Supplement to "A penalized likelihood approach to estimate within-household contact networks from egocentric data"

Gail E. Potter and Niel Hens

In this supplement we present cross-validation plots in section 1, penalized MLE estimates for additional household compositions in section 2, and penalized MLE results and simulation study results using the adjacency penalty in section 3.

1 Cross-Validation Plots

Figure 1: Mean of maximized likelihood from cross-validation procedure for a range of λ-values for households of various age compositions. The vertical lines show the CV-selected λ-values.
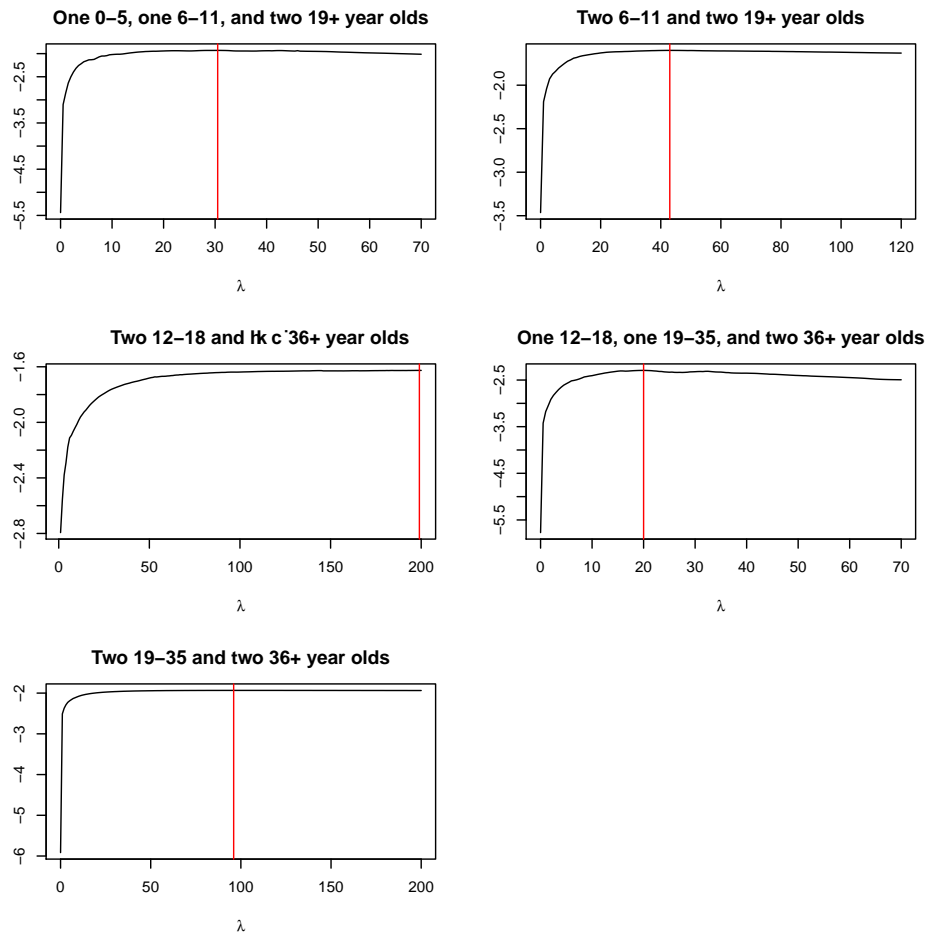
2 Network estimates for additional household compositions

In this section we report estimates of the household network probability distribution by maximum likelihood, penalized likelihood, and the independence model for six different household age compositions. For ease of comparison between different age compositions, we include estimates reported in the main text. Tables 1–6 show the estimated probability distribution by the three different models. Among the 64 possible networks, we show only those whose probability estimates were at least 0.03 in at least one of the three models. We graphically depict the most likely networks in Figures 2–7. The figures show a decrease in network density as the age of household members increases. The probability of the complete network decreases substantially as the age of the two youngest members increases. Its estimate is 0.65 in households with two 0–5 year olds and two adults. In the next three age compositions, where at least one of the children is school-aged and both are still under 18, its estimates ranges from 0.24–0.34. In households with one adult (19+) child, the probability of the complete network is estimated to be 0.17, and its estimate is 0.01 for households with two adult children. We also see shifts in the second most likely network. This network includes all contacts except between the two adults when both children are under twelve. When the two children are teenagers, the second most likely network includes all contacts except between one of the teenagers and one of the parents. In households with at least one adult child, sparser networks are much more common. When the youngest member is a teenager and the second youngest is over 18, the most likely network is the empty network, and the second most likely is the complete network. When the two youngest members are both adults, the most likely networks include very few contacts. This low frequency of contact fits our expectations of social behavior in all-adult households, where members are more independent and the youngest members do not require caretaking.

Table 1: Estimated probability distribution of contact network for households with two 0–5 year olds and two 19+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 23.5$), and unpenalized likelihood estimates are shown.

