Socially Connected Neighborhoods and the Spread of Sexually Transmitted Infections

Lauren Newmyer, Megan Evans, and Corina Graif

List of Appendices

Appendix A: Map of Chicago

Figure A.1 shows the 77 communities in Chicago that our study assesses. This figure represents these communities geographically as well as showing the names of each. The map was created by authors using data from the Chicago data portal of Tiger shape files of community boundaries.

Figure A.2 maps the prevalence of STIs across Chicago neighborhoods for three selected years during our study, 2002, 2007, and 2014. Neighborhoods are color coded based on the yearly quartile distribution of STIs. Darker gradients of green indicate a lower prevalence of STIs, meaning healthier communities; in contrast, the lighter green communities have a higher prevalence of STIs. These maps demonstrate visible geographical clustering of STIs. While some small differences emerge in the central and northeastern communities of Chicago, the trends in STIs appear relatively stable across space and time.

Figure A.1 Map of Chicago Communities

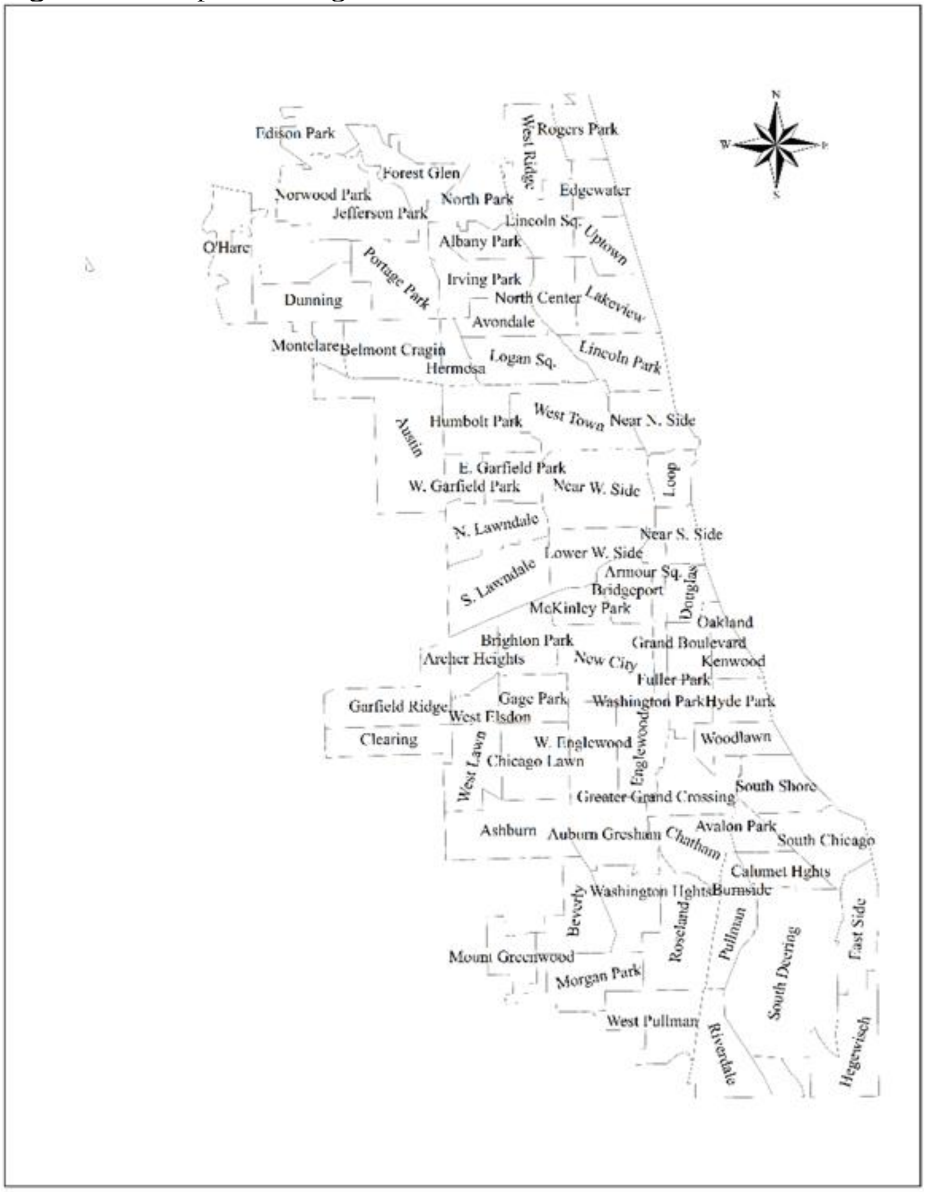
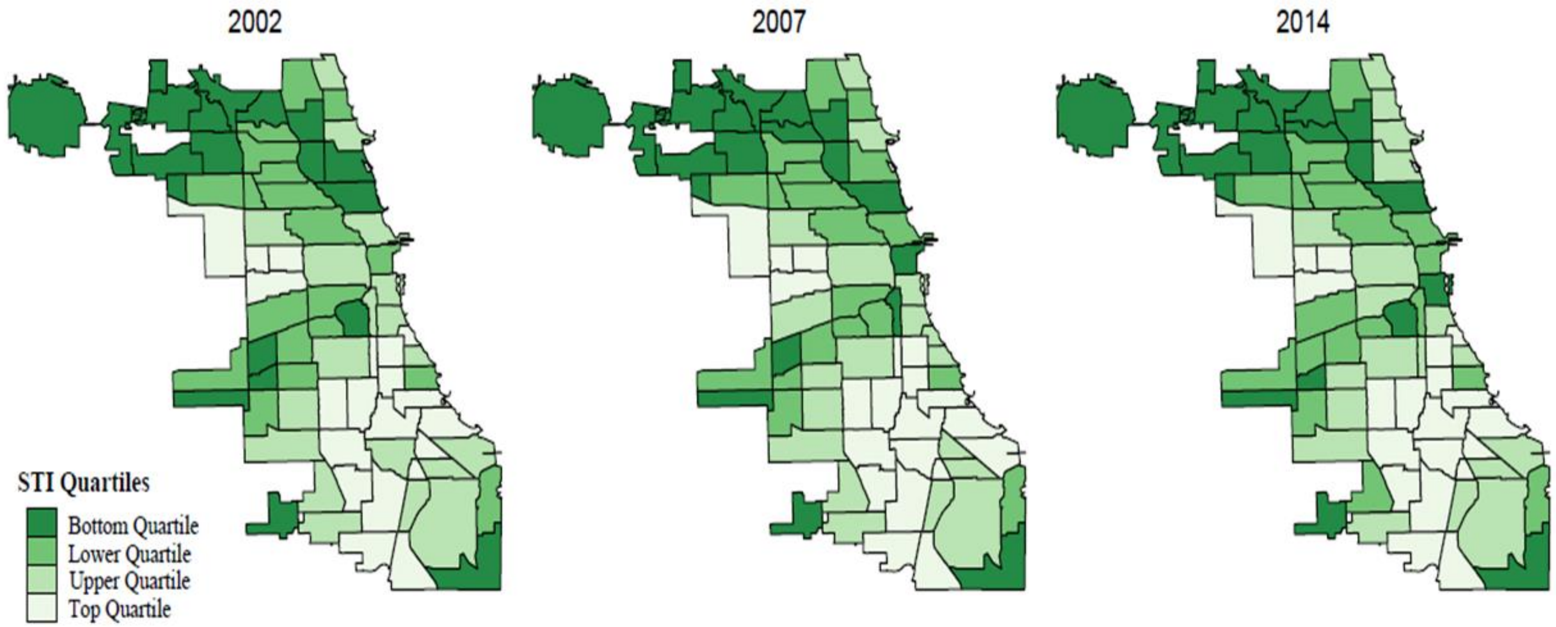


Figure A.2 Map of STI Rates in Chicago across Time



Appendix B: Descriptive Statistics

Table B.1 presents the descriptive statistics of the commuting networks over time. Overall, the commuting networks do not differ substantially over time regarding indegree, outdegree, or density. Across 2002-2014, the average total number of ties (average indegree and outdegree) between neighborhoods is 7.30. In this period, the average minimum number of ties a work community receives is 2.69 ties and the maximum is 14.31 (the measure of indegrees). Regarding how many ties a residential neighborhood may send to a work neighborhood (the measure of outdegree), these measures range from 0 up to 76. On average the measure of density is .10 from 2002-2014, with no year varying widely. This finding means that out of all possible ties that could be formed in this network on average 10% of them are present.

