Supplementary Online Content

Green B, Horel T, Papachristos AV. Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Intern Med*. Published online January 3, 2017. doi:10.1001/jamainternmed.2016.8245

This supplementary material has been provided by the authors to give readers additional information about their work.

eText. Supplementary Text

Glossary of social network and epidemiological terms

Cascade: A series of events in which an individual with a particular state transmits that state to his or her associates, who in turn transmit that state to their associates, and so on. A cascade can also refer to the nodes and edges involved in a particular transmission history.

Connected Component: A subgraph of a network in which every node is connected to every other node along edges contained in the subgraph. The nodes in a connected component are not reachable along network paths by any nodes outside the connected component.

Co-Offending: When two or more individuals jointly engage in a crime together.

Co-Offending Network: A social network created by linking unique individuals to each other through acts of co-offending. In co-offending networks, the nodes represent the unique individuals and the ties connect individuals who have co-offended.

Degree of separation: The distance between two nodes, measured as shortest number of edges that must be traversed to connect them. A friend is one degree of separation away (and is known as a first-degree neighbor), a friend of a friend is two degrees away, and so on.

Edge (or tie): A defined relationship between two nodes, in this study co-offending together.

Infection: In this study, infection refers to being shot, i.e., someone becomes infected when they become a gunshot subject.

Infector: The person most responsible for causing another individual to become infected (i.e., shot) by exposing that subject to gun violence. The infector is not assumed to be the one who shoots the subject, but rather the one who exposes him or her to the risk of gunshot violence.

Largest Connected Component: The connected component of a network that contains the most nodes.

Network Neighbors: Nodes that are connected through a path in the network. First-degree neighbors are directly connected by an edge, second-degree neighbors are connected indirectly through one intermediary (i.e., by two edges), and so on.

Node: A unique individual in a network; in this study representing individuals arrested in Chicago between 2006 and 2014.

Offending: Committing a crime.

Social Contagion: The transmission of a state (e.g., beliefs; in this study, being a gunshot subject) from one person to another through social interaction.

Social Network: A defined set of social relationships or interactions (edges) among a set of social actors (nodes).

Subject: Someone who has been shot by a gun (excluding suicides, accidents, and shootings involving law-enforcement personnel).

Violence event: An event in which someone becomes a subject of gun violence (including events involving subjects who have previously been shot).

1. Gun violence in Chicago

Our study examined co-offending networks in a single city: Chicago, IL. While studies of co-offending networks in other cities find many similarities with those in Chicago,[1-5] it is nonetheless important to situate our study within the broader context of Chicago's gun violence problem.

Over the past two decades, Chicago, like many other U.S. cities,[6,7] has seen unprecedented declines in crime and is currently experiencing some of its lowest rates of violent crime since the mid-1960s. eFigure 1 illustrates this point by plotting the annual index crime rate in Chicago from 1965 to 2013.[8] Rates of index crime remained relatively stable between 1966-1973 at approximately 3,300 per 100,000 and then jumped dramatically around 1973 to 5,882 per 100,000. Index crime rates fell again until about 1983 when they jumped to more than 8,000 per 100,000 during the mid-1980s. The apex of index crimes in Chicago occurred in 1991 when the rate was 10,647.9 per 100,000. After 1991 index crime fell at a steady rate throughout the late-1990s into the present day. In 2013, for example, the index crime rate was 4,251.2 per 100,000—approximately the same rate as in 1966.

Homicides in Chicago followed the same general pattern as overall index crimes. eFigure 2 plots Chicago's homicide rate (per 100,000) from 1965 to 2013. eFigure 2 clearly shows a parallel decline to that of overall index crime. Homicide increased rapidly

around 1967 before leveling off in the mid-1970s and early-1980s (with some fluctuations throughout that period). Homicide reached its apex in 1992 with a rate of 31.8 per 100,000. After that point, homicide declined drastically (despite some periodic spikes) until 2004. Between 2004 and the present, homicide rates hovered around 14 per 100,000, though they jumped in 2012 to 17.6 per 100,000. The homicide rate in 2013 was 14.1 per 100,000—the lowest overall rate since 1966.

The main variable of interest in the present study is *fatal and non-fatal gunshot victimization*, excluding self-inflicted and accidental gunshot injuries as well as legal interventions (i.e. police-related shootings). eFigure 3 plots the monthly combined number of fatal and non-fatal gunshot injuries during the observation period, 2006 to 2014. The expected seasonal variation of gun violence,[9,10] with peaks in the summer months, is also apparent. Average monthly rates during the study period ranged from 71.25 shootings in February to 245.5 in July.

