefore the law of large number holds, and the second part of the theorem is proven. ■