Social network analysis and exponential random graph modelling of dengue spatiotemporal linkage networks in an urban Malaysian district

Siti Najiha Md Asari[1], Fatimah Abdul Razak[2], Mohd Rahim Sulong[3], Wan Ming Keong[3], Nazarudin Safian[1]*

Affiliations

[1]Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia (UKM), Kuala Lumpur, Malaysia.

[2]Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia.

[3]Vector-Borne Disease Control Sector, Disease Control Division, Ministry of Health Malaysia, Putrajaya, Malaysia

Corresponding author

Nazarudin Safian

Department of Public Health Medicine,

Faculty of Medicine, Universiti Kebangsaan Malaysia,

Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras,

Kuala Lumpur, Malaysia

E-mail: nazarudin@hctm.ukm.edu.my

Supplementary Information

Supplementary Table S1 Exponential Random Graph Model (ERGM) estimates for the network formation in Petaling District

Term	Model 1: Basic Model		Model 2: Attribute-Based Model		Model 3: Attribute and Structural Model		
	Estimate (SE)	**Exponentiated coefficient (95% CI)**	**Estimate (SE)**	**Exponentiated coefficient (95% CI)**	**Estimate (SE)**	**Exponentiated coefficient (95% CI)**	***p*-value**
Edges	−8.27 (0.007)	0.00026 (0.00025–0.00027)	–9.822 (0.024)	0.000054 (0.000051–0.000056)	−9.818 (0.026)	0.000055 (0.000052–0.000057)	< 0.0001
Occupation (nodefactor)							
Employed	–	–	Ref	–	Ref	–	–
Students	–	–	0.071 (0.012)	1.07 (1.04–1.10)	0.071 (0.012)	1.07 (1.05–1.10)	< 0.0001
Unemployed	–	–	0.010 (0.013)	1.01 (0.98–1.04)	0.011 (0.013)	1.01 (0.98–1.04)	0.415
Housing type (level-specific nodematch)	–	–	1.900 (0.020)	6.68 (6.42–6.94)	1.896 (0.020)	6.66 (6.40–6.93)	< 0.0001
High-rise–High-rise							
Low-rise–Low-rise	–	–	1.466 (0.022)	4.33 (4.15–4.52)	1.467 (0.023)	4.34 (4.15–4.54)	< 0.0001
Citizenship (level-specific nodematch)	–	–	0.335 (0.018)	1.40 (1.35–1.45)	0.334 (0.018)	1.40 (1.35–1.45)	< 0.0001
Citizen–Citizen							
Non-Citizen–Non-citizen	–	–	1.376 (0.033)	3.96 (3.71–4.22)	1.376 (0.033)	3.96 (3.70–4.22)	< 0.0001
gwdegree (decay 0.25)	–	–	–	–	−0.019 (0.041)	0.98 (0.91–1.06)	0.648
Model Fit							
Log-likelihood	−198,737.5		−191,852.5		−191,852.0		
AIC	397,477		383,719		383,720		
BIC	397,494		383,833		383,850		

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Ref = Reference category; SE = Standard Error; CI = Confidence Interval. The table presents coefficients, standard errors, exponentiated coefficients and 95% confidence intervals for three sequentially fitted

models: Model 1: Basic structural model (edges only); Model 2: Attribute-based model including occupation, housing type, and citizenship; and Model 3: Full model incorporating both node attributes and structural effects (gwdegree). Model fit statistics comprising the log-likelihood, AIC and BIC are reported for each specification.

Supplementary Table S2 MCMC convergence diagnostics for the final ERGM

MCMC	**Value**
Geweke z-scores	+0.476 to +1.683
Joint *p*	0.873
Autocorrelation	0.113–0.568 at lag 122,000
Cross-correlation	0.007–0.883

Cross-correlations between MCMC samples reflect the degree to which parameter estimates co-vary within the sampling chain. High cross-correlations between the edges term and covariate terms (up to 0.883) are expected in ERGMs, as the edges parameter functions as a network-level intercept and is structurally correlated with all other model terms. These values do not indicate model instability and should be interpreted alongside the Geweke joint *p* (0.873) and autocorrelation: 0.113–0.568 at the first stored lag, declining towards zero at longer lags, which together indicate no evidence of non-stationarity or residual serial dependence.

