Online Supplement to Innovation, Cooperation, and the Structure of Three Regional Sustainable Agriculture Networks in California

Regional Environmental Change

Michael Levy and Mark Lubell*

2017-09-30

*Corresponding author. Department of Environmental Science and Policy, University of California, Davis. malevy@ucdavis.edu

Supplementary Methods

ERGMs

The ERGM likelihood is:

$$P(Y = \mathbf{y}) = \frac{exp(\theta^{\top}\Gamma(\mathbf{y}, x_y))}{\sum_{z\in\mathcal{Y}} exp(\theta^{\top}\Gamma(\mathbf{z}, x_z))}, \mathbf{y} \in \mathcal{Y} \tag{1}$$

where Γ is a vector of sufficient statistics calculated on the network $\mathbf{y}$ and covariate information on the network x_y, θ is the corresponding vector of parameter values, and $\mathcal{Y}$ is the set of possible networks (Frank and Strauss 1986; Morris et al. 2008; Lusher et al. 2012).

We obtained parameter estimates for our ERGMs using Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) (Hunter et al. 2008; Handcock et al. 2016). For each network, we used a burn-in interval of 80,000 draws followed by 10,000 samples taken at an interval of 5,000 draws. MCMC and goodness-of-fit diagnostics are presented below.

Geometrically-weighted statistics

The geometrically-weighted statistics are sums with decreasing weights for additional units on a given node or edge. For example, any edge augments the GWD statistic, but if the edge is added to higher degree nodes, the statistic increases less than if it were added to a lower degree node. Thus, GWD "may be thought of as a sort of anti-preferential attachment model term" (Hunter 2007), and in this manner parameterizes the degree distribution of a network. The rate of decreasing return as edges are added to higher degree nodes is governed by the θ_s decay parameter. Small values of θ_s yield a greater disparity between edges to low- and high-degree nodes. For intuition on the relationship between GWD, θ_s, and the degree distribution, see ref (Levy 2016). GWESP is similar to a statistic that counts the number of edges with a shared partner (i.e. three times the number of triangles), but with decreasing effect of each shared partner after the first. The rate of decrease is controled by θ_t, with smaller values producing a rapidly diminishing effect of additional shared partners. Specifically, the GWD and GWESP statistics are given by:

$$GWD = e^{\theta_s} \sum_{i=1}^{N-1} [1 - (1 - e^{-\theta_s})^i] D_i(\mathbf{y}); \tag{2}$$

$$GWESP = e^{\theta_t} \sum_{i=1}^{N-2} [1 - (1 - e^{-\theta_t})^i] EP_i(\mathbf{y}), \tag{3}$$

where D_i is the number of nodes of degree i and EP_i is the number of edges with i shared partners. We fixed decay parameter values and held them constant across all three networks to facilitate comparison of the parameter estimates. For GWD, $\theta_s = 3.0$ ensures that the statistic captures popularity effects among high degree nodes. For GWESP, $\theta_t = 0.7$ produces a moderately declining marginal effect of shared partners.

Modeled edge probabilities

The probability of observing a tie between nodes i and j is:

$$P(y_{ij}|\mathbf{y}_{ij}^{-}, \theta) = logistic(\theta^{\top} \delta_{\Gamma}(\mathbf{y}_{ij})), \tag{4}$$

where $\delta_{\Gamma}(\mathbf{y}_{ij})$ is a vector of change statistics associated with the ij edge given by $\Gamma(\mathbf{y}_{ij}^{+}) - \Gamma(\mathbf{y}_{ij}^{-})$, where $\mathbf{y}_{ij}^{+}$ is the network with the ij tie present and $\mathbf{y}_{ij}^{-}$ with the ij tie absent, and logistic is the inverse logit function: $(1 + e^{-x})^{-1}$ (Hunter et al. 2008; Desmarais and Cranmer 2012).

We parse edge probabilities by both nodal and dyadic attributes. Parsing edge probabilities by nodal attributes (profession and degree) results in each dyad being included twice, once for each actor in the dyad.