Contact network						Estimate			95% C.I.		
c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
0	1	0	0	0	1	0.04	0	0	[0, 0.10]	[0, 0]	[0, 0]
0	1	1	0	0	1	0.04	0.06	0	[0, 0.10]	[0, 0.15]	[0, 0.02]
0	1	1	1	1	1	0	0.01	0.05	[0, 0]	[0, 0.01]	[0, 0.10]
1	1	1	0	0	1	0	0	0.02	[0, 0]	[0, 0.02]	[0, 0.07]
1	1	1	0	1	0	0	0.01	0.02	[0, 0]	[0, 0.07]	[0, 0.08]
1	1	1	0	1	1	0.07	0.08	0.13	[0, 0.21]	[0.01, 0.21]	[0.03, 0.23]
1	1	1	1	0	0	0	0.01	0.02	[0, 0]	[0, 0.03]	[0, 0.06]
1	1	1	1	0	1	0.05	0.06	0.1	[0, 0.17]	[0, 0.17]	[0, 0.18]
1	1	1	1	1	0	0.14	0.12	0.09	[0, 0.48]	[0, 0.42]	[0, 0.33]
1	1	1	1	1	1	0.65	0.65	0.54	[0.30, 0.89]	[0.35, 0.88]	[0.26, 0.83]

Table 2: Estimated probability distribution of contact network for households with one 0–5 year old, one 6–11 year old, and two 19+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 30.5$), and unpenalized likelihood estimates are shown.

Contact network						Estimate			95% C.I.		
c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
0	1	1	1	0	1	0.05	0.01	0.01	[0, 0.12]	[0, 0.05]	[0, 0.03]
0	1	1	1	1	1	0.05	0.03	0.02	[0, 0.12]	[0, 0.10]	[0, 0.07]
1	0	0	0	0	1	0.09	0.05	0	[0, 0.28]	[0, 0.11]	[0, 0.01]
1	0	0	0	1	1	0.03	0.01	0	[0, 0.13]	[0, 0.05]	[0, 0.03]
1	0	0	1	1	0	0.07	0.03	0.01	[0, 0.27]	[0, 0.13]	[0, 0.04]
1	0	1	0	0	1	0.09	0.04	0.01	[0, 0.30]	[0, 0.17]	[0, 0.05]
1	0	1	0	1	1	0.03	0.02	0.03	[0, 0.16]	[0, 0.10]	[0, 0.08]
1	0	1	1	0	1	0	0.01	0.03	[0, 0]	[0, 0.04]	[0, 0.08]
1	0	1	1	1	0	0.07	0.07	0.05	[0, 0.28]	[0, 0.21]	[0, 0.16]
1	0	1	1	1	1	0	0.04	0.07	[0, 0]	[0, 0.13]	[0, 0.18]
1	1	1	0	0	1	0	0.03	0.03	[0, 0]	[0, 0.16]	[0, 0.12]
1	1	1	0	1	0	0	0.02	0.04	[0, 0]	[0, 0.07]	[0, 0.11]
1	1	1	0	1	1	0	0.03	0.06	[0, 0]	[0, 0.09]	[0, 0.16]
1	1	1	1	0	0	0	0.03	0.05	[0, 0]	[0, 0.10]	[0, 0.14]
1	1	1	1	0	1	0.07	0.05	0.07	[0, 0.24]	[0, 0.16]	[0, 0.21]
1	1	1	1	1	0	0.20	0.14	0.11	[0, 0.55]	[0, 0.35]	[0, 0.29]
1	1	1	1	1	1	0.25	0.24	0.16	[0, 0.63]	[0.03, 0.50]	[0.03, 0.40]

Table 3: Estimated probability distribution of contact network for households with two 6–11 year olds and two 19+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 43$), and unpenalized likelihood estimates are shown.

Contact network						Estimate			95% C.I.		
c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
0	0	0	0	1	0	0.03	0.02	0	[0, 0.20]	[0, 0.07]	[0, 0.01]
0	0	0	1	1	0	0.04	0.01	0	[0, 0.12]	[0, 0.05]	[0, 0.02]
0	0	0	1	1	1	0.04	0.03	0	[0, 0.12]	[0, 0.11]	[0, 0.04]
0	1	1	1	1	1	0	0.02	0.03	[0, 0]	[0, 0.07]	[0, 0.08]
1	0	0	0	1	0	0.03	0.01	0	[0, 0.12]	[0, 0.04]	[0, 0.03]
1	0	0	1	1	1	0	0.01	0.03	[0, 0]	[0, 0.06]	[0, 0.11]
1	0	1	0	1	0	0.03	0.03	0.01	[0, 0.12]	[0, 0.11]	[0, 0.06]
1	0	1	0	1	1	0.12	0.05	0.04	[0, 0.3]	[0, 0.14]	[0, 0.1]
1	0	1	1	1	0	0	0.02	0.04	[0, 0]	[0, 0.05]	[0, 0.11]
1	0	1	1	1	1	0	0.05	0.1	[0, 0]	[0, 0.12]	[0.01, 0.18]
1	1	0	1	1	1	0	0.03	0.06	[0, 0]	[0, 0.09]	[0, 0.13]
1	1	1	0	1	0	0	0.02	0.03	[0, 0]	[0, 0.09]	[0, 0.11]
1	1	1	0	1	1	0	0.07	0.07	[0, 0]	[0.01, 0.20]	[0.01, 0.19]
1	1	1	1	0	0	0.14	0.03	0.01	[0, 0.31]	[0, 0.10]	[0, 0.06]
1	1	1	1	0	1	0	0.05	0.03	[0, 0]	[0, 0.16]	[0, 0.11]
1	1	1	1	1	0	0	0.1	0.08	[0, 0]	[0, 0.35]	[0, 0.28]
1	1	1	1	1	1	0.56	0.32	0.2	[0.35, 0.75]	[0.11, 0.55]	[0.04, 0.5]