Supplementary Table S3 Goodness-of-fit statistics for the final ERGM.

Statistics	Observed	Simulated mean	Monte Carlo *p*-value
Edges	21,424	21,471.29	0.850
Nodefactor: Occupation			
Students	8,700	8,791.94	0.275
Unemployed	7,063	7,106.84	0.665
Nodematch: Citizenship			
Citizen	16,442	16,472.48	0.865
Non-citizen	1,202	1,195.28	0.785
Nodematch: Housing type			
High-rise	12,437	12,465.91	0.720
Low-rise	5,821	5,846.80	0.825
gwdegree. Fixed 0.25	15,138	15,146.79	0.850

Comparing observed network statistics with simulated means and Monte Carlo *p* values for structural, occupation, citizenship, housing-type, and gwdegree terms.

Supplementary Table S4 Sensitivity analysis of network structure to alternative linkage definitions in the Petaling dengue network

Metric	**100 m, 7 days**	**100 m, 14 days**	**100 m, 21 days**	**200 m, 7 days**	**200 m, 14 days**	**200 m, 21 days**	**300 m, 7 days**	**300 m, 14 days**	**300 m, 21 days**
Number of nodes	6,477	8,642	10,109	10,131	12,978	14,508	11,538	15,145	17,101
Number of edges	6,520	11,052	15,363	12,266	21,424	30,152	19,638	34,954	49,570
Density	0.00031	0.00030	0.00030	0.00024	0.00025	0.00029	0.00029	0.00031	0.00034
Average degree	2.01	2.56	3.04	2.42	3.3	4.16	3.4	4.62	5.8
Maximum degree	17	25	37	24	33	46	32	43	64
Number of components	2,280	2,666	2,762	2,992	2,957	2,596	2,945	2,321	1,692
Largest component size	71	114	132	130	211	411	220	916	1126
Average component size	2.84	3.24	3.66	3.39	4.4	5.6	3.9	6.52	10

Networks were reconstructed under combinations of spatial radius (100 m, 200 m, 300 m) and onset interval (7 days, 14 days, 21 days). Shown are network size, density, degree structure, and component structure. The primary analysis definition was 200 m/14 days.

Supplementary Table S5 Sensitivity analyses of ERGM estimates across spatial radii (100 m, 200 m, 300 m) with a fixed 14-day window