Supplementary Results

Survey Response Rates

	Central Coast	Napa	Lodi
Sample Size	1489	611	500
Number Responders	358	237	227
Response Rate	0.240	0.388	0.454

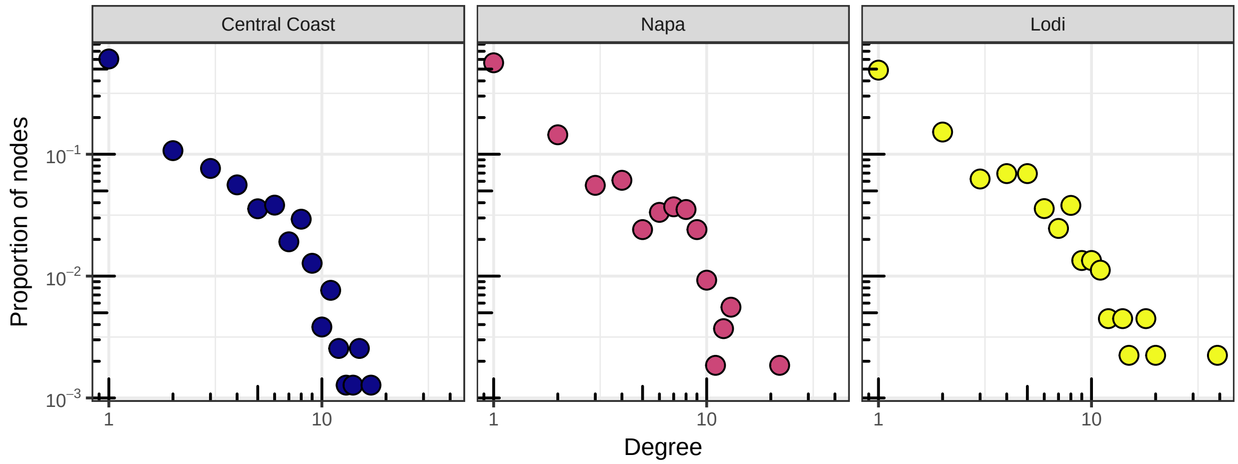
Table 1: Survey response details for three regional vineyard management communication networks.

Descriptive Statistics

	Central Coast	Napa	Lodi
Actors	785	540	447
Average Number Ties	2.4	2.6	3.0
Density	0.0031	0.0048	0.0068
Connectedness	0.73	0.79	0.84
HMPL	8.3	6.1	5.0
Degree Centralization	0.019	0.036	0.081
Clustering	0.034	0.034	0.057

Table 2: Descriptive statistics for three regional vineyard managment communication networks.

Degree Distributions



ERGM Robustness

Here, we check the robustness of our conclusions to decisions made in the analytical pipeline. In summary, our methodological decisions had very little impact on the substantive interpretation of the models. Specifically, here we estimate ERGMs on networks with two differences to the models presented in the manuscript: 1) We originally treated ties as undirected, based on the logic that communication about vineyard management is generally bi-directional. In the first table below, we maintain the directionality of ties and estimate distinct geometrically-weighted *in*-degree and *out*-degree distribution terms. And, 2) Survey respondents in Lodi were only allowed to nominate four growers and four other individuals with whom they communicated, while respondents in Napa and Central Coast were able to nominate eight of each. To enable regional comparisons, we originally trimmed the nominations in Napa and Central Coast to no more than four of each; here we include all nominations in all regions.

Table 3: Parameter estimates for ERGMs of directed viticulture knowledge networks. Outreach and Outreach-Grower are main effect terms for incoming ties, and Respondent is a main effect term for sent ties. Geographic terms are homophily effects at the zip-code level.