Table B.1 Descriptive Statistics for Commuting Network

	2002-2014	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Average Indegree/Outdegree	7.30	8.30	7.57	7.56	8.10	7.26	6.87	7.87	6.88	7.01	7.16	6.55	6.82	6.90
Min Indegree	2.69	2	2	2	2	2	3	4	3	3	3	3	3	3
Max Indegree	14.31	14	14	13	14	13	15	13	18	14	13	14	15	16
Min Outdegree	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Max Outdegree	76	76	76	76	76	76	76	76	76	76	76	76	76	76
Density	0.10	0.11	0.10	0.10	0.11	0.10	0.09	0.10	0.09	0.09	0.09	0.09	0.09	0.09

Appendix C: TERGM Results and Goodness-of-Fit Graphs for TERGMS

We estimate TERGMs to better understand how communities are tied through commutes and the implications they have for disease spillovers. Additionally, these models allow us to better assess the role of selection in infection patterns. ERGMs allow for the statistical modeling of the patterns of relationships in a network using node-level characteristics (such as a community's STI rate) and graph-level indices (such as reciprocity (if commuters commute to one neighborhood does that neighborhood also send commuters to theirs) or transitivity (i.e., a friend of my friend is also my friend) (Frank & Strauss, 1986; Robins et al., 2007). Through a simulation procedure, an ERGM evaluates whether network patterns are significantly different than what would be expected to occur by chance. Modeling both node-level characteristics and graph-level indices is advantageous because it allows a researcher to assess what is uniquely shaping a network structure. By using Markov chain Monte Carlo (MCMC) simulations, the ERGM estimates the true likelihood function (Hunter, 2007). The results represent the log likelihood that a network statistic is more or less likely to occur than by chance (Leifeld et al., 2018; Robins et al., 2007). Positive and significant coefficients indicate that a network structure is more likely to occur than by random chance, while negative and significant coefficients indicate that a network structure is less likely to occur. Although computationally different, ERGMs are often compared to logistic regression because the dependent variable is binary (1=tie; 0=no tie) and the coefficients are estimated in log odds. We use a longitudinal extension of the ERGM, the TERGM to assess the network structures which predict our commuting network over time, rather than cross-sectionally. We account for inter-temporal dependence as we predict the existence of commuting ties across a thirteen-year period. Accounting for inter-temporal dependence is important as the work communities which home communities are tied to through commuting remain relatively stable across time.

Both network and spatial data in general are inherently interdependent which violates typical regression assumptions. In networks, nodes are connected to each other through a tie, which causes them to have correlated error terms. ERGMs overcome the limitation of interdependence by explicitly modeling the structural dependencies of the data. We account for endogenous effects in our model by accounting for network density, reciprocity, popularity spread, transitivity (closed triads), and the preconditions for transitivity (open triads). All these features are commonly found in complex directed networks (Wasserman & Faust, 1994). Network density indicates the number of edges in the network; reciprocity indicates the propensity for actors to have mutual ties with one another; popularity spread indicates the tendency for some communities to be more popular and central in the network than others; and open and closed triangles indicate the propensity for transitivity to occur in networks (Hunter, 2007; Snijders et al., 2006).

Our TERGMs predict the presence of commuting ties in our network of 77 communities. We create an asymmetric, binary inter-neighborhood commuting network by calculating whether a commuting tie exists between two communities. LEHD's LODES data provides the origin-destination links for block groups, representing the number of persons from each home community commuting to each work community. We aggregate these data to community areas and normalize the numbers to the home community's 2000 total population. From this 77 by 77 valued matrix, we create an asymmetric, binary inter-neighborhood commuting matrix. We define a commuting tie as existing between two communities if at least 0.5% of a community's population commutes to a work community. If less than 0.5% of a community's residents commute to a work community then we consider a commuting tie as nonexistent between them. We create an inter-neighborhood commuting network for each of the years in our study, i.e., 2002 to 2014.

To estimate our TERGMs we use maximum pseudolikelihood with bootstrapped (bootstrapping sample size=1000) confidence intervals. We use the *xergm* package in R (Leifeld et al., 2018). The TERGM results are comparable to ERGM results but present the average effect of the network statistics across the thirteen-year period (Leifeld et al., 2018). In addition to the endogenous effects discussed above, we account for two temporal network statistics: a linear time trend and memory. We control for the trend for networks to become denser (a linear time covariate) and maintain stable dyadic ties over time (memory). The TERGM model is represented by the following equation,

$$\mathrm{P}(N^{\mathrm{K}+1}, \dots, N^{\mathrm{T}} | N^1, \dots, N^{\mathrm{K}}, \theta) = \prod_{t=K+1}^{T} \frac{\exp\,(\theta^{\intercal} h(N^t, N^{t-1}, \dots, N^{t-K}))}{c(\theta, N^{t-K}, \dots, N^{t-1})}$$

where we are predicting the probability of observing the networks N between times K+1 and T by taking the product of the probabilities of the individual networks conditional on the others, where N is the adjacency matrix of our commuting network in which $N_{ij} = 1$ if community i sends a commuting tie to community j and 0 if community i does not send a commuting tie to community j. The vector of model coefficients is represented by θ, $h(N)$ represents the vector of statistics accounting for endogenous and exogenous network dependencies, and $c(\theta) = \sum_{i=1}^{\mathcal{N}} \exp(\theta^{\intercal}(N_i))$ represents the set of all possible permutations of the network given the same number of nodes, $\mathcal{N}$, where $\theta^{\intercal}$ represents the transpose of vector θ.

We chose a TERGM over other approaches like SIENA in part because the nodal unit of analysis in this study is a community area. SIENA is often used to model individual actors as nodal units of analyses and assumes frequent change from one time period to another, constraints that are less of an issue with TERGM (Block, Stadtfeld, & Snijders, 2016; Desmarais & Cranmer, 2012). The TERGM approach that we use comes with several limitations as well, such as it functions as a pooling approach rather than allowing a formal causal distinction between processes leading to the formation vs. dissolutions of ties (Leifeld, Cranmer, & Desmarais, 2018).

Table C.1 presents the results of our TERGMs which estimate selection factors that contribute to commuting ties between neighborhoods. These models estimate effects of different characteristics of sending (home origin) communities and of receiving (work destination) communities. We estimate both receiver and sender effects for neighborhood measures of the STI rate, controlling for socioeconomic and demographic community-level factors like disadvantage, residential stability, racial and ethnic diversity, and the density of local workers. Model 1 includes all community-level predictors for estimation. Due to the high correlation between disadvantage and STI rates, we also estimate Model 2 to assess the robustness of our results when disadvantage is removed. We also examine whether the inclusion of STI covariates in Model 1 can explain in part the disadvantage coefficient in Model 3. We see few differences when excluding disadvantage or STI covariates. Model 1 coefficients will be discussed here in detail, as Models 2 and 3 yield largely similar patterns in results.

We find that many of the community-level variables are significant predictors of pairwise ties. A receiver effect refers to a covariate estimate that predicts whether a neighborhood is commuted to for work. A sender effect is similar but refers to the home community; this is the neighborhood that a commuter is residentially located in, and thus commutes from. Increasing residential stability in a work community decreases the odds that a commuter will travel to that neighborhood for work. Accordingly, increasing residential stability in a home community increases the likelihood that a neighborhood will send commutes out. Neighborhoods with high residential stability are less likely to be commuted to for work and residents are more likely to commute for work from these communities to other areas. The density of local workers in a community is both a positive significant predictor for both receiving and sending communities. A community with a high density of local workers both attracts other commuters and sends out its own commuters. Importantly, we find no significance of STI rates, socioeconomic disadvantage, and racial and ethnic diversity for receiver or sender communities.