In short, Chicago's rates of gun violence are higher than the national average, but are currently at some of the city's lowest rates in recent decades. In many ways, crime patterns in Chicago are similar to those in other cities, including the spatial distribution of crime and the concentration of crime in small networks.[11-15] The PI has written extensively on gun violence in Chicago and refers readers seeking further information to other relevant publications.[11,12,16-18]

eMethods. Supplementary Methods

1. Data description

The data used in this study were provided to the PI as part of a non-disclosure agreement with the Chicago Police Department (CPD). The PI controlled access to the data and supervised all data analyses and processing. Two sources of data were used in the present study. The first set of data entailed every arrest recorded by the CPD from January 1, 2006 to March 31, 2014 (N=1,458,957). These data include detailed incident-level information for each arrest, including the demographic information of each arrested individual. Note that these data include *only* arrests, and not non-custodial stops or contacts. The second dataset details incident-level data on all fatal and non-fatal shootings during the same period: these include homicides and non-fatal injuries (excluding self-inflicted and accidental injuries as well as those occurring during legal intervention). As with the arrest information, the shooting data also contained detailed information about each shooting, including the demographic information of the victims. Events and individuals are uniquely identified across both datasets using internal alphanumeric codes created by CPD (which we refer to as Event Codes and Identity Codes, respectively), thereby allowing us to match events and people over time and across datasets.

1.1 Data limitations

These data were not without limitations. First, police data have known biases, including: (a) undercounts of the true volume of crime because most crimes go unreported; (b) problems caused by data-entry errors or the use of aliases; and (c) biases in criminal justice practices and polices, including racial and neighborhood profiling, that might skew the true geographic and socio-demographic distribution of crime.[19-21] Regarding this last point, we make no claims of whether arrests were justified, but simply rely on them as the systematic recording of an observed behavior. Second, since crime is generally underreported, our co-offending data most likely *underestimates* the social ties related to risky behavior. And, third, without comparative data from other cities, it is difficult to know how representative the Chicago co-offending network is of co-offending more generally.

2. Data processing

In this section we describe how we processed police records into a co-offending social network with detailed attributes for each person. See Figure 1 for a diagram of this process.

2.1 Missing data

We first cleaned the arrest records to infer and fill in missing information.

Twenty-three percent (N=331,592) of the arrest records had no labeled Event Code (EC). In these instances, we determined which entries came from the same event by matching entries based on their date, time, and location. If multiple entries had identical information in all of these fields, we assumed they were from the same arrest and generated a new, unique EC for all such entries. We also provided unique ECs for all other entries where no match was found.

A small portion of entries (N=2,595, 0.2%) contained no recorded Identity Code (IC). For some entries (N=873), we identified the proper IC by finding other arrest records where the offender had the same birthdate, race, sex, and home address. For the rest of the cases (N=1,722), we assigned new, unique ICs. Where possible, we assigned the same IC to records that matched on birthdate, race, sex, and home address.

Next, we cleaned the entries (N=1,087, 0.07%) where there was no recorded birthdate for the offender. In 85% of these cases (N=927) we found the correct birthdate based on the other arrest records of the same individual, where the proper birthdate was recorded. Where this was not possible (N=160), we removed the entry from the data.

We performed a similar procedure to identify the residential neighborhood of offenders when this data was missing. 35,137 records (2.4%) involved individuals who lived in Chicago but did not have a labeled neighborhood. (We defined neighborhoods based on which police district the offender lived in. We ignored smaller geographic divisions such as police beat because the data within each group was too sparse.) We identified the district for 28,098 of these records based on other arrests of the same individual. This left 7,039 records from Chicago where the neighborhood was not recorded, plus 162,026 records from people who do not live in Chicago. We pooled these 169,065 offenders into a single "null" neighborhood.

Finally, there were a few entries (N=130, 0.009%) with no labeled sex for the offender. We determined the correct sex for 3 of these cases by finding other arrest records with the same identity code. We imputed the other 127 individuals to be male, as 85% of the arrested individuals were male.