Table 4: Estimated probability distribution of contact network for households with two 12–18 year olds and two 36+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 199$), and unpenalized likelihood estimates are shown.

c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
		Contact network					Estimate			95% C.I.	
0	0	0	1	1	0	0.07	0.01	0	[0, 0.17]	[0, 0.03]	[0, 0.02]
0	0	1	0	0	1	0.05	0	0	[0, 0.16]	[0, 0.02]	[0, 0]
0	0	1	1	1	1	0.03	0.04	0.03	[0, 0.16]	[0, 0.09]	[0, 0.09]
0	1	0	0	0	0	0.06	0	0	[0, 0.15]	[0, 0.02]	[0, 0]
0	1	1	1	1	1	0	0.03	0.06	[0, 0]	[0.01, 0.1]	[0.02, 0.12]
1	0	0	1	1	1	0	0.02	0.04	[0, 0]	[0, 0.06]	[0, 0.08]
1	0	1	0	0	1	0.07	0.01	0	[0, 0.17]	[0, 0.02]	[0, 0.01]
1	0	1	1	1	0	0.04	0.04	0.04	[0, 0.23]	[0, 0.12]	[0, 0.11]
1	0	1	1	1	1	0	0.12	0.11	[0, 0]	[0.02, 0.27]	[0.02, 0.26]
1	1	0	1	1	0	0.11	0.03	0.02	[0, 0.26]	[0, 0.09]	[0, 0.09]
1	1	0	1	1	1	0	0.07	0.07	[0, 0]	[0.01, 0.16]	[0.02, 0.16]
1	1	1	1	0	1	0	0.03	0.05	[0, 0]	[0.01, 0.1]	[0.01, 0.1]
1	1	1	1	1	0	0	0.07	0.07	[0, 0]	[0, 0.17]	[0, 0.16]
1	1	1	1	1	1	0.56	0.34	0.22	[0.35, 0.76]	[0.1, 0.51]	[0.07, 0.48]

Table 5: Estimated probability distribution of contact network for households with one 12–18 year old, one 19–35 year old, and two 36+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 20$), and unpenalized likelihood estimates are shown.

c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
		Contact network					Estimate			95% C.I.	
0	0	0	0	0	0	0	0.19	0	[0, 0]	[0, 0.15]	[0, 0.01]
0	0	0	1	1	0	0.14	0.02	0	[0, 0.33]	[0, 0.13]	[0, 0.03]
0	1	0	0	0	0	0.15	0.02	0.01	[0, 0.30]	[0, 0.15]	[0, 0.04]
0	1	0	0	1	1	0	0	0.03	[0, 0]	[0, 0.02]	[0, 0.06]
0	1	0	1	0	1	0	0.02	0.04	[0, 0]	[0, 0.09]	[0.01, 0.10]
0	1	0	1	1	0	0	0.04	0.04	[0, 0]	[0, 0.16]	[0, 0.12]
0	1	0	1	1	1	0.15	0.11	0.08	[0, 0.37]	[0.01, 0.27]	[0.02, 0.19]
0	1	1	0	1	1	0	0.01	0.03	[0, 0]	[0, 0.02]	[0, 0.09]
0	1	1	1	0	1	0.12	0.05	0.05	[0, 0.30]	[0, 0.16]	[0.01, 0.15]
0	1	1	1	1	0	0	0.02	0.05	[0, 0]	[0, 0.10]	[0, 0.13]
0	1	1	1	1	1	0.08	0.1	0.1	[0, 0.37]	[0.01, 0.31]	[0.02, 0.26]
1	1	0	1	0	1	0	0	0.03	[0, 0]	[0, 0.01]	[0.01, 0.06]
1	1	0	1	1	0	0	0.01	0.03	[0, 0]	[0, 0.06]	[0, 0.08]
1	1	0	1	1	1	0	0.03	0.06	[0, 0]	[0, 0.15]	[0.01, 0.14]
1	1	1	0	1	0	0.04	0.02	0.01	[0, 0.14]	[0, 0.1]	[0, 0.04]
1	1	1	0	1	1	0.04	0.04	0.02	[0, 0.14]	[0, 0.15]	[0, 0.08]
1	1	1	1	0	1	0	0.02	0.03	[0, 0]	[0, 0.10]	[0, 0.10]
1	1	1	1	1	0	0	0.02	0.03	[0, 0]	[0, 0.14]	[0, 0.11]
1	1	1	1	1	1	0.28	0.17	0.07	[0, 0.50]	[0.02, 0.38]	[0.01, 0.24]