	Central Coast	Napa	Lodi
Anti-centralization of incoming ties (GWIDegree, $\theta_s = 3.0$)	-3.49^{***}	-3.70^{***}	-3.65^{***}
	(0.19)	(0.16)	(0.15)
Anti-centralization of outgoing ties (GWODegree, $\theta_s = 3.0$)	-2.28^{***}	-2.69^{***}	-3.03^{***}
	(0.36)	(0.34)	(0.29)
Triangles (GWESP, $\theta_t = 0.7$)	0.50^{***}	0.58^{***}	0.38^{***}
	(0.08)	(0.07)	(0.07)
Outreach Nominated	1.62^{***}	-0.11	0.93^{***}
	(0.14)	(0.27)	(0.11)
Outreach-Grower Nominated	0.81^{***}	1.07^{***}	0.86^{***}
	(0.10)	(0.15)	(0.10)
Shared zip-code	1.54^{***}	0.61^{***}	0.99^{***}
	(0.10)	(0.10)	(0.10)
Possibly shared zip-code	-1.34^{***}	0.17	-0.27^{*}
	(0.15)	(0.17)	(0.12)
Survey Respondent Nominations	6.52^{***}	8.84^{***}	3.31^{***}
	(0.38)	(1.26)	(0.20)

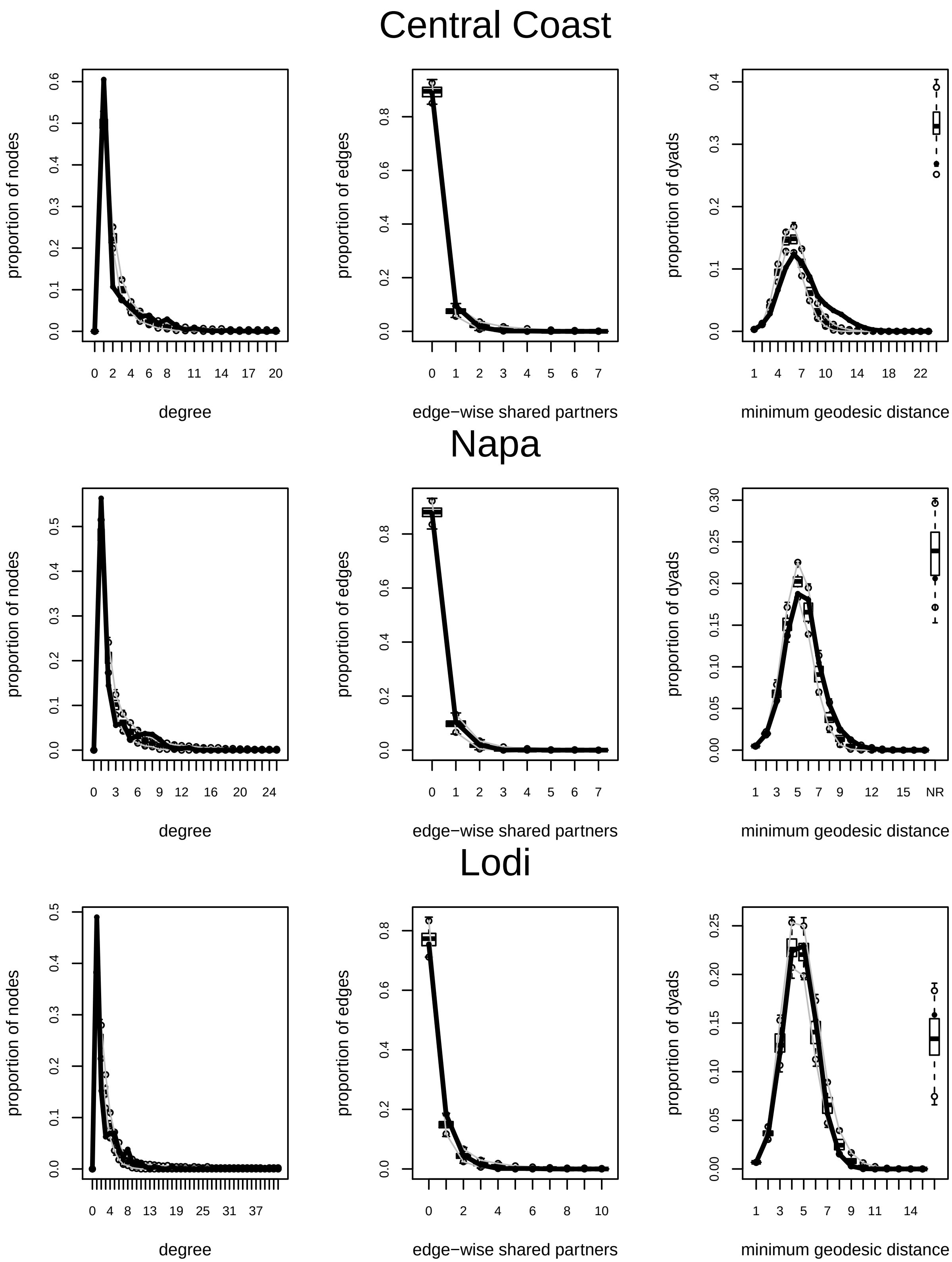
*p<0.05; **p<0.01; ***p<0.001

Table 4: Parameter estimates for ERGMs of viticulture knowledge networks with no nominations trimmed. Outreach, Outreach-Grower and Respondent are main-effect terms. Geographic terms are homophily effects at the zip-code level.

	Central Coast	Napa	Lodi
Anti-centralization (GWDegree, $\theta_s = 3.0$)	-2.29^{***}	-2.57^{***}	-2.52^{***}
	(0.15)	(0.16)	(0.18)
Triangles (GWESP, $\theta_t = 0.7$)	0.20^{***}	0.15^{**}	0.40^{***}
	(0.05)	(0.05)	(0.06)
Outreach	0.12	-0.82^{***}	0.22^{**}
	(0.13)	(0.17)	(0.07)
Outreach-Grower	0.25^{***}	0.28^{***}	0.23^{***}
	(0.04)	(0.05)	(0.05)
Shared zip-code	1.24^{***}	0.64^{***}	0.89^{***}
	(0.09)	(0.09)	(0.09)
Possibly shared zip-code	-0.17	0.54^{***}	-0.15
	(0.12)	(0.13)	(0.13)
Survey respondent	1.25^{***}	1.27^{***}	0.52^{***}
	(0.07)	(0.10)	(0.08)