We also use measures of dissimilarity to determine how homophily drives patterns of commuting ties between neighborhoods. Neighborhoods with dissimilar rates of residential stability are less likely to be connected. This finding means that commuters often work in areas of similar residential stability as their residential neighborhood. Neighborhoods with similar residential stability are more likely to be tied by their residents' commutes. We find the importance of dissimilarity in racial and ethnic diversity for ties in our network. Commuters are significantly more likely to commute to areas with different levels of racial and ethnic diversity than their home community. We find that STIs play a role in the shaping of commuting ties. Each increase in dissimilarity of STI rates decreases the odds that communities are tied through commutes by $((e^{.26}-1)x100))$ 29.69% on average. Commuters tend to work in environments that have similar rates of STIs as their residential neighborhoods. Lastly, we find no significance of dissimilarity for socioeconomic disadvantage and the density of local workers for ties.

Model 1 results indicate that network structures are significant contributors to commuting ties between neighborhoods. There is a significantly lower likelihood of tie formation in this commuting network, which is shown by a negative edge term. This result is common as real-world ties are rarely randomly distributed and are expected to occur at lower frequency than the maximum possible. We also find evidence of clustering and transitivity in our commuting network. The network structure of geometrically weighted edgewise-shared partners is both significant and positive, meaning that neighborhoods are likely tied together in clusters and there is a propensity for transitivity in the Chicago commuting network. Regarding popularity spread, this term is negative indicating that communities with many receiving ties are less likely to receive additional commuting ties on average. Lastly, we find no significant effect for reciprocity. Commuters are not more likely to work in communities that send residents to work in their home community. To assess the importance of spatial features and temporal effects, we include multiple measures. Spatial proximity is a positive and significant predictor of commuting ties over time. This finding indicates, as expected, that commuters have higher odds of commuting to neighborhoods that are geographically contiguous to them. Additionally, we also find that whether neighborhoods are accessible to each other via shared public transit lines significantly and positively predicts commuting ties. Regarding temporal effects, we find that whether a tie existed in the previous year between communities significantly predicts the tie being present in following time periods. Commuting ties once formed remain generally stable.

To assess the goodness-of-fit for our TERGMs we simulate 1,000 networks generated at random from the specified coefficients in Table C.1. We then graph these simulations to visualize how close our simulated networks match the real commuting network (Figures C.1, C.2, and C.3). The black line shows how the values from our simulated networks align with the values in our real commuting network. We find that our simulated networks match closely to the real network regarding edgewise shared partners, geodesic distance, and degree for all three models in Table C.1.

Table C.1 Bootstrapped Temporal Exponential Random Graph Models Predicting Commuting Ties, 2002-2014

	Model 1	Model 2	Model 3
Network Structure			
Edges	-3.008* (-3.339, -2.640)	-3.077* (-3.448, -2.631)	-3.181* (-3.520, -2.761)
Reciprocity	-.153 (-.500, .143)	-.110 (-.467, .172)	-.162 (-.516, .132)
Geometrical Weighted In-Degree (Popularity Spread)	-2.155* (-2.552, -1.775)	-2.143* (-2.559, -1.748)	-2.087* (-2.524, -1.696)
Geometrically Weighted Edgewise Shared Partner (Closed Triads)	1.006* (.765, 1.237)	.998* (.783, 1.209)	.999* (.769, 1.217)
Geometrically Weighted Dyadic Shared Partner (Open Triads)	-.079* (-.095, -.063)	-.078* (-.094, -.064)	-.078* (-.094, -.064)
Receiver Effects (Work Community)			
Sexually Transmitted Infections Rate	.165 (-.104, .434)	.115 (-.117, .316)	
Community Disadvantage	-.056 (-.428, .283)		.003 (-.255, .247)
Residential Stability	-.426* (-.600, -.277)	-.422* (-.599, -.278)	-.449* (-.624, -.297)
Racial and Ethnic Diversity	-.137 (-.274, .003)	-.138 (-.257, .001)	-.163* (-.269, -.044)
Density of Local Workers	.240* (.178, .303)	.252* (.199, .308)	.243* (.186, .302)
Sender Effects (Home Community)			
Sexually Transmitted Infections Rate	.209 (-.020, .397)	.072 (-.101, .208)	
Community Disadvantage	-.196 (-.359, .007)		-.096 (-.233, .039)
Residential Stability	.341* (.173, .529)	.372* (.189, .573)	.332* (.166, .516)
Racial and Ethnic Diversity	.038 (-.058, .114)	.052 (-.019, .114)	.009 (-.070, .082)
Density of Local Workers	.097* (.031, .158)	.112* (.049, .170)	.099* (.038, .161)
Dissimilarity			
Sexually Transmitted Infections Rate	-.260* (-.360, -.131)	-.249* (-.373, -.104)	
Community Disadvantage	.010 (-.105, .126)		-.117 (-.253, .018)
Residential Stability	-.268* (-.401, -.151)	-.249* (-.409, -.135)	-.262* (-.416, -.147)
Racial and Ethnic Diversity	.088* (.009, .194)	.096* (.011, .215)	.070* (.004, .164)
Density of Local Jobs	-.053 (-.114, .013)	-.064 (-.118, .004)	-.056* (-.114, -.001)
Spatial Effects			
Spatial Proximity	1.217* (.789, 1.682)	1.205* (.775, 1.680)	1.221* (.781, 1.720)

Transportation	.733* (.636, .844)	.740* (.639, .854)	.766* (.658, .892)
Time Effects			
Linear Time Trend	.013 (-.039, .054)	.014 (-.040, .058)	.017 (-.034, .060)
Memory	2.131* (1.957, 2.373)	2.137* (1.960, 2.359)	2.137* (1.970, 2.365)
N= 77 Nodes and 77,077 Possible Edges			

Notes: *Denotes that confidence interval does not overlap with zero; coefficients with 95% confidence intervals in parenthesis