Table 6: Estimated probability distribution of contact network for households with two 19–35 year olds and two 36+ year olds. Dyad-independent, penalized likelihood (CV-selected $\lambda = 96$), and unpenalized likelihood estimates are shown.

c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	indep.	MLE	pen.MLE	indep.
							Estimate			95% C.I.	
0	0	0	0	0	0	0	0.05	0.01	[0, 0]	[0, 0.06]	[0, 0.03]
0	0	0	0	0	1	0.26	0.08	0.06	[0.08, 0.45]	[0.02, 0.21]	[0.01, 0.18]
0	0	0	0	1	0	0.12	0.03	0.01	[0, 0.27]	[0, 0.09]	[0, 0.05]
0	0	0	0	1	1	0	0.09	0.08	[0, 0]	[0.03, 0.19]	[0.03, 0.18]
0	0	0	1	0	1	0.11	0.07	0.06	[0, 0.30]	[0.01, 0.16]	[0.01, 0.14]
0	0	0	1	1	1	0	0.08	0.09	[0, 0]	[0.02, 0.17]	[0.03, 0.18]
0	0	1	0	1	1	0	0.03	0.03	[0, 0]	[0, 0.09]	[0.01, 0.09]
0	0	1	1	0	1	0	0.02	0.03	[0, 0]	[0, 0.04]	[0, 0.06]
0	0	1	1	1	1	0	0.03	0.04	[0, 0]	[0, 0.07]	[0.01, 0.09]
0	1	0	0	0	1	0	0.04	0.05	[0, 0]	[0.01, 0.11]	[0.01, 0.11]
0	1	0	0	1	1	0	0.04	0.06	[0, 0]	[0.01, 0.09]	[0.03, 0.12]
0	1	0	1	0	1	0	0.05	0.05	[0, 0]	[0.01, 0.11]	[0.01, 0.12]
0	1	0	1	1	1	0	0.06	0.07	[0, 0]	[0.02, 0.14]	[0.02, 0.14]
0	1	1	0	0	1	0.04	0.02	0.02	[0, 0.18]	[0, 0.06]	[0, 0.05]
0	1	1	0	1	1	0.06	0.03	0.03	[0, 0.27]	[0, 0.10]	[0, 0.08]
0	1	1	1	0	1	0.05	0.03	0.02	[0, 0.21]	[0, 0.07]	[0, 0.06]
0	1	1	1	1	1	0.25	0.05	0.03	[0.03, 0.40]	[0.01, 0.13]	[0, 0.08]
1	1	0	0	0	1	0.06	0.01	0.01	[0, 0.22]	[0, 0.04]	[0, 0.02]
1	1	0	1	1	0	0.07	0.01	0	[0, 0.18]	[0, 0.03]	[0, 0.01]

Figure 2: Estimated probability distribution for households with two 0–5 year olds and two 19+ year olds. Labels are: c1 = younger child, c2 = older child, mom = female adult, dad = male adult
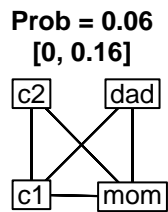
Prob = 0.65
[0.41, 0.89]

Prob = 0.12
[0, 0.32]

Prob = 0.08
[0, 0.2]