2.2 Forming a social network

We used the arrest records to generate a bipartite network that connects arrest events and people (Figure 1). That is, the network connects each person to every arrest in which he or she was involved. Equivalently, the network connects each arrest event to all of the people arrested. The network is clearly bipartite since people cannot be linked with

people, and arrests cannot be linked with arrests. This network has a total 1,189,225 arrest event nodes, 462,516 person nodes, and 1,458,957 edges.

We performed a bipartite projection on the person nodes to obtain a one-mode social network, where nodes represent each person who was arrested during the study period. This network contained 462,516 nodes and 467,506 edges. Unweighted edges connect every pair of people who were arrested in the same event during the study period, connecting individuals based on their association with the same crime.

Edges connected pairs of individuals who co-offended together at some point during the study period. Due to the one-mode structure, incidents in which more than two people co-offended together were represented by edges between every pair of individuals involved rather than all individuals to a common incident. More than two-thirds of the arrests involving multiple people had only two participants, hoewever. Another 18% contained three people, leaving only 13% of arrests that involved more than three people. This shows that our one-mode co-offending network is a reasonable representation of co-offending dynamics.

We treated the co-offending network as static rather than forming each edge at the date of first co-offense. Although the co-offending events occur at specific points in time, previous research on co-offending has shown that the individuals involved typically already have close existing relationships.[22] Because we can identify the presence of these

relationships but not the date when those relationships formed, we generate a static network that includes every co-offense throughout the study period.

While it is possible to build a weighted social network with edge weights corresponding to the number of co-arrests between individuals, we did not do this for three reasons. First, it is difficult to determine if two people appear together in the arrest data multiple times because they actually co-offended together multiple times or simply faced multiple charges from the same co-offense. Second, there are very few edges between individuals who co-offended multiple times together—a finding consistent with prior research on co-offending networks.[3] 94% of edges have weight=1, 5% of edges have weight=2, and all larger edge values account for the remaining 1% of edges. Finally, we found no evidence that high-weight edge facilitate the transmission of gunshot victimization. For each edge weight represented in the network, we looked at the percentage of pairs where both individuals were infected. The probability that such a pair exists actually goes *down* as the edge weight increases. In particular, out of the 269 highest-weight edges (weight>5), there is not a single pair where both people were infected. This leads us to believe that the few high-weight edges that exist have little or no special effect on the contagion of violence.

2.3 Adding victim attributes

We used gunshot victimization records to determine our dependent variable of whether or not any individual in the data was a victim of a fatal or non-fatal gunshot injury during the study period. For each victim, we record the date of every fatal and non-fatal gunshot victimization associated with that individual. Eleven percent (N=1,251) of the victims of non-fatal gunshots had multiple victimizations during the study period, with a maximum of five. Twelve percent (N=247) of the victims of fatal gunshots had previously during the study period been the victim of a non-fatal gunshot.

Although we restricted ourselves to gunshot victims who are also in the co-offending network, we still captured the vast majority of victims in our analysis: 93% of nonfatal victims members of the co-offending arrest network, and 80% of fatal victims are in the network.

2.4 Largest connected component

Decomposing the co-offending network into disjoint connected components yielded many small components and one giant component. This is similar to the pattern observed in other empirical networks.[23,24] More than half of the nodes (56%) are isolated, corresponding to people who were never arrested with anyone else. Of the 284,876 connected components, only one contains more than 30 nodes. This largest connected

component contains 30% of the nodes in the network (N=138,163), 89% of the edges (N=417,635), and 74% of the victims (N=9,773). As is standard in social network analyses,[23] we take the largest connected component (LCC) as the focus of our study.

The largest connected component resembled a typical social network. The network's degree distribution followed a power-law distribution with scaling exponent 1.39. This means that the LCC is a scale-free network, which is very common among social networks.[24] The LCC has a clustering coefficient of 0.6 and an average path length of 8.3. In comparison, an Erdös-Réyni random graph of the same size has a clustering coefficient of 0.00003 and an average path length of 6.78. Since the LCC is highly clustered with a similar average path length compared to a random graph, it is a small-world network.[25]

3. Homophily and confounding versus contagion

Understanding how gunshot victimization might make its way through a network requires understanding different reasons for how patterns of gunshot violence might emerge in a network: failing to account for all possible explanations can lead to overestimating the effects of social contagion.[26-28] We consider three potential explanations: individuals associate with similar peers (homophily), individuals are exposed to the same environmental factors (confounding), and individuals influence one another's behavior over time (contagion).[26,27,29,30] To distinguish between these explanations, we analyzed the temporal patterns of victimization with those generated by simulations that account for homophily and confounding but not contagion. We ran 10,000 Monte Carlo

simulations of the study period, assigning to each victim a new victimization date that is consistent with his or her exposure to violence based on risk factors and environmental influences. By shuffling the infection dates between victims as described below, the simulations generated a set of networks that 1) retained the aggregate patterns of gun violence, as measured by the number of victimizations each day, and 2) accounted for the effects of homophily and confounding but assume no social contagion.