Figure C.1 Goodness-of-Fit Statistics for TERGM Model 1

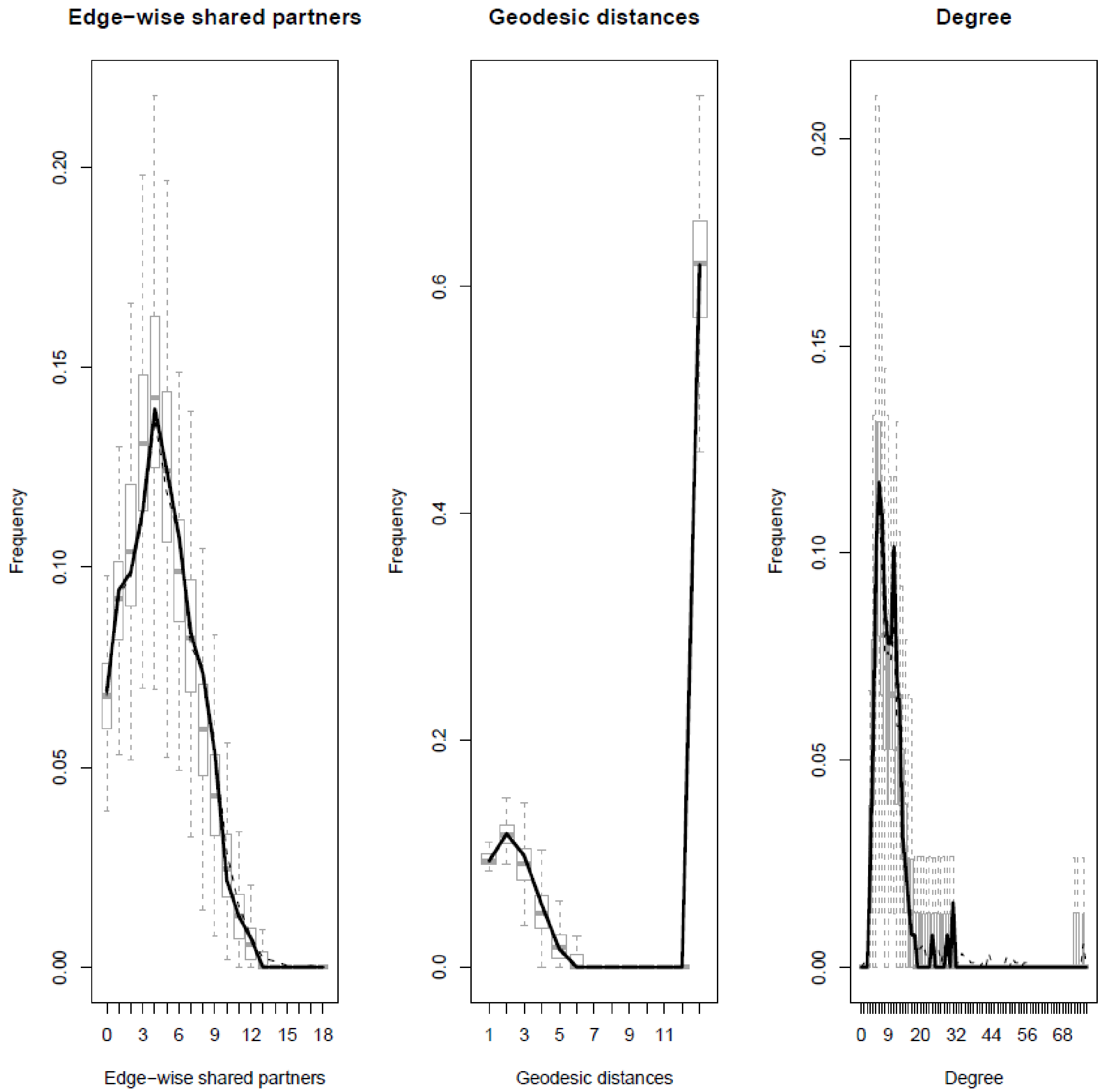


Figure C.2 Goodness-of-Fit Statistics for TERGM Model 2

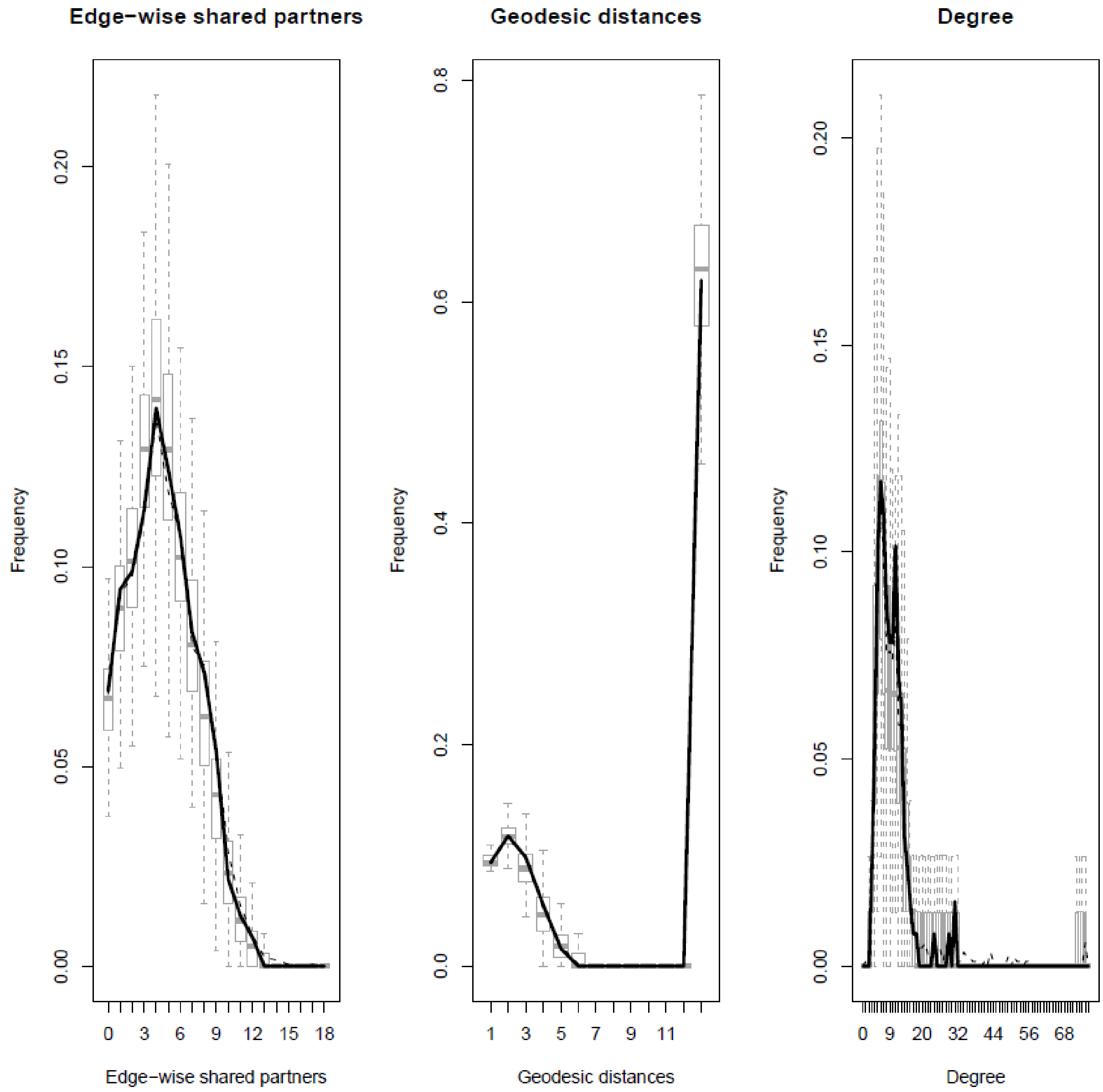
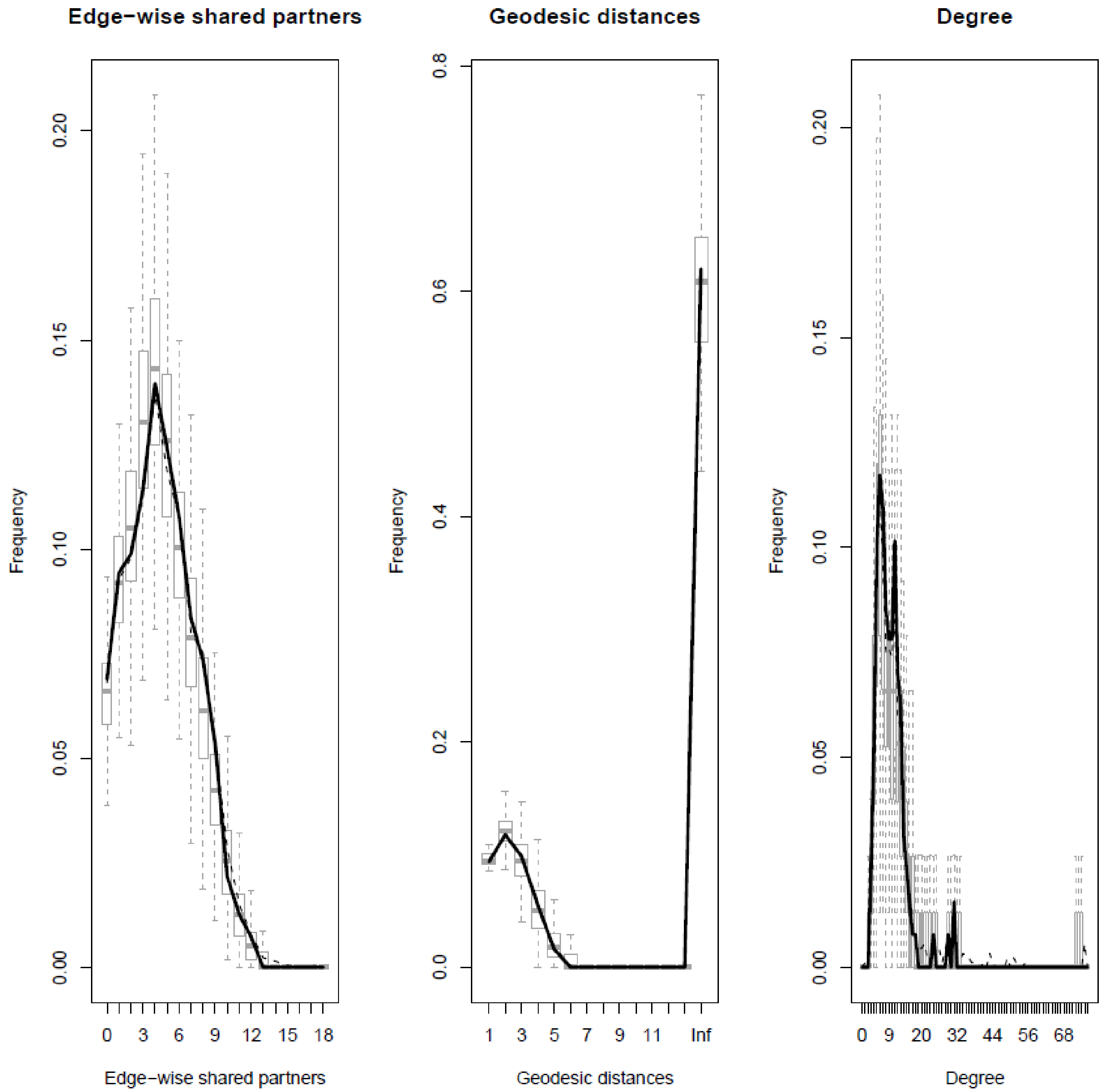