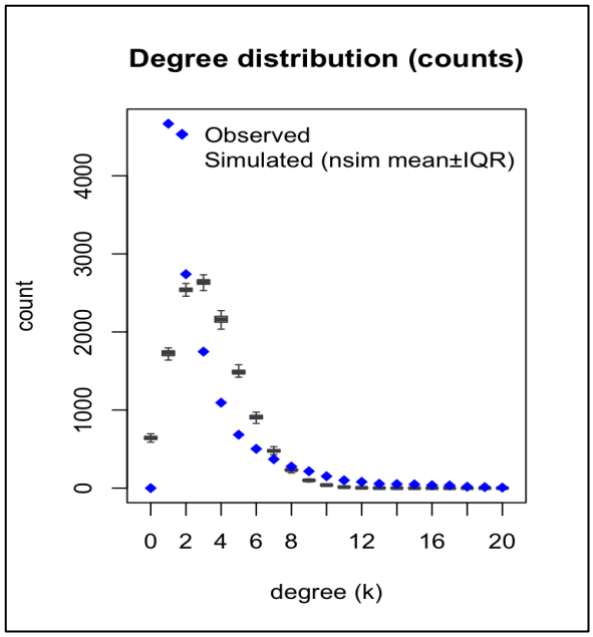
Term	Model 3: 100 m 14 days		Model 3: 200 m 14 days		Model 3: 300 m 14 days	
	Estimate (SE)	**Exponentiated coefficient (95% CI)**	**Estimate (SE)**	**Exponentiated coefficient (95% CI)**	**Estimate (SE)**	**Exponentiated coefficient (95% CI)**
Edges	−10.7 (0.05)	0.000022 (0.000020–0.000025)	−9.818 (0.026)	0.000055 (0.000052–0.000057)	−9.3 (0.02)	0.000091 (0.000089–0.000095)
Occupation (nodefactor)						
Employed	Ref	–	Ref	–	Ref	–
Students	0.128 (0.019)	1.14 (1.1–1.2)	0.071 (0.012)	1.07 (1.05–1.10)	0.073 (0.01)	1.08 (1.06–1.1)
Unemployed	0.013 (0.02)	1.01 (0.97–1.06)	0.011 (0.013)	1.01 (0.98–1.04)	0.008 (0.0098)	1.01 (0.99–1.03)
Housing type (level-specific nodematch)						
High-rise–High-rise	2.7 (0.04)	14.4 (13.3–15.57)	1.896 (0.020)	6.66 (6.40–6.93)	1.48 (0.014)	4.41 (4.29–4.53)
Low-rise–Low-rise	2.2 (0.043)	9.40 (8.6–10.2)	1.467 (0.023)	4.34 (4.15–4.54)	1.11 (0.015)	3.04 (2.95–3.13)
Citizenship (level-specific nodematch)						
Citizen–Citizen	0.34 (0.026)	1.40 (1.33–1.48)	0.334 (0.018)	1.40 (1.35–1.45)	0.32 (0.014)	1.40 (1.34–1.42)
Non-Citizen–Non-citizen	1.58 (0.04)	4.8 (4.44–5.28)	1.376 (0.033)	3.96 (3.70–4.22)	1.29 (0.026)	3.65 (3.47–3.84)
Gwdegree (decay 0.25)	0.86 (0.05)	2.40 (2.14–2.63)	−0.019 (0.041)	0.98 (0.91–1.06)	−0.89 (0.039)	0.41 (0.38–0.44)
Model Fit	Log-likelihood	AIC	Log-likelihood	AIC	Log-likelihood	AIC
	−95,767	191,550	−191,852.0	383,720	−313,014	626,043

AIC = Akaike Information Criterion; Ref = Reference category; SE = Standard Error; CI = Confidence Interval; Values are estimated (SE) and exponentiated coefficient (95% CI). Models include nodefactor (occupation), level-specific nodematch (housing type; citizenship), and gwdegree (decay 0.25).