Homophily would explain the temporal clustering of victims in the network if people co-offend with others who have similar risk factors and therefore are likely to be shot at similar times. Many prior studies have shown strong relationships between certain risk factors and exposure to violence,[31] a relationship that our data corroborates. In our simulations we controlled for whatever traits cause two individuals to co-offend together by holding constant the network structure and victim identities: each victim has the same neighbors in both the real and simulated data. Confounding would explain the pattern of victimizations if features such as age and neighborhood expose similar individuals to violence at the same time. Our simulations controlled for confounding by shuffling victimization dates only between individuals who are the same age, gender, and ethnicity; live in the same neighborhood; and both either belong or do not belong to a gang (if an individual does not match with anyone else across all of these features, that person's infection date is not altered). We also controlled for the fact that violence rates fluctuate, following a predictable seasonal trend of rising in the summer and declining in the winter.[9,10] Furthermore, some years have more incidents of violence than others and crime in the US and Chicago declined during the observation period.[7,11] In order to

control for violence rates over time, we simulated the exact same number of infections per day as observed in the data. Together, these controls ensured that we accurately represented each person's exposure to violence as it relates to individual and environmental risk factors.

This approach allowed us to determine the extent to which the observed concentration of victims could be explained without any social contagion. If the concentration of victims was primarily due to homophily and environmental confounding, the simulations would accurately recreate the observed pattern of gunshot violence. On the other hand, if social contagion was responsible for some victimizations, we expected to see that the observed victimizations appear closer together in time than the simulations could explain.

Because we held constant the set of victims and infection dates, we could simulate an infection process that lacks social contagion by shuffling the matching between victim identity and victimization date. Given our method's similarity to the previously-developed "Shuffle Test,"[29] we refer to our approach by the same name.

Our Shuffle Test ran as follows:

1) Take the LCC and identify the gunshot victims from the data.
2) Divide all the victims into groups that share the same birth year, gender, ethnicity, residential neighborhood, and gang membership status (i.e. belong to a gang or not).

3) Within each group, randomly permute the infection dates associated with each individual. Individuals in groups by themselves retain the same infection date.

This yielded a new version of the LCC, with the same victim population and overall set of infection dates as the raw data. Each victim was infected at different times during the simulated study period compared to the observed data, but in a manner that is consistent with the rate at which he or she was exposed to violence.

For each simulation, we measured how many days passed between infections within every pair of first-degree associates who were both victims (N=9,568). If one or both victims were infected multiple times during the study period, we take the minimum time difference between infections. As our network and set of victims are fixed based on the data, the quantity and identities of such pairs remain constant in every simulation. If these pairs of victim are shot equally close together in time in the data and simulations, then we will be able to conclude that homophily and confounding are sufficient to explain the data. Alternatively, if the data exhibits a higher degree of temporal clustering, this will imply that explanations beyond homophily and confounding are necessary.

Our comparisons of simulations with the data show that homophily and confounding cannot fully explain the concentration of gunshot victims within the network (eFigure 7). As reported in the main text, these pairs are shot on average 60 days closer together than the simulations can explain. We similarly found that the median time difference between victimizations is 75 days shorter in the data than in the simulations. We then evaluated

how many pairs become victims within a specific, short period of time. While 7.6% of pairs in the data became victims within 30 days of one another (N=726), there were only 4.0% (3.7%-4.4% 95CI) such pairs in the simulations. Homophily and confounding, then, explained only 52.6% of the gunshot victimization that occurred between associates within 30 days. Similarly, 17% of pairs in the data became victims within 100 days of one another, compared to only 12% in the simulations. These results indicated that victims are clustered both temporally and topologically in a manner that homophily and confounding cannot fully explain. This suggested that considering social contagion may help explain when and where victimizations occur in the network. We turn in the next section to modeling this social contagion directly.