Figure C.3 Goodness-of-Fit Statistics for TERGM Model 3



Appendix D: Supplementary Analyses for Combined Matrices

We combine matrices to assess how multiple connections across Chicago contribute to STI rates in neighborhoods. Our four additional matrices include neighboring communities that are 1) geographically contiguous or connected through commuting ties; 2) geographically contiguous or connected through public transit; 3) connected through public transit or commuting ties; and 4) geographically contiguous, or connected through commuting ties, or public transit ties. We incorporate these four new definitions of networks as spatial weight matrices in our fixed effects spatial autoregressive models. In Table D.1, we show how different combinations of the matrices of geographic contiguity, mass transit, and commuting shape the effects for spatial network lag of STI rates. Using measures of AIC and BIC to assess model fit, we find that the model which accounts for all three of these networks has the best model fit. Communities that have multiple connections are most at risk of exposure to STI spread from the communities they are connected.

Table D.1 Spatial Lag and Error Models Predicting STI Rates with Combined Matrices

	Network STI Risk	AIC	BIC
Geographic Contiguity Network	.60*** (.06)	-921.79	-823.62
Public Transit Network	.88*** (.01)	-3391.84	-3293.66
Commuting Network	.96*** (.01)	-1268.55	-1170.38
Contiguity and Commuting Network	3.52*** (.15)	-6617.41	-6529.05
Contiguity and Transit Network	6.09*** (.42)	-7594.63	-7496.45
Transit and Commuting Network	3.35*** (.09)	-8337.62	-8244.35
Contiguity, Transit, and Commuting Network	3.08*** (.10)	-8412.05	-8372.78

Notes: Models are estimated including all covariates from Table 4; Coefficients with standard errors in parentheses
*p<.05, **p<.01,*** p<.001

Appendix E: Supplementary Analyses for Relevant Demographic Variables

STI rates are more prevalent among lower socioeconomic statuses and racial and ethnic minorities. These demographic patterns likely link to residential patterning of STIs. Further, a neighborhood's age structure, marriage rates, and average household size are relevant for that neighborhood's overall risk of contracting STIs and may influence access to public transit and commuting patterns.

In Tables E.1 through E.3, we predict our fixed effects spatial and network autoregressive models using more detailed information on the racial composition of the community, and the neighborhood's age structure, marriage rates, and average household size. We also consider how the teen birth rate, the age-adjusted total fertility rate, total population logged, and population density could influence neighborhood STIs. By controlling for these variables, we can examine the robustness of our results in determining how spatial proximity, commuting, and transportation ties influence STI rates. Table E.1 presents the results with our geographic contiguity network; Table E.2 presents the results with our public transit network, and Table E.3 presents the results with our commuting network.

Our results remain robust to these additional controls. Additionally, we find that several of them are integral in predicting STI rates. We consistently find that communities which experience an increase of Black residents also experience an increase in the prevalence of STIs. This finding is consistent with prior findings of a higher concentration of STIs among the Black population (Adimora & Schoenbach, 2005, 2013; Harling et al., 2014; Thomas & Thomas, 1999).

Across all three spatial weight specifications, we find that a high teen birth rate and a high age-adjusted total fertility rate increases a neighborhood's STI prevalence. Additionally, we find that a high average household size increases STI prevalence, while a high percentage of married residents decreases STI prevalence. We also find the percentage of the population age most at risk, i.e., those aged 15 to 25, significant. Contrary to expectations, a high percentage of the population at risk of STI contraction decreases STI prevalence. Although the main results remain robust to all these additional controls, high correlations between many of the added variables and thus, the increased multicollinearity risk, make these models less ideal for inclusion in the main tables. Additionally, the models retained in the main tables presented consistently better fit scores: for the spatial network (Table 2 in text AIC: -915.45 and Table E.1 AIC: -891), public transit network (Table 2 in text AIC: -3391.84 and Table E.2 AIC: -3378), and the commuting network (Table 2 in text AIC: -1269.08 and Table E.3 AIC: -1176). The models in the main tables also control for prior STI rates, which inherently absorbs the effects of other prior STI determinants from the longer models.

Table E.1 Spatial Lag and Error Models with Geographic Contiguity Network Predicting STI Rates with Relevant Demographic Variables, 2002-2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Concentrated Disadvantage	.014	-.070*	.015	.037	-.044	.005	.026	.014	.014	-.037	
	(.029)	(.033)	(.029)	(.029)	(.024)	(.029)	(.024)	(.029)	(.030)	(.035)	
Residential Stability	-.030	-.050	-.004	-.049	-.049*	.002	-.038	.003	-.030	-.109*	-.113**
	(.036)	(.038)	(.036)	(.035)	(.024)	(.037)	(.023)	(.038)	(.036)	(.044)	(.042)
Diversity	-.072*		-.056	-.074*	-.041	-.059	-.033	-.076*	-.072*		
	(.036)		(.036)	(.035)	(.030)	(.036)	(.030)	(.036)	(.036)		
Local Worker Density	.007	-.002	.019	.013	.017	.014	.018	.028	.007	.045	.043
	(.032)	(.032)	(.032)	(.032)	(.035)	(.032)	(.035)	(.033)	(.032)	(.034)	(.032)
Percentage White		-.002								.003	.002
		(.006)								(.006)	(.006)
Percentage Black		.015**								.015*	.013*
		(.005)								(.006)	(.005)
Percentage Hispanic		.001								.007	.007
		(.006)								(.006)	(.006)
Teen Birth Rate			.004***							.003**	
			(.001)							(.001)	
Age-Adjusted Total Fertility Rate				.0002***						.0002***	.0002***
				(.00003)						(.00003)	(.00003)
Average Household Size					.134**					.128	.128
					(.049)					(.075)	(.071)
Population Married						-.009**				-.011**	-.010**
						(.003)				(.004)	(.003)
Population Aged 15 to 25							-.007**			-.038***	-.035***
							(.003)			(.007)	(.007)
Total Population Logged								-.238**		.047	