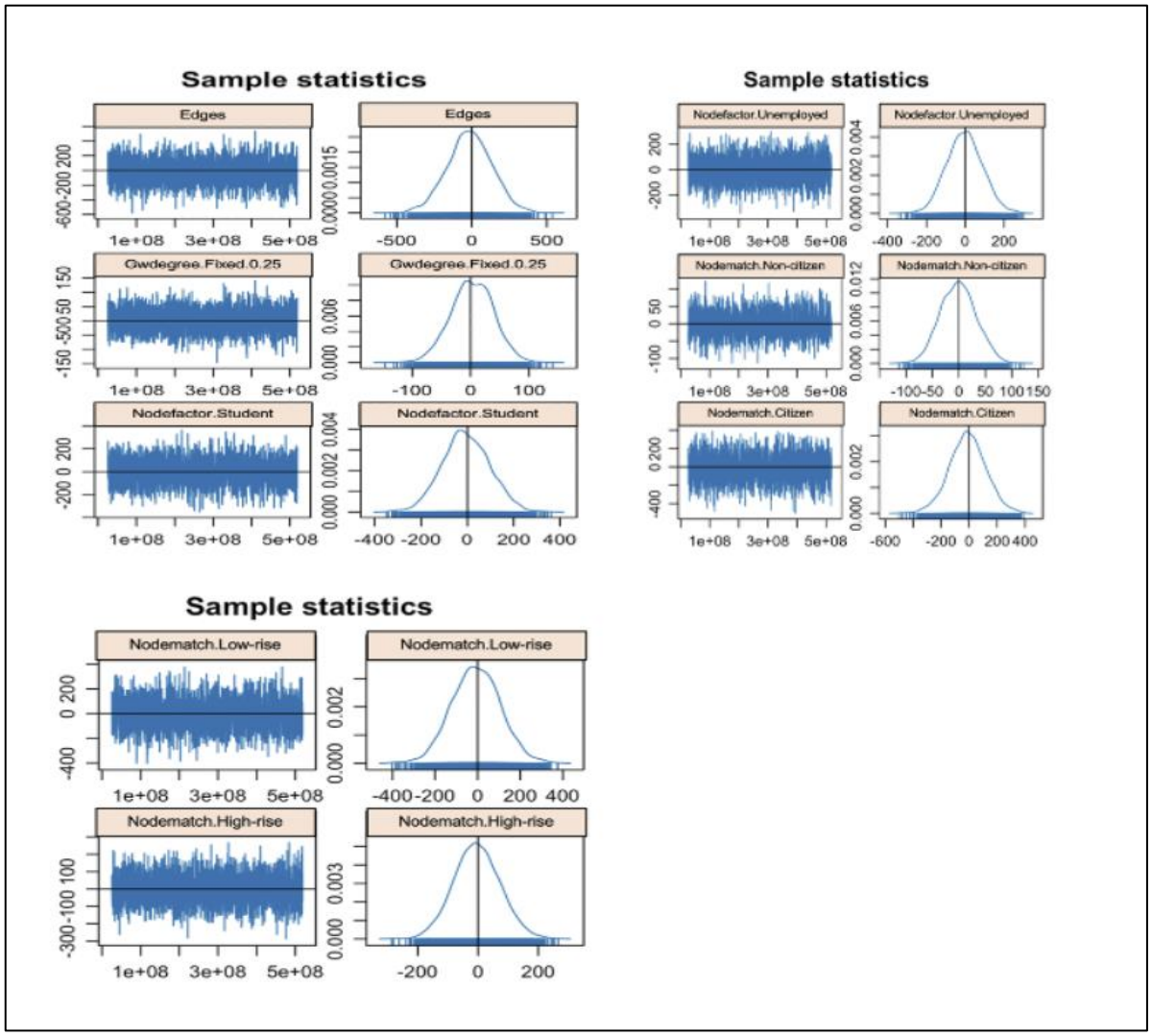
Supplementary Table S6 Covariate evaluation and model stability checks

Category	Included in final ERGM	Evaluated but not retained	Not available in the dataset
Demographics	Citizenship; Occupation	Age; Sex; Ethnicity (non-convergent)	–
Residential / built environment	Housing type	–	Population density proxy
Space/time context	Linkage definition (200 m/14-day)	–	Subdistrict

Age and sex were not retained as they were not statistically significant and reduced model stability. Ethnicity was excluded due to high collinearity with citizenship status, resulting in MCMC convergence instability. Nodemix terms were evaluated but did not improve model stability; level-specific nodematch terms were retained as the primary specification. Population density proxy and subdistrict were unavailable at the residential unit level in the e-Dengue dataset. ERGM, exponential random graph model; MCMC, Markov chain Monte Carlo.



Supplementary Figure S1 Goodness-of-fit for the final ERGM based on degree distribution. Observed degree counts are plotted against the mean and interquartile range from simulated networks. Because the analytic network excluded isolates, degree-0 nodes appear only in the simulated distribution. Fit should be assessed from degree 1 upwards.



Supplementary Figure S2 MCMC trace and density plots for sample statistics in the final ERGM. MCMC trace plots (left panels) show the sampled values of each network statistic across iterations; stable horizontal mixing without drift indicates satisfactory chain convergence. Density plots (right panels) show the empirical distribution of sampled statistics; unimodal, approximately symmetric distributions indicate adequate mixing. Trace and density plots are presented for all terms in the final ERGM: edges, geometrically weighted degree (gwdegree, decay = 0.25), nodefactor for occupation (student, unemployed), nodematch for citizenship (Malaysian citizen, non-citizen), and nodematch for housing type (high-rise, low-rise). MCMC burn-in of 500,000 iterations was applied before sampling; sample size was 20,000 with a thinning interval of 4,000. ERGM, exponential random graph model; MCMC, Markov chain Monte Carlo.