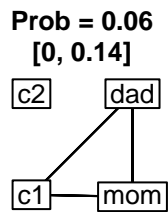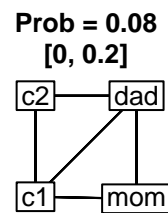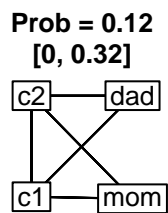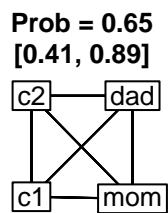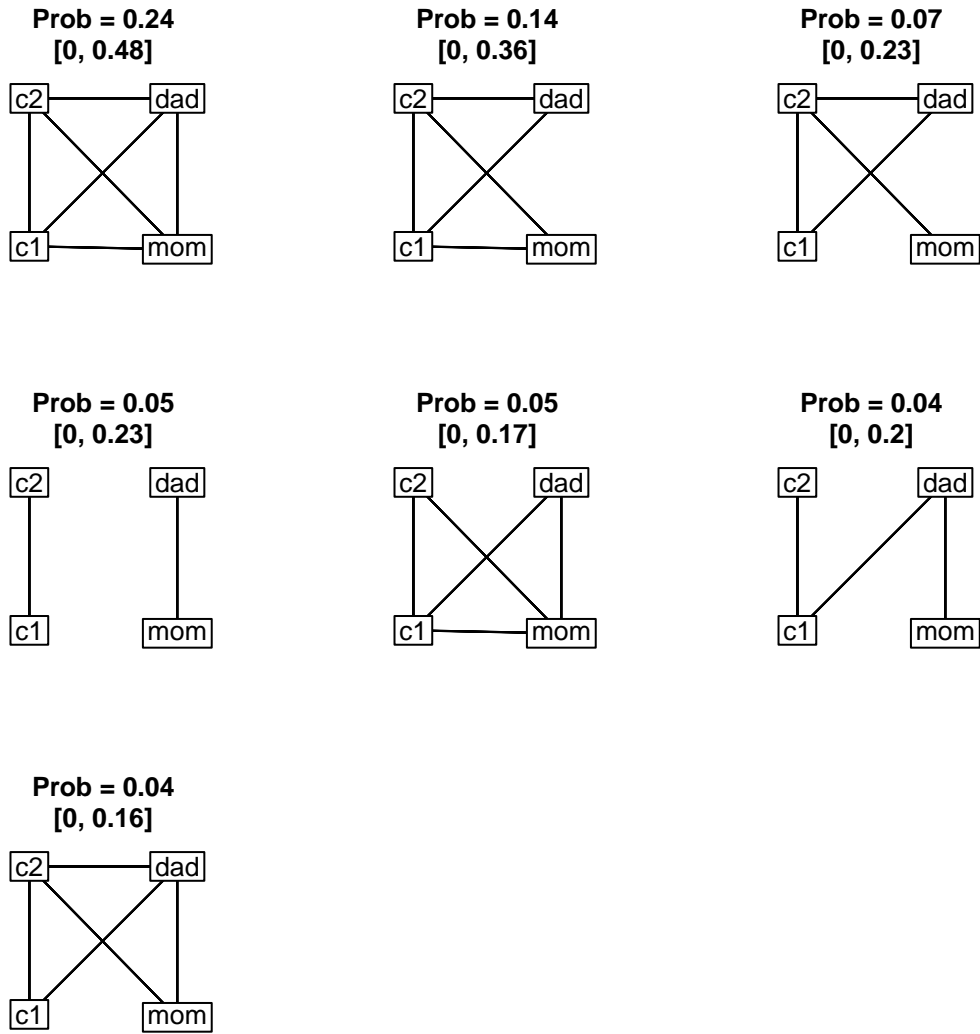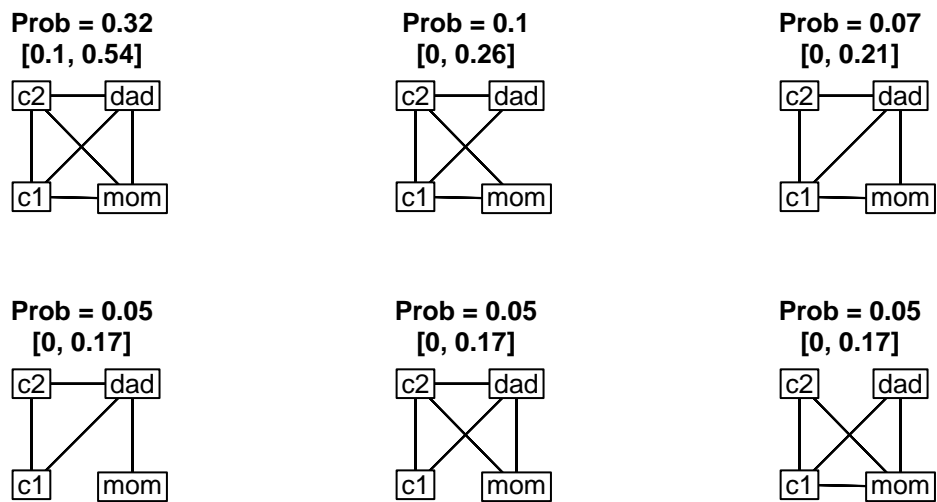**Prob = 0.06**
[0, 0.14]

Prob = 0.06
[0, 0.16]

Figure 3: Estimated probability distribution for households with one 0–5 year old, one 6–11 year old, and two 19+ year olds. Labels are: c1 = younger child, c2 = older child, mom = female adult, dad = male adult