								(.092)		(.116)	
Population Density									3.79e-10 (1.06e-07)	1.51e-07 (1.07e-07)	1.32e-07 (1.04e-07)
Network STI Risk	- .777***	- .772***	- .785***	-.735***	.759***	- .772***	.751***	- .778***	-.777***	-.740***	-.722***
	(.085)	(.083)	(.084)	(.087)	(.042)	(.085)	(.044)	(.084)	(.085)	(.083)	(.085)
Error Variance Parameter	.761***	.763***	.768***	.757***	- .787***	.76***	-.77***	.764***	.761***	.759***	.750***
	(.042)	(.041)	(.041)	(.043)	(.085)	(.042)	(.087)	(.041)	(.042)	(.042)	(.044)
Constant	.138***	.136***	.136***	.136***	.137***	.137***	.138***	.137***	.138***	.131***	.132***
	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)
N	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001
AIC	-808	-834	-826	-847	-812	-814	-812	-813	-806	-894	-891
BIC	-715	-731	-728	-748	-714	-716	-714	-715	-708	-757	-769

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<0.05, **p<0.01,*** p<0.001 (two-tailed)

Table E.2 Spatial Lag and Error Models with Public Transit Network Predicting STI Rates with Relevant Demographic Variables, 2002-2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Concentrated Disadvantage	.008	-.114**	.041	.065*	-.033	-.022	.034	-.006	.004	-.020	
	(.030)	(.036)	(.031)	(.031)	(.033)	(.031)	(.034)	(.030)	(.030)	(.041)	
Residential Stability	-.051	-.100*	-.042	-.075*	-.091*	-.008	-.070	-.007	-.056	-.112*	-.131**
	(.037)	(.040)	(.037)	(.036)	(.039)	(.039)	(.038)	(.040)	(.038)	(.047)	(.045)
Diversity	-.128**		-.138***	-.143***	-.123**	-.104*	-.119**	-.124**	-.122**		
	(.041)		(.041)	(.040)	(.041)	(.041)	(.041)	(.041)	(.042)		
Local Worker Density	.020	.033	.054	.037	.034	.026	.019	.049	.021	.112**	.073
	(.042)	(.041)	(.042)	(.041)	(.042)	(.041)	(.042)	(.043)	(.042)	(.043)	(.040)
Percentage White		.001								.003	.006
		(.006)								(.007)	(.006)
Percentage Black		.020***								.012	.017***
		(.005)								(.006)	(.005)
Percentage Hispanic		.007								.008	.010
		(.006)								(.006)	(.006)
Teen Birth Rate			.005***							.002*	
			(.001)							(.001)	
Age-Adjusted Total Fertility Rate				.0002***						.0003***	.0003***
				(.00004)						(.00004)	(.00004)
Average Household Size					.204**					.143	.099
					(.069)					(.084)	(.078)
Population Married						-.014***				-.017***	-.017***
						(.004)				(.004)	(.004)
Population Aged 15 to 25							-.006			-.031***	-.030***
							(.003)			(.007)	(.007)
Total Population Logged								-.263**		-.195	

								(.093)		(.126)	
Population Density									1.13e-07 (1.33e-07)	2.09e-07 (1.28e-07)	1.74e-07 (1.24e-07)
Network STI Risk	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)
Error Variance Parameter	.883*** (.006)	.884*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.883*** (.006)	.884*** (.006)	.884*** (.006)
Constant	.166*** (.004)	.163*** (.004)	.164*** (.004)	.162*** (.004)	.165*** (.004)	.165*** (.004)	.166*** (.004)	.165*** (.004)	.166*** (.004)	.156*** (.004)	.156*** (.004)
N	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001
AIC	-3277	-3309	-3295	-3319	-3284	-3287	-3278	-3283	-3276	-3380	-3378
BIC	-3184	-3205	-3196	-3221	-3185	-3189	-3180	-3185	-3177	-3243	-3255

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<0.05, **p<0.01,*** p<0.001 (two-tailed)

Table E.3 Spatial Lag and Error Models with Commuting Network Predicting STI Rates with Relevant Demographic Variables, 2002-2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Concentrated Disadvantage	.033	-.099**	.058	.056	-.003	.025	.056	.004	.034	-.015	
	(.032)	(.037)	(.032)	(.032)	(.034)	(.032)	(.036)	(.033)	(.033)	(.040)	
Residential Stability	-.031	-.053	-.033	-.072*	-.068	.011	-.046	.020	-.030	-.076	-.082
	(.037)	(.040)	(.036)	(.036)	(.038)	(.040)	(.038)	(.038)	(.037)	(.047)	(.046)
Diversity	-.087*		-.108*	-.113**	-.078	-.065	-.085*	-.087*	-.088*		
	(.043)		(.043)	(.042)	(.043)	(.044)	(.043)	(.043)	(.043)		
Local Worker Density	.038	.039	.058	.034	.048	.048	.039	.076*	.039	.083*	.062
	(.037)	(.036)	(.036)	(.036)	(.037)	(.037)	(.037)	(.038)	(.037)	(.038)	(.035)
Percentage White		-.001								.008	.008
		(.006)								(.007)	(.007)
Percentage Black		.017***								.016**	.017***
		(.005)								(.006)	(.005)
Percentage Hispanic		.001								.007	.007
		(.006)								(.007)	(.007)
Teen Birth Rate			.006***							.003*	
			(.001)							(.001)	
Age-Adjusted Total Fertility Rate				.0002***						.0003***	.0003***
				(.00004)						(.00004)	(.00004)
Average Household Size					.217**					.177*	.162*
					(.070)					(.082)	(.078)
Population Married						-.010**				-.011**	-.010**
						(.004)				(.004)	(.004)
Population Aged 15 to 25							-.005			-.023**	-.023**
							(.003)			(.008)	(.008)
Total Population Logged								-.314***		-.069	
								(.078)		(.114)	

Population Density									-2.99e-08 (1.28e-07)	9.96e-08 (1.28e-07)	8.01e-08 (1.24e-07)
Network STI Risk	.964*** (.010)	.965*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)	.964*** (.010)
Error Variance Parameter	.964*** (.010)	.965*** (.009)	.965*** (.010)	.965*** (.010)	.965*** (.010)	.964*** (.010)	.964*** (.010)	.965*** (.009)	.964*** (.010)	.966*** (.009)	.966*** (.009)
Constant	.165*** (.004)	.161*** (.004)	.162*** (.004)	.161*** (.004)	.164*** (.004)	.164*** (.004)	.165*** (.004)	.163*** (.004)	.165*** (.004)	.155*** (.004)	.156*** (.004)
N	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001
AIC	-1083	-1123	-1106	-1125	-1090	-1088	-1083	-1096	-1081	-1178	-1176
BIC	-989	-1020	-1008	-1027	-992	-990	-985	-998	-982	-1040	-1053

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at $p<.001$. *$p<0.05$, **$p<0.01$,*** $p<0.001$ (two-tailed)

Appendix F: Supplementary Analyses Examining STI Prevalence by Disease and Gender

As STI rates vary by disease and gender, these variations might map onto communities differently. Tables E.1 through E.4 predict the four STI rates which make up our composite STI prevalence measure separately – gonorrhea rates for males aged 15 to 44, gonorrhea rates for females aged 15 to 44, chlamydia rates for males aged 15 to 44, and chlamydia rates for females aged 15 to 44. Our substantive conclusions from Table 1, in main text, remain consistent. Interestingly, our community sociodemographic measures differently predict the varying STI rates. We find that residential stability serves to increase the prevalence of gonorrhea for males and females, while it decreases chlamydia rates among females.