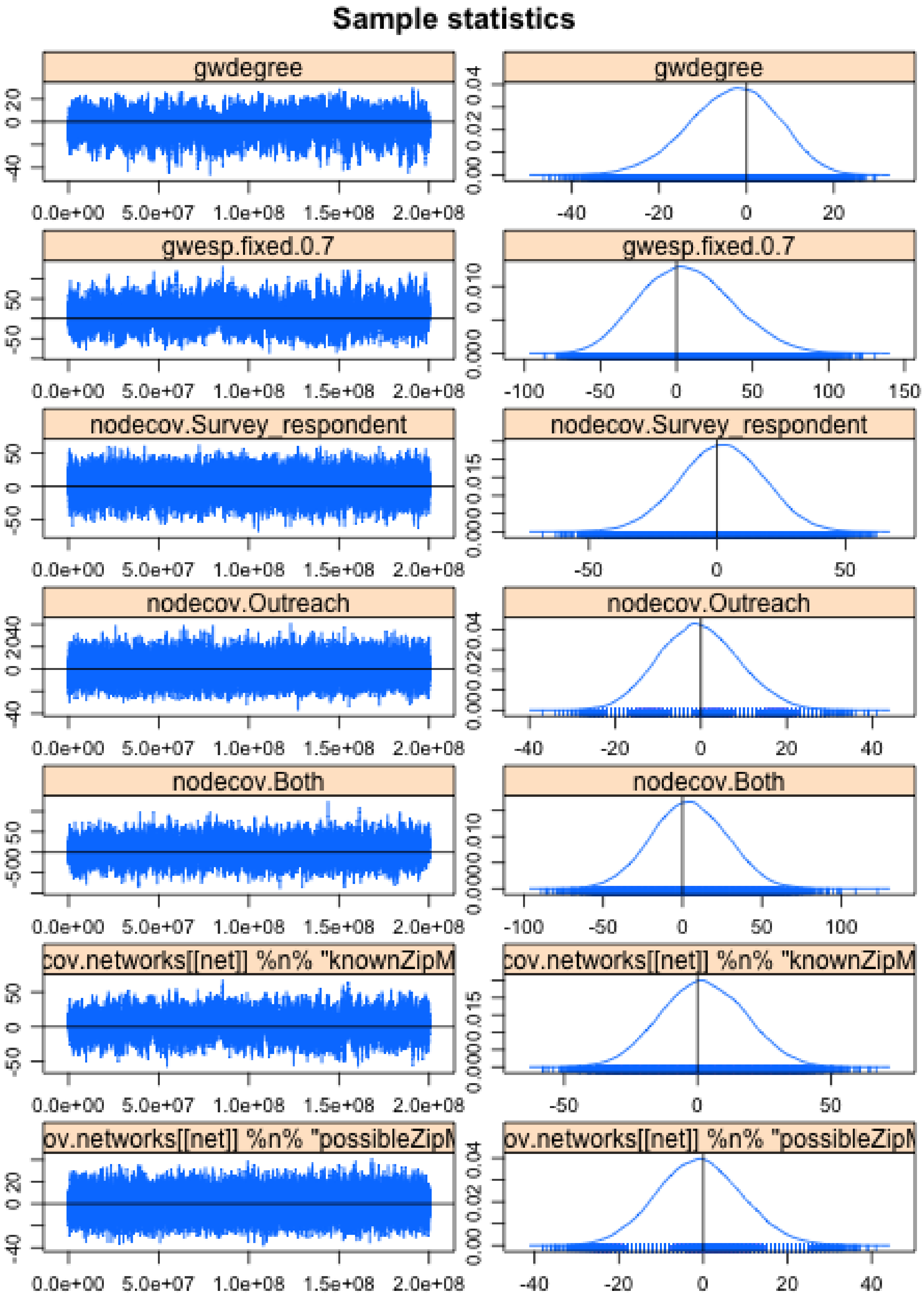
*p<0.05; **p<0.01; ***p<0.001

ERGM Goodness-of-fit

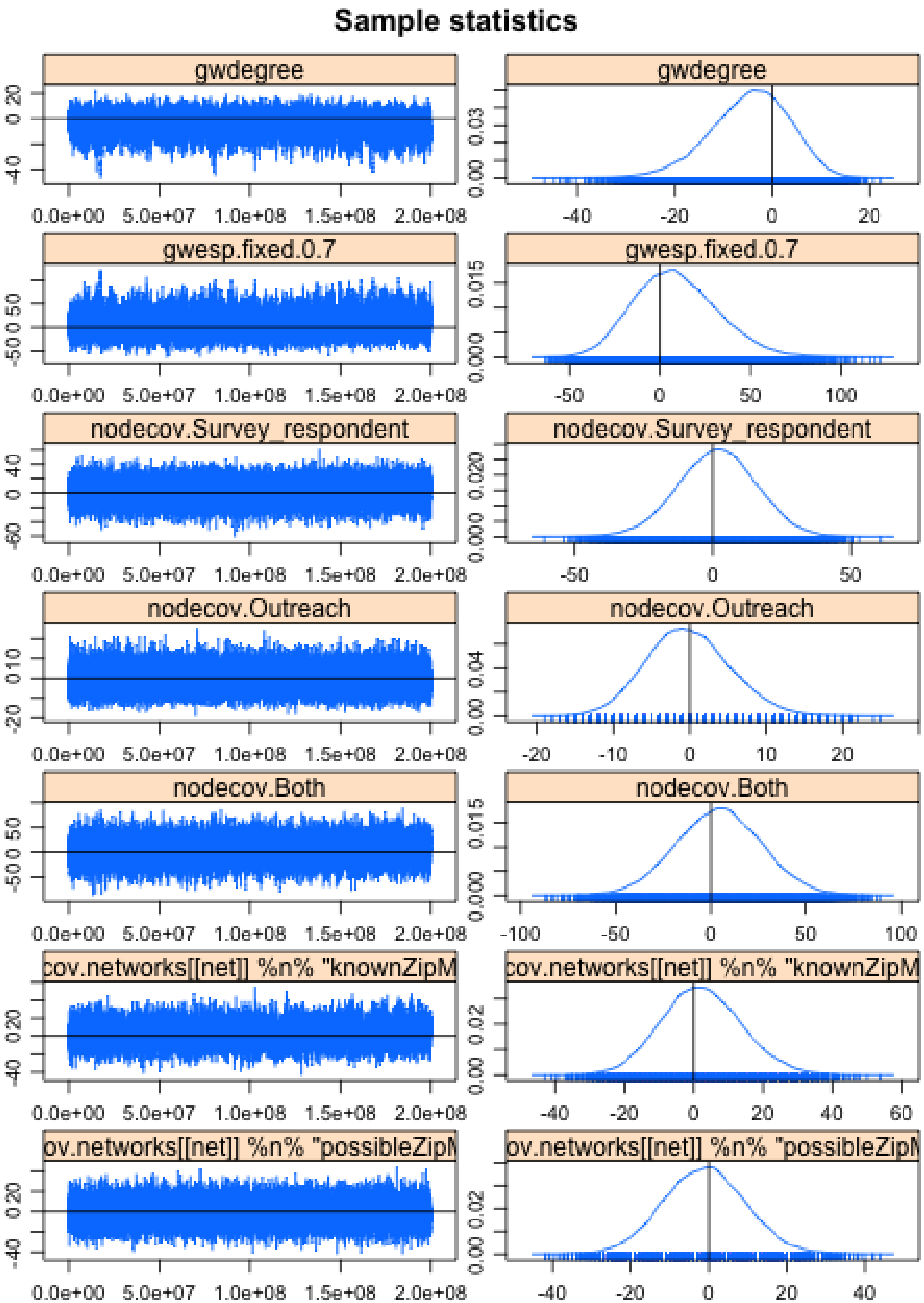


ERGM MCMC Diagnostics

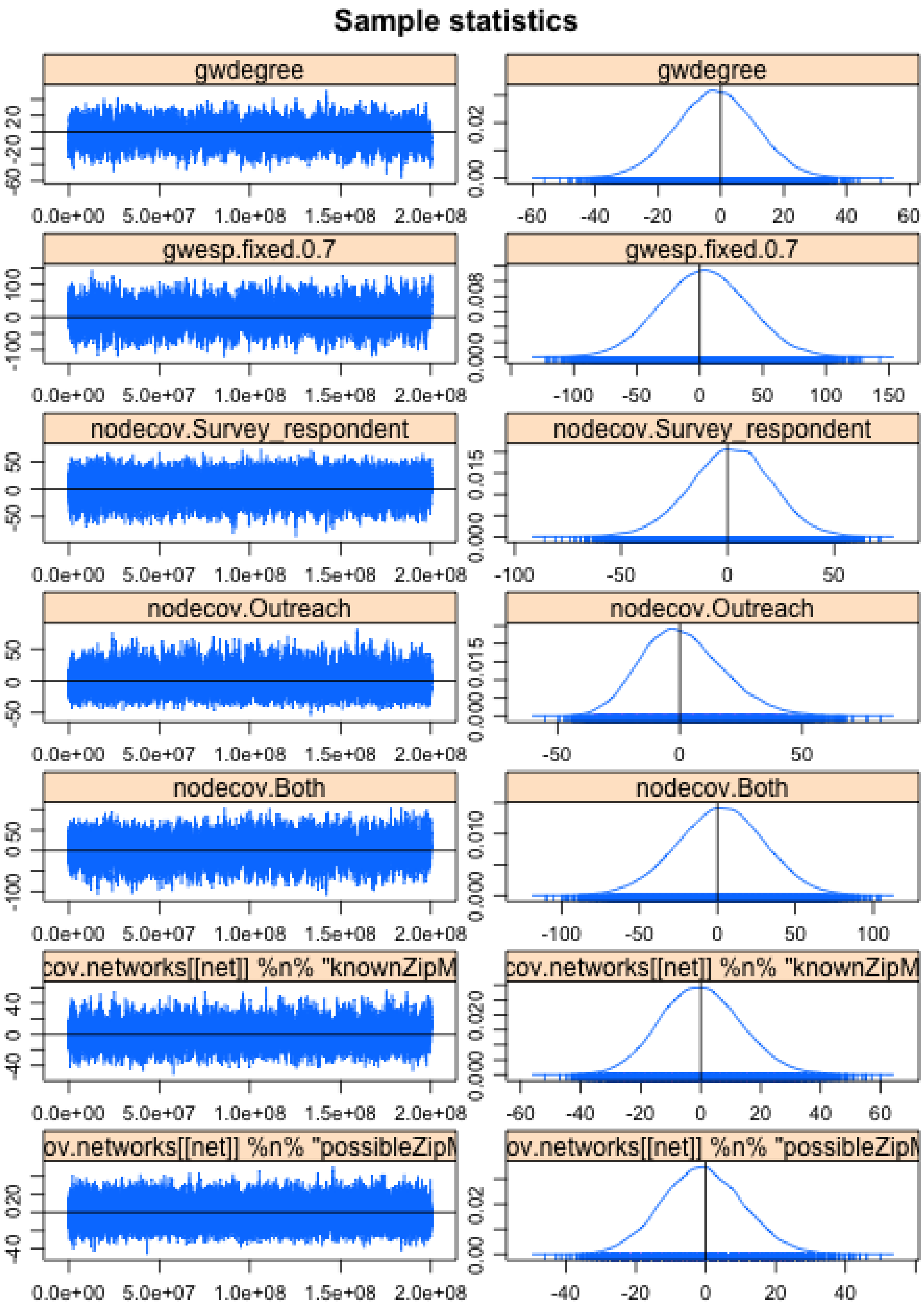
Central Coast



Napa


Sample statistics
gwdegree
gwdegree
gwesp.fixed.0.7
gwesp.fixed.0.7
nodecov.Survey_respondent
nodecov.Survey_respondent
nodecov.Outreach
nodecov.Outreach
nodecov.Both
nodecov.Both
cov.networks[[net]] %n% "knownZipM
cov.networks[[net]] %n% "knownZipM
ov.networks[[net]] %n% "possibleZipM
ov.networks[[net]] %n% "possibleZipM
0.0e+00 5.0e+07 1.0e+08 1.5e+08 2.0e+08

Lodi


Sample statistics
gwdegree
gwesp.fixed.0.7