Figure 4: Estimated probability distribution for households with two 6–11 year olds and two 19+ year olds. Labels are: c1 = younger child, c2 = older child, mom = female adult, dad = male adult
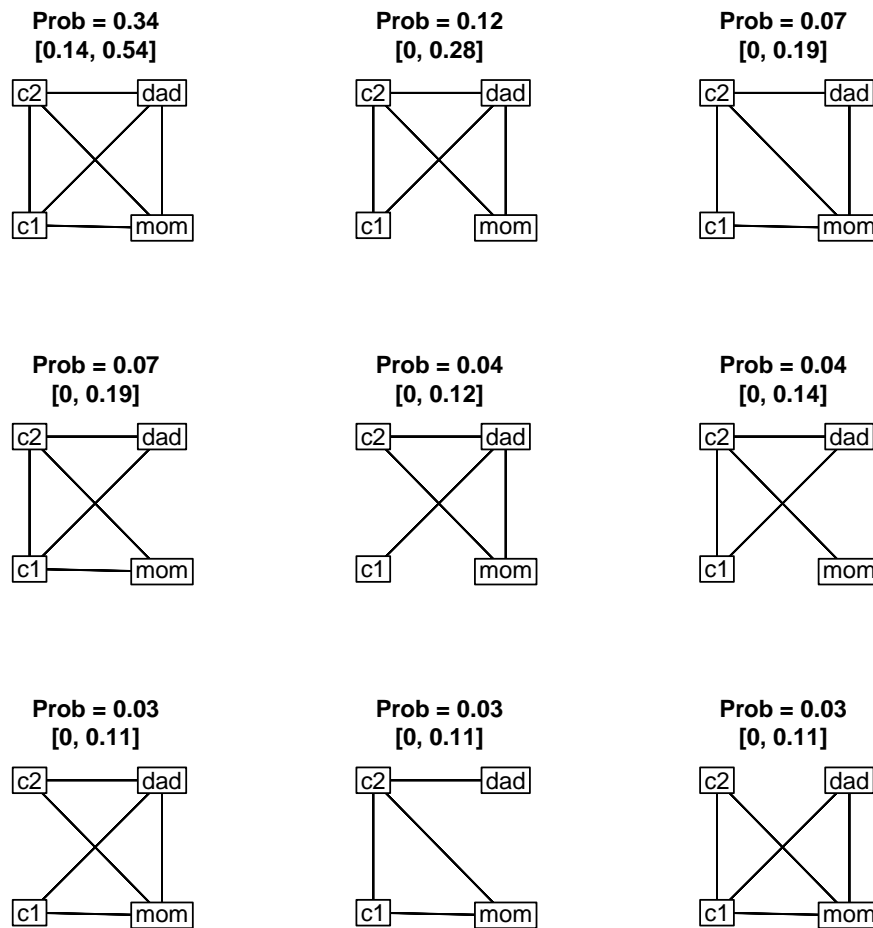
Prob = 0.32
[0.1, 0.54]

Prob = 0.1
[0, 0.26]

Prob = 0.07
[0, 0.21]

Prob = 0.05
[0, 0.17]

Prob = 0.05
[0, 0.17]

Prob = 0.05
[0, 0.17]

Figure 5: Estimated probability distribution for households with two 12–18 year olds and two 36+ year olds. Labels are: c1 = younger child, c2 = older child, mom = female adult, dad = male adult