Table F.1 Spatial Lag and Error Models Predicting Gonorrhea Rates for Males Aged 15 to 44, 2002-2014

	Without Network	Spatial Network	Commuting Network	Public Transit Network
Prior Gonorrhea Rate	.372***	.245***	.299***	.329***
	(.0289)	(.027)	(.046)	(.031)
Concentrated Disadvantage	-19.6	-37.7	64.7	-38.4
	(51)	(38.3)	(56)	(51.8)
Residential Stability	455***	232***	193*	394***
	(58.8)	(47.2)	(84.5)	(65.9)
Diversity	35.6	-24.3	232*	3.86
	(68.2)	(54.9)	(103)	(70.3)
Local Worker Density	-120	-82.9	-2.12	-146*
	(69)	(62.7)	(59.3)	(71.5)
Constant	271***	243***	421***	284***
	(6.31)	(6.47)	(14.4)	(6.62)
Network STI Risk		.639***	5.35***	.883***
		(.046)	(.411)	(.006)
Error Variance Parameter		-.542***	3.5***	.883***
		(.091)	(.226)	(.006)
N	1001	1001	1001	1001
AIC	13014	12925	10766	10483
BIC	13102	13023	10864	10582

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<.05, **p<.01,*** p<.001

Table F.2 Spatial Lag and Error Models Predicting Gonorrhea Rates for Females Aged 15 to 44, 2002-2014

	Without Network	Spatial Network	Commuting Network	Public Transit Network
Prior Gonorrhea Rate	.237***	.154***	.092*	.197***
	(.031)	(.025)	(.044)	(.033)
Concentrated Disadvantage	15.4	-9.1	128***	-11.8
	(39.9)	(28)	(38.5)	(41.2)
Residential Stability	186***	70.9*	12.7	118*
	(43.4)	(30.8)	(53.7)	(49.9)
Diversity	-39.3	-22.8	266***	-83.6
	(52.8)	(40.2)	(78.3)	(55.4)
Local Worker Density	-74.2	-65.2	-78.6*	-60.3
	(53.2)	(47)	(39.7)	(56)
Constant	210***	184***	356***	224***
	(4.89)	(5.27)	(12.7)	(5.2)
Network STI Risk		.73***	4.55***	.883***
		(.044)	(.28)	(.006)
Error Variance Parameter		-.683***	4.94***	.883***
		(.091)	(.31)	(.006)
N	1001	1001	1001	1001
AIC	12540	12468	10291	10039
BIC	12629	12566	10390	10137

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<0.05, **p<0.01,*** p<0.001 (two-tailed)

Table F.3 Spatial Lag and Error Models Predicting Chlamydia Rates for Males Aged 15 to 44, 2002-2014

	Without Network	Spatial Network	Commuting Network	Public Transit Network
Prior Chlamydia Rate	.218***	.126***	.202***	.215***
	(.032)	(.024)	(.0327)	(.034)
Concentrated Disadvantage	-90.7	-48.2	-14.7	-54.3
	(58.7)	(41.2)	(61.5)	(58.2)
Residential Stability	-37	14	35.8	-18
	(64.2)	(43.4)	(69.3)	(71.1)
Diversity	-69.4	-46.1	-62.5	-77.9
	(79.1)	(59.3)	(82)	(79.3)
Local Worker Density	88.6	108	194**	51.2
	(79.9)	(67.7)	(69.9)	(80.5)
Constant	315***	265***	312***	321***
	(7.32)	(7.38)	(7.27)	(7.46)
Network STI Risk		.78***	.966***	.883***
		(.039)	(.009)	(.006)
Error Variance Parameter		-.558***	.966***	.883***
		(.091)	(.009)	(.006)
N	1001	1001	1001	1001
AIC	13288	13136	12867	10705
BIC	13377	13234	12965	10803

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<.05, **p<.01,*** p<.001

Table F.4 Spatial Lag and Error Models Predicting Chlamydia Rates for Females Aged 15 to 44, 2002-2014

	Without Network	Spatial Network	Commuting Network	Public Transit Network
Prior Chlamydia Rate	.384***	.316***	.375***	.394***
	(.032)	(.031)	(.033)	(.033)
Concentrated Disadvantage	73.8	21.8	16.2	74
	(79.8)	(67.3)	(86)	(81.8)
Residential Stability	-423***	-243**	-403***	-409***
	(90.9)	(80.5)	(99.6)	(102)
Diversity	-203	-94.7	-195	-262*
	(106)	(96.8)	(114)	(110)
Local Worker Density	38.5	-12.4	3.3	90.9
	(107)	(103)	(96)	(111)
Constant	421***	398***	431***	443***
	(9.8)	(10.8)	(10)	(10.3)
Network STI Risk		.522***	.965***	.883***
		(.074)	(.010)	(.006)
Error Variance Parameter		-.393**	.966***	.884***
		(.12)	(.009)	(.00)
N	1001	1001	1001	1001
AIC	13826	13791	13460	11301
BIC	13915	13889	13558	11399

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001.
*p<.05, **p<.01,*** p<.001

Appendix G: Supplementary Analyses Examining STI Prevalence by Varying Commuting Thresholds

We define a commuting tie using a 0.5% cutoff for several reasons. Community areas in Chicago, on average, have a population of 35,000 residents. A 0.5% cutoff represents, on average, 175 residents (0.005 X 35,000). It is important for us to define a commuting tie existing between two communities based on a sufficiently large enough sample to avoid noise and have a substantively meaningful number of residents commuting to the work community. When selecting smaller thresholds, most of the communities become connected to one another, making it more difficult to disentangle meaningful connections from noisy data. When selecting higher thresholds, only a few hub work communities remain connected to the other home communities, leading to possible concerns about a small number of outliers driving the results. Still, we found it valuable to explore analyses which examine the results using varying cutoff thresholds which represent both weak and strong commuting ties.

These supplementary analyses are presented in Table G.1. We incorporate two commuting networks with a weak tie threshold cutoff, 0.1% and 0.25%, as well as two commuting networks with a strong tie threshold cutoff, 1% and 2.5%. We find with small cutoffs there is a stronger effect and with strong cutoffs there is no effect. This is likely because the small commuting tie thresholds lead communities to be connected to most of the other communities in the network, while the strong commuting tie thresholds leads communities to only be connected to a very select few hub communities such as O'Hare and the Loop.

We explore this possibility using descriptive statistics by examining the number of isolates, the average indegree and outdegree, and the range for indegree and outdegree in each commuting network. Table G.2 presents these results. The average indegree and outdegree changes dramatically between the weakest and strongest tie thresholds. Further, while in the weakest network the minimum number of work communities a home community is connected to is 21, in the strongest network there are two isolates who are not connected to any work community.

Table G.1 Spatial Lag and Error Models Predicting STI Prevalence by Varying Commuting Thresholds, 2002-2014

	0.1% Cutoff	0.25% Cutoff	1% Cutoff	2.5% Cutoff
Prior STI Prevalence	-.087	.306***	.392***	.391***
	(.047)	(.061)	(.030)	(.031)
Concentrated Disadvantage	.076*	-.023	-.056*	-.054
	(.034)	(.041)	(.028)	(.028)
Residential Stability	-.064*	-.561***	-.010	-.014
	(.029)	(.060)	(.030)	(.030)
Diversity	-.035	.059	-.061	-.064
	(.050)	(.060)	(.037)	(.03)
Local Worker Density	-.221***	-.208***	-.002	-.009
	(.045)	(.046)	(.037)	(.041)
Constant	5.77***	7.12***	.055	-.154
	(.282)	(.47)	(.081)	(.362)
Network STI Risk	84.2***	11.3***	.065	.078
	(15.5)	(.919)	(.082)	(.434)
Error Variance Parameter	1.45***	.471***	.145***	.146***
	(.243)	(.026)	(.003)	(.003)
N	1001	1001	1001	1001
AIC	-9066	-5909	-901	-900
BIC	-8968	-5811	-802	-802

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001.
*p<.05, **p<.01,*** p<.001

Table G.2 Network Descriptive Statistics for Varying Commuting Thresholds

	.1% Cutoff	.25% Cutoff	.5% Cutoff	1% Cutoff	2.5% Cutoff
Indegree					
Mean	33.2	16.3	8.2	4.1	1.5
SD	23.8	20.7	16.8	14.8	8.9
Min	0	0	0	0	0
Max	76	76	76	76	74
Outdegree					
Mean	33.2	16.3	8.2	4.1	1.5
SD	4.7	4.1	2.7	1.3	.7
Min	21	6	2	1	0
Max	43	26	14	7	3
Isolates	0	0	0	0	2

Appendix H: Supplementary Analyses Examining Absolute STI Cases

Table H.1 evaluates models using the absolute number of STI cases rather than a standardized combined measure of STI rates. In these models, we control for the logged total population. The results are substantively the same as those presented in the main text which use a standardized measure.