nodecov.Survey_respondent
nodecov.Outreach
nodecov.Both
cov.networks[[net]] %n% "knownZipM
ov.networks[[net]] %n% "possibleZip
0.0e+00 5.0e+07 1.0e+08 1.5e+08 2.0e+08

References

Desmarais BA, Cranmer SJ (2012) Micro-level interpretation of exponential random graph models with application to estuary networks. Policy Studies Journal 40:402–434.

Frank O, Strauss D (1986) Markov graphs. Journal of the american Statistical association 81:832–842.

Handcock MS, Hunter DR, Butts CT et al (2016) Statnet: Software Tools for the Statistical Analysis of Network Data.

Hunter DR (2007) Curved exponential family models for social networks. Social Networks 29:216–230. doi: 10.1016/j.socnet.2006.08.005

Hunter DR, Handcock MS, Butts CT et al (2008) Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Journal of statistical software 24:nihpa54860.

Levy MA (2016) Gwdegree: Improving interpretation of geometrically-weighted degree estimates in exponential random graph models. The Journal of Open Source Software. doi: 10.21105/joss.00036

Lusher D, Koskinen J, Robins G (2012) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications. Cambridge University Press

Morris M, Handcock MS, Hunter DR (2008) Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. Journal of statistical software 24:1548–7660.