Figure 6: Estimated probability distribution for households with one 12–18 year old, one 19–35 year olds and two 36+ year olds. Labels are: c1 = 12–18 year old, c2 = 19–35 year old, mom = female 36+ year old, dad = male 36+ year old
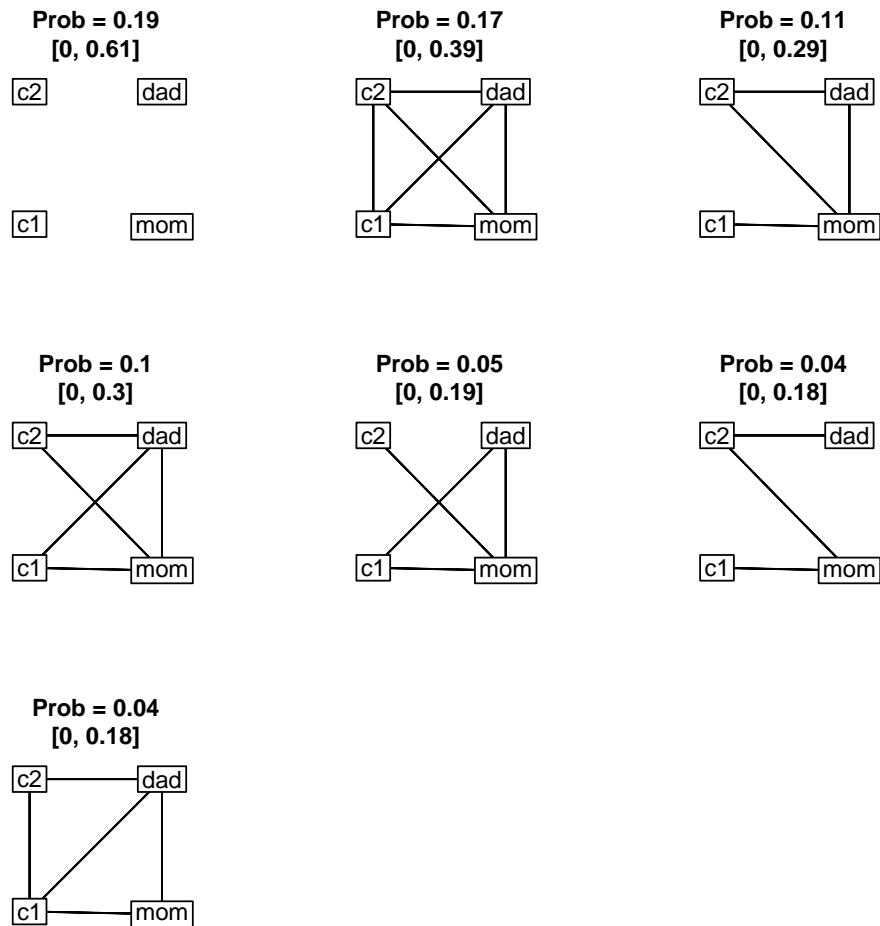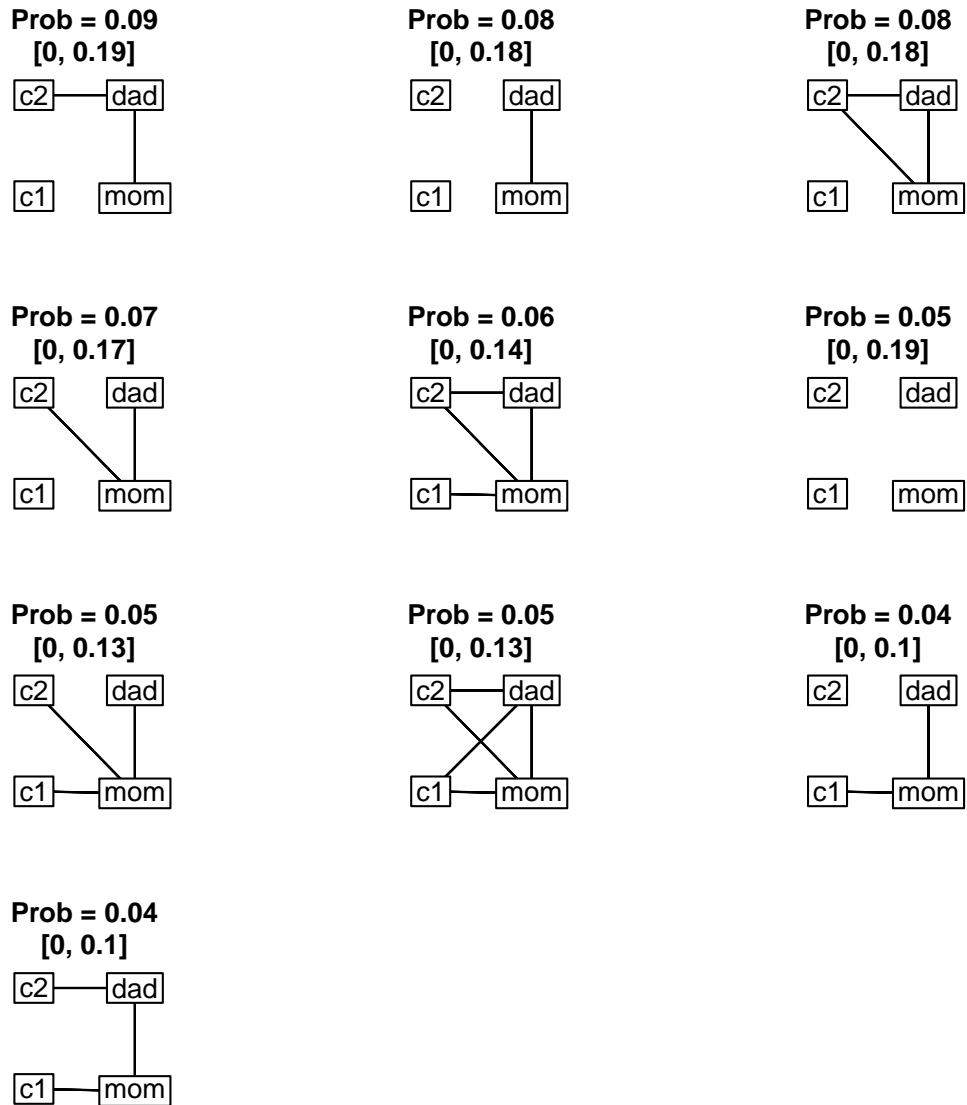
Figure 7: Estimated probability distribution for households with two 19–35 year olds and two 36+ year olds. Labels are: c1 = younger 19–35 year old, c2 = older 19–35 year old, mom = female 36+ year old, dad = male 36+ year old

3   Adjacency penalty results

Table 7 shows maximum likelihood and penalized likelihood estimates using the adjacency penalty for households with two 0–5 year olds and two 19–59 year olds. The low probability estimate for the complete network and the large difference between the penalized and unpenalized estimates suggest that this penalty oversmooths. Estimates for households with two 12–18 and two 36+ year olds are shown in Table 8. Again, oversmoothing is evident.

Table 7: Penalized MLE computed with CV-selected λ=2.4 using the adjacency penalty for households with two 0–5 year olds and two 19–59 year olds. (Networks with probability< 0.01 not depicted.)