4. Hawkes contagion model

We modeled the contagion of violence using a multi-dimensional Bayesian Hawkes process over the co-offending network. We first present the general definition of Hawkes Processes, then instantiate and adapt it to the contagion of gun violence over a network.

4.1 Hawkes processes

Hawkes processes are a class of self-exciting temporal point processes originally introduced by Alan Hawkes in the early 1970s,[32] and have recently become common as a way to model contagion and diffusion processes. Applications include the spread of seismic events,[33] information spread in social networks,[34] and stock market trading dynamics.[35]

A convenient way to describe temporal point processes is through their conditional intensity function, which describes the instantaneous probability of occurrence of an event at any given time *t*. For Hawkes processes, the conditional intensity function can be written as the sum of endogenous time-varying intensities (capturing the intra-network influence of the events preceding time *t*) and an exogenous intensity (capturing the influence of all extra-network factors).

Formally, for a *D*-dimensional Hawkes process with *N* infection events, let us introduce the set of events $\mathcal{E} = \{(t_i, k_i)\}_{1 \le i \le N}$ where t_i denotes the time of event *i* and k_i the dimension (or coordinate) on which it occurs. The conditional intensity function is defined as follows:

$$\lambda_k(t) = \mu_k + \sum_{i=1}^{N} \phi_{k_i,k}(t - t_i) \quad (1)$$

where $M = (\mu_k)_{1 \le k \le D}$ is the vector of exogenous intensities (also known as background rates) and the functions $\Phi = (\phi_{i,j})_{1 \le i,j \le D}$ is the matrix of endogenous kernel functions (also known as exciting functions). For a pair of coordinates *(u,v)*, $\phi_{u,v}(t)$ models the influence of coordinate *u* over coordinate *v* after *t* time has passed since *u* was infected. The kernel functions are non-negative and causal (i.e. $\phi_{u,v}(t) = 0 \forall t < 0$). In particular, this implies that the summation in Equation (1) is only over the indices *i* such that $t_i < t$.

From this definition, we see that the Poisson process can be characterized as a special case of the Hawkes process, with a constant exogenous intensity and no dependence on past events. That is, $\lambda(t) = \lambda$.

We refer the reader to other sources[36,37] for a formal discussion of the conditional intensity function and its proper interpretation in a Hawkes process. From these we apply the following formula for the log-likelihood of events $\mathcal{E} = \{(t_i, k_i)\}_{1 \le i \le N}$ given M and Φ over observation period $[0, T]$:

$$\mathcal{L}(\mathcal{E} | M, \Phi) = \sum_{i=1}^{N} \log \lambda_{k_i}(t_i) - \sum_{k=1}^{D} \int_0^T \lambda_k \, dt \tag{2}$$

The first sum calculates the log-likelihood of every infection event that *did* occur, and the second sum calculates the log-likelihood that each individual was *not* infected at all other times.

4.1.1 Contagion of gun violence as a Hawkes process

We model the contagion of gun violence as a Bayesian Hawkes Process by defining the following features: each network vertex (i.e. each individual) occupies its own coordinate of the Hawkes Process and each gunshot victimization is an event of the process occurring on the coordinate that corresponds to the victim (repeated victimizations of the same individual correspond to multiple events on the same node, and are treated the same as single victimizations).

Exogenous intensity. We assume that the exogenous intensity is the same for every individual in the network, and attribute the observed fluctuations of violence rates (eFigure 3) to a seasonal effect independent of peer contagion. For this reason, we fit a time-varying function $\mu(t)$ to the data and use it for the common exogenous intensity (described in Section 4.2.1).

Endogenous exciting functions. The exciting function $\phi_{u,v}(t)$ models the effect of person *u* on person *v* after *t* time has passed since *u* was infected and captures two common assumptions regarding the spread of contagions (eFigure 4).