Table H.1 Spatial Lag and Error Models Predicting Absolute STI Cases, 2002-2014

	Without Network	Spatial Network	Commuting Network	Public Transit Network
Total Population Logged	71.1*	30	59.1*	94.5**
	(31.5)	(29.6)	(27.5)	(31.4)
Prior STI Cases	.52***	.453***	.485***	.529***
	(.030)	(.035)	(.033)	(.032)
Concentrated	-5.94	-4.09	-5.4	-9.41
Disadvantage	(9.79)	(9.12)	(11.2)	(10)
Residential Stability	47.3***	35.4**	49***	32.9*
	(11.6)	(11)	(13.2)	(13.2)
Diversity	-17.7	-22.7	-3.22	-22.7
	(13)	(12.1)	(14.6)	(13.4)
Local Worker Density	-5.66	5.31	41.8**	-14.9
	(13.4)	(12.8)	(12.8)	(14)
Constant	51.6***	48.8***	55.5***	54.2***
	(1.2)	(1.2)	(1.3)	(1.26)
Network STI Risk		.414***	.955***	.882***
		(.084)	(.012)	(.006)
Error Variance Parameter		-.0776	.958***	.884***
		(.123)	(.011)	(.006)
N	1001	1001	1001	1001
AIC	9949	9876	9679	7421
BIC	10042	9979	9782	7524

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<.05, **p<.01,*** p<.001

Appendix G: Supplementary Analyses Motivating the Fixed Effects Models

Our analyses use fixed effects spatial and network autoregressive models. Though the geographic patterning of STI prevalence remains relatively stable over time, we find important variation between 2002 and 2014. We find 35% of communities experienced a change between the STI tercile rankings at least once during the duration of our study, with 67% of these experiencing change between rankings 1 and 2 (lowest and medium STI categories) and 33% between 2 and 3 (medium and highest STI categories).

Table I.1 presents the results of the Hausman test comparing the fixed effects and random effects spatial and network autoregressive models. This test assesses whether a fixed or random effects model is more appropriate for the data and analysis. The fixed effects examine the within-unit change of communities; in contrast, a random effects model accounts for both within and between differences of communities and presents the average effects across the 77 communities and time period. The Hausman test considers whether the more stringent model, the fixed effects model which focuses solely on within variation, is most appropriate. These results indicate that a fixed effects model is the more appropriate modeling strategy to examine STI rates over time for the network with no spatial weights matrix, the commuting network, and the public transit network. A random effects model would be more appropriate to model STI prevalence with our geographic contiguity matrix. However, for consistency we used a fixed effects model in our main results. Table I.2 presents the results of the random effects spatial and network autoregressive models and the results are substantively the same.

Table I.1 Hausman Test Comparing Fixed Effect and Random Effect Spatial and Network Autoregressive Models

	Chi-Squared	**P-Value**
Without Network	407.15***	0.000
Geographic Contiguity Network	18.3	0.370
Commuting Network	122.09***	0.000
Public Transit Network	1618.31***	0.000

Note: There were issues with the random effects spatial autoregressive model with the public transit network. Sometimes the model would not converge, whiles others, the model failed to meet the asymptotic assumptions of the Hausman test. The results reported above are from when the model converged and met the assumptions. *p<.05, **p<.01,*** p<.001

Table I.2 Spatial Lag and Error Models Predicting STI Prevalence using Random Effects, 2002-2014

	Without Network	Spatial Network		Commuting Network		Public Transit Network	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Prior STI Prevalence	.943***	.544***	.533***	.327***	.307***	-4.34***	-1.11***
	(.0116)	(.0712)	(.0571)	(.0396)	(.0375)	(.563)	(.122)
Concentrated Disadvantage	.0284**	.00905		-.035		1.15***	
	(.0109)	(.0304)		(.0239)		(.327)	
Residential Stability	-.00147	-.0449*	-.0463**	-.169***	-.161***	-3.3***	-1.15***
	(.00583)	(.018)	(.0177)	(.0307)	(.0299)	(.444)	(.0993)
Diversity	-.0218**	-.0961***	-.0982***	-.0412	-.00347	-1.15**	-.603***
	(.00768)	(.0234)	(.0227)	(.047)	(.0398)	(.37)	(.122)
Local Worker Density	-.00302	-.0554*	-.0579**	-.0392	-.0373	1.01*	.284
	(.00626)	(.0224)	(.021)	(.0238)	(.0237)	(.474)	(.17)
Constant	.00183	.0000856	-.00069	1.43***	1.41***	.211	.0143
	(.0193)	(.0267)	(.0272)	(.0994)	(.0996)	(1.25)	(.421)
Network STI Risk		.364***	.379***	3.45***	3.41***	24.8***	8.57
		(.0822)	(.0617)	(.163)	(.161)	(2.51)	(.)
Error Variance Parameter		-.179	-.204*	7.51***	7.65***	68.6***	44.9***
		(.133)	(.103)	(.0201)	(.0727)	(6.92)	(2.27)
N	1001	1001	1001	1001	1001	1001	1001
AIC	-683	-682	-684	-2561	-2561	-7114	-7256
BIC	-585	-574	-580	-2453	-2458	-7011	-7158

Note: Standard errors in parentheses. All models include year dummies. The Wald test of spatial autocorrelation is significant across all models at p<.001. *p<.05, **p<.01,*** p<.001

Appendix J: References in Appendices

Adimora, A. A., & Schoenbach, V. J. (2005). Social Context, Sexual Networks, and Racial Disparities in Rates of Sexually Transmitted Infections. *The Journal of Infectious Diseases*, *191*, S115–S122. https://doi.org/10.1086/425280

Adimora, A. A., & Schoenbach, V. J. (2013). Social Determinants of Sexual Networks, Partnership Formation, and Sexually Transmitted Infections. In S.O. Aral, K. A. Fenton, & J. A. Lipshutz (Eds.), *The New Public Health and STD/HIV Prevention: Personal, Public and Health Systems Approaches* (pp. 13–31). https://doi.org/10.1007/978-1-4614-4526-5

Block, P., Stadtfeld, C., & Snijders, T. A. B. (2016). Forms of Dependence: Comparing SAOMs and ERGMs from Basic Principles. *Sociological Methods and Research*, 1–38. https://doi.org/10.1177/0049124116672680

Desmarais, B. A., & Cranmer, S. J. (2012). Micro-Level Interpretation of Exponential Random Graph Models with Application to Estuary Networks. *The Policy Studies Journal*, 40(3), 402–434.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832–42.

Harling, G., Subramanian, S. V., Bärnighausen, T., & Kawachi, I. (2014). Income Inequality and Sexually Transmitted in the United States: Who Bears the Burden? *Social Science & Medicine*, *102*, 174–182. https://doi.org/10.1016/j.socscimed.2013.11.025

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, 29(2), 216–230. https://doi.org/10.1016/j.socnet.2006.08.005

Leifeld, P., Cranmer, S. J., & Desmarais, B. A. (2018). Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. *Journal of Statistical Software*, 83(6). https://doi.org/10.18637/jss.v083.i06

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (P*) models for social networks. *Social Networks*, 29:173–91.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36, 99–153.

Thomas, J. C., & Thomas, K. K. (1999). Things Ain't What They Ought to be: Social Forces Underlying Racial Disparities in Rates of Sexually Transmitted Diseases in a Rural North Carolina County. *Social Science & Medicine*, *49*, 1075–1084.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge, U.K.: Cambridge University Press.