Contact network						Estimate		95% C.I.	
c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	MLE	pen.MLE
0	1	0	0	0	1	0.04	0.01	[0, 0.11]	[0, 0.1]
0	1	1	0	0	1	0.04	0.01	[0, 0.11]	[0, 0.1]
1	1	1	0	1	1	0.07	0.05	[0, 0.22]	[0, 0.19]
1	1	1	1	0	1	0.05	0.01	[0, 0.18]	[0, 0.15]
1	1	1	1	1	0	0.14	0.25	[0, 0.45]	[0.1, 0.29]
1	1	1	1	1	1	0.65	0.34	[0.3, 0.95]	[0.25, 0.4]

Figure 8 shows the average mean squared error and variance-bias decomposition for simulations based on characteristics of households with two 0–5 year olds and two 19+ year-olds using the adjacency penalty. Figure 9 shows results for simulations based on characteristics of households with two 12–18 year olds and two 36+ year-olds. In both household types, average mean squared error decreases, while squared bias increases and variance decreases. The plots of probability parameter estimates show the oversmoothing arising from this penalty: estimates for large values of λ differ substantially from the unpenalized MLE.

Table 8: Penalized MLE computed with CV-selected λ=2.8 using the adjacency penalty for households with two 12–18 year olds and two 36+ year olds. (Networks with probability< 0.01 not depicted.)

Contact network						Estimate		95% C.I.	
c1-c2	c1-m	c1-d	c2-m	c2-d	m-d	MLE	pen.MLE	MLE	pen.MLE
0	0	0	1	0	0	0	0.03	[0, 0.09]	[0, 0.09]
0	0	0	1	1	0	0.07	0.04	[0, 0.12]	[0, 0.10]
0	0	1	0	0	1	0.05	0.05	[0, 0.14]	[0, 0.13]
0	0	1	1	1	1	0.03	0.05	[0, 0.12]	[0, 0.11]
0	1	0	0	0	0	0.06	0.05	[0, 0.11]	[0, 0.10]
1	0	1	0	0	1	0.07	0.02	[0, 0.17]	[0, 0.07]
1	0	1	1	1	0	0.04	0.06	[0, 0.21]	[0, 0.14]
1	0	1	1	1	1	0	0.13	[0, 0.40]	[0.04, 0.21]
1	1	0	1	1	0	0.11	0.06	[0, 0.23]	[0, 0.13]
1	1	0	1	1	1	0	0.08	[0, 0.19]	[0, 0.16]
1	1	1	0	0	1	0	0.05	[0, 0.10]	[0, 0.12]
1	1	1	1	1	0	0	0.10	[0, 0.17]	[0, 0.20]
1	1	1	1	1	1	0.56	0.25	[0, 0.70]	[0.19, 0.34]

Figure 8: Simulation results based on the characteristics of households with two 0–5 year olds and two 19+ year-olds. Left panel: mean average squared error versus λ, middle panel: squared bias versus λ, right panel: variance versus λ.
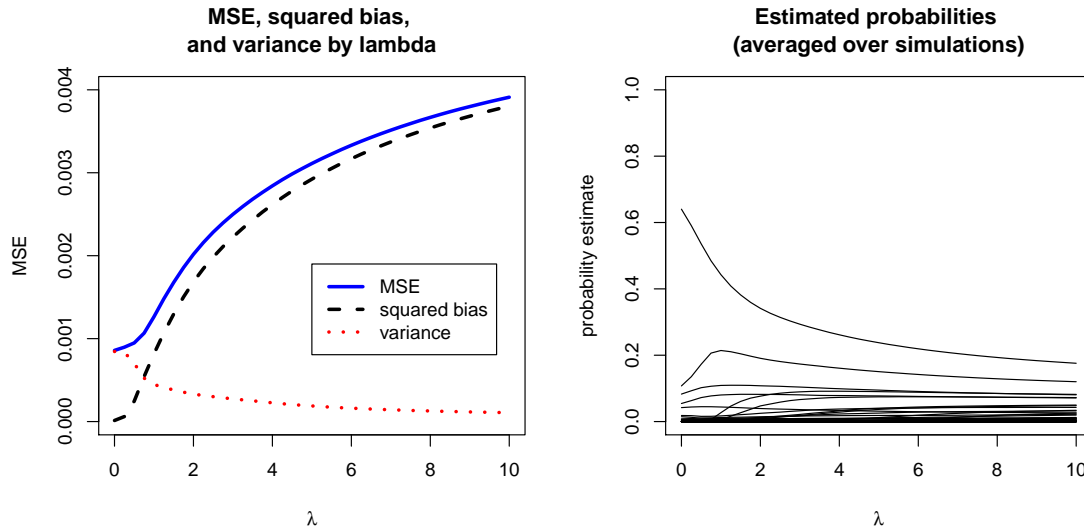

Figure 9: Simulation results based on the characteristics of households with two 12–18 year olds and two 36+ year-olds. Left panel: mean average squared error versus λ, middle panel: squared bias versus λ, right panel: variance versus λ.