1. Time: consistent with previous models used to infer the spread of contagions over social networks,[34,38,39] we assume that the impact of earlier infections on future events decays as the time passed since the original infection increases. Additionally, influence can only travel forward in time: an infection has no impact on those that came before it. As is common for Hawkes processes,[34-39] we assume an exponential decay and obtain the following formula for the temporal component of the exciting functions:

$$f_\beta(t) = \begin{cases} \beta e^{-\beta t} & \text{if } t > 0 \\ 0 & \text{if } t \le 0 \end{cases} \quad (3)$$

2. Network structure: epidemiologists commonly assume that contagious events are localized and that the transmission probability increases closer to the source.[40-43] In our case, we assume that violence is more likely to spread between people who

are closely linked in the network and measure the distance between individuals based on network topology. Based on previous studies of violence in social networks,[12,30] we assume that infections are able to be transmitted across a network distance of up to three degrees of separation; people who are further away in the network have no effect on one another. Hence, we obtain the following formula for the structural component:

$$g_\alpha(u,v) = \begin{cases} \alpha \mathrm{dist}(u,v)^{-2} & \text{if } \mathrm{dist}(u,v) \le 3 \\ 0 & \text{if } \mathrm{dist}(u,v) > 3 \end{cases} \quad (4)$$

Finally, we obtain the exciting function by combining the above two components:

$$\phi_{u,v}(t) = f_\beta(t) g_\alpha(u,v) \quad (5)$$

4.1.2 Model likelihood

Using Equation (2) and the model presented in Section 4.1.1, we can now write the log-likelihood of observed infection events $\mathcal{E} = \{(t_i, k_i)\}_{1 \le i \le N}$. V denotes the set of vertices in the network, and $[0,T]$ marks the study period.

Since some individuals were the victims of fatal gunshots during the study period, they were not susceptible to infection during the entire study period. For these victims, the second summand of Equation (2) only needs to be integrated until their time of death. Denoting by T_v the time of death of vertex v (T_v=T if the individual didn't die during the study period), we obtain:

$$\mathcal{L}(\alpha,\beta,\mu|\mathcal{E}) = \sum_{i=1}^{N} \log\lambda_{k_i}(t_i) - \sum_{v\in V}\int_0^{T_v} \lambda_v(t)dt \quad (6)$$

4.2 Inferring model parameters

In this section, we describe how we learned the optimal parameters to describe the Hawkes model described in Section 4.1.

4.2.1 Exogenous intensity

Because the seasonal variations in gunshot rates remain consistent throughout the study period (eFigure 3), we assume these are not purely driven by noise or social contagion. We model these seasonal variations with a periodic sinusoidal function.

Let *M(t)* denote the expected number of total victimizations occurring on day t . We assume the following form:

$$M(t) = A\left[1+\rho\sin(\omega t+\phi)\right] \quad (7)$$

Because violence fluctuates annually, we know that the period is one year, i.e. $\omega = 2\pi/365.24$. We learn the remaining three parameters $\{A,\rho,\phi\}$ using non-linear least squares estimates with the Gauss-Newton algorithm. This yields:

$$M(t) = 3.73\left[1+0.43\sin\left(\frac{2\pi}{365.24}t+4.36\right)\right] \quad (8)$$

eFigure 5 depicts the number of infections on each day of the study period along with the function $M(t)$.

Because we do not yet know the importance of the exogenous intensity in spreading gunshot violence, we only keep $\{\rho,\phi\}$ from the fitted parameters. In other words, we only keep the parameters characterizing the seasonal fluctuations; the base amplitude A of the exogenous intensity will be inferred together with the kernel function parameters in the following section.

Finally, we relate the aggregate number of infections $M(t)$ to the node-level exogenous intensity $\mu(t)$. By definition:

$$M(t)=\sum_{v\in V}\int_{t-1}^{t}\mu(s)ds=|V|\int_{t-1}^{t}\mu(s)ds \tag{9}$$

where we used that the exogenous intensity is identical across all nodes. Assuming that $\mu(t)$ is approximately constant over the course of one day, we get $M(t)=|V|\mu(t)$. Hence we obtain the following form for the exogenous intensity:

$$\mu(t)=\mu_0\left[1+0.43\sin\left(\frac{2\pi}{365.24}t+4.36\right)\right] \tag{10}$$

where $\mu_0=A/|V|$.

4.2.2 Learning the optimal model parameters

Using the exogenous intensity from Section 4.2.1, the log-likelihood now depends on three parameters $\{\alpha,\beta,\mu_0\}$. Finding the maximum likelihood estimate of these parameters amounts to solving the following optimization problem:

$$\arg\max_{\alpha,\beta,\mu_0} \mathcal{L}(\alpha,\beta,\mu_0|\mathcal{E}) \tag{11}$$

Unfortunately, the function $\mathcal{L}(\alpha,\beta,\mu_0|\mathcal{E})$ is not jointly concave in its three arguments. We will, however, exploit the following fact.

Proposition 1. The function $(\alpha,\mu_0)\mapsto\mathcal{L}(\alpha,\beta,\mu_0|\mathcal{E})$ is concave.

Proof. Expanding the terms in Equation (6), it is clear that the second sum is linear in $\{\alpha,\mu_0\}$. Hence it is sufficient to show that for $1\le i\le N$:

$$h(\alpha,\mu_0)=\log\left(\mu_0\left[1+0.43\sin\left(\frac{2\pi}{365.24}t+4.36\right)\right]+\sum_{j:t_j<t_i}\alpha\,\mathrm{dist}(u,v)^{-2}f_\beta(t)\right) \tag{12}$$

is concave. For this, we see that the operand of the log function is linear in $\{\alpha,\mu_0\}$. By composition with the concave function log, we obtain that *h* is concave and thus conclude the proof. □

We observed numerically that $\mathcal{L}$ has many local optima; hence we solve Equation (11) using the following heuristic:

1. We perform a brute force grid search to locate good starting points for the refining heuristic.

2. Starting from the best point obtained during the first step, we refine the solution by alternated minimization:
    a. Optimize over $\{\alpha,\mu_0\}$ for a fixed value of β. By Proposition 1 we were able to use standard convex optimization (gradient descent, in this case) to solve this step exactly.
    b. Optimize over β for a fixed value of $\{\alpha,\mu_0\}$, using simulated annealing.

Other heuristics were considered: using gradient descent as well for the optimization over β, or using global gradient descent to optimize over $\{\alpha,\beta,\mu_0\}$ at the same time. All heuristics led to the same optimal solution, indicating that our initial grid search was precise enough to identify good starting points. We obtained the following values of the parameters at the optimum:

$$\alpha = 7.82\times10^{-3}, \quad \beta = 3.74\times10^{-3}, \quad \mu_0 = 1.19\times10^{-5} \tag{13}$$

4.2.3 Validation on simulated data

In order to validate our approach for learning the Hawkes model parameters, we evaluated the method described in Section 4.2.2 on synthetic data. Starting from the same co-offending network as in the dataset (i.e. the LCC), we generated synthetic contagion events by simulating the Hawkes contagion model described in Section 4.1.[44] The model parameters are fixed to the values obtained in Equation (13).

We then computed the maximum likelihood estimator described in Section 4.2 on the synthetic contagion events and compared the inferred parameters to the true values used during the simulation. To analyze how our estimates converge as we observe more data, we truncated the synthetic dataset at increasing time horizons between 0 and 3,000 days (our study period spanned 3,012 days) and trained the maximum likelihood estimator separately on each truncated dataset.

We performed this procedure five times to generate five independent sets of synthetic contagion events (eFigure 6). We observed that the inferred parameters for α and β vary for short study periods but quickly converged toward the optimal value as the study period increases. The learned parameters for μ_0 are close to optimal even for short study periods. After 3,000 days, all inferences for α were within 10.8% of the true value, all inferences for β were within 12.7%, and all inferences for μ_0 were within 2.1%. The mean parameters for each parameter from the five trials at 3,000 days were all within 1.8% of the optimal value. These results indicate that our parameter inference method was able to reliably determine the parameters of a Hawkes model over the study period length used.

4.3 Inferring the pattern of infections

Given fitted values of the parameters of the Hawkes contagion model, we then determined whether each infection event *(t,v)* was primarily caused by the exogenous

background rate or endogenous peer contagion. Using Equation (1), we compared the value of the exogenous and endogenous intensities at the time *t* of infection, and attributed the infection event to the larger of the two quantities. In other words, we compared:

$$\mu(t) = 1.19 \times 10^{-5}\left[1 + 0.43\sin\left(\frac{2\pi}{365.24}t + 4.36\right)\right] \tag{14}$$

with

$$\sum_{u \neq v} \phi_{u,v}(t) \tag{15}$$

and attributed the infection to peer contagion if (15) >(14).

For infection events *(t,v)* attributed to peer contagion, we could single out a single peer $\hat{u}$ event as the individual most responsible for transmission. This was achieved by choosing the peer $\hat{u}$ with the strongest social influence on *v* at time *t*. That is,

$$\hat{u} = \arg\max_{u}\left[\phi_{u,v}(t)\right] \tag{16}$$

We thus uncovered the pattern of infections: each infection event is attributed to either the exogenous intensity or a single past infection event. We draw an edge from each infection event to all other infections that it spawns. We note that edges are directed forward in time, making cycles impossible, meaning that every connected component in this infection network is a tree. We referred to each tree in the forest of infections as a cascade.

Our model identified 7,016 victimizations that were caused primarily by social contagion. We found that 46% of infections came from first-degree neighbors, 41% came from second-degree neighbors, and the remaining 13% of infections came from third-degree neighbors. Victims were shot on average 125 days after their infector, with a median time difference of 83 days.

We found 4,107 distinct cascades of victimization through the LCC during the study period. The distribution of the cascade sizes extracted from our dataset can be seen in eFigure 9. Consistent with previous findings in related domains,[45,46] this distribution follows a power-law of exponent 1.8.

4.3.1 The timing of infections and co-offending

Previous research suggests that co-offending represents strong and enduring relationships between individuals.[22] We therefore treated co-offending as evidence of an existing relationship between two individuals involved rather than as a point-in-time estimate of when that relationship formed, and accordingly generated static edges in the network representing that two individuals co-offended together at any time during the study period. Nonetheless, it is useful to evaluate the temporal relationship between when an individual is infected by an associate compared to when the two first co-offended together, to ensure that the typical timing of these two events supports this modeling decision. We considered all contagion events (as inferred in Section 4.3) between first-degree neighbors and found that 77.1% of all infected individuals had been co-arrested

with their infector before being shot (eFigure 8). Many of these infections occur in the immediate aftermath of being arrested with a recent victim. Another 10.8% of victims were shot in the year immediately preceding their first co-arrest with the infector. These results indicate that, even discounting prior research studying the close ties that generally exist between people before co-offending, our findings of contagion are not merely artifacts of the static network.

4.4 Causality in the Hawkes model

The notion of causality has been the subject of many debates.[47] With this in mind, we should qualify the previous section in which we assign a single cause to certain gunshot victimizations.

In discussing the definition of causality, Ned Hall proposed the following thought experiment:[48]

> *Suzy and Billy, two friends, both throw rocks at a bottle. Suzy is quicker, and consequently it is her rock, and not Billy's, that breaks the bottle. But Billy, though not as fast, is just as accurate: Had Suzy not thrown, or had her rock somehow been interrupted mid-flight, Billy's rock would have broken the bottle moments later.*

According to some interpretations of causality, within this scenario Suzy and Billy are jointly responsible for the bottle breaking: they were both throwing rocks at it, and the fact that Suzy's rock reached the bottle first is coincidental. However, it also clear that Suzy's rock shattered the bottle. Even if we had not observed the rock that first hit the bottle, since Suzy was throwing rocks more quickly than Billy we could say that the rock that shattered the bottle was more likely to have been thrown by Suzy.

The Hawkes contagion model can be re-interpreted in light of this example: as they become infected, victims begin to "throw rocks" at their associates with a frequency that decreases over time. Being shot due to peer infection is equivalent, in this metaphor, to being hit by a rock thrown by an associate. Since we do not observe whose rock hits first, the only thing we can say for certain is that at the time of victimization an individual was subject to the combined throws of his or her previously-infected neighbors. This combined effect is expressed mathematically by the sum in Equation (1).

It is now clear which interpretation to give the cascades extracted in Section 4.3: it is a simplification where we designate the cause for victimization to be the associate who was "throwing rocks" with the highest frequency at the time of infection. This simplification is acceptable in that this associate is the *most likely* to be the direct cause of infection. Nonetheless, based on another interpretation of causality we would instead consider the throws from every associate to be jointly the cause of victimization.

5. Predicting victimization

If gunshot injuries in the network are caused by social contagion, then how much can knowledge of the co-offending network improve our ability to predict future individual victimization? Traditional models of individual victimization tend to rely on individual, contextual, or ecological risk factors,[49] yet our findings suggest that adding in additional temporal and network features might improve such prediction.

In this section we compare the Hawkes contagion model with a traditional demographics model by evaluating how effectively each model predicts who will be shot on a given day.

Given that law enforcement and social services must make targeted interventions with limited resources, predictions of gunshot victims are only actionable if they precisely identify a small population that faces the highest risk to be shot. With this in mind, the proper evaluation for any model is its ability to identify future victims as part of the population's highest-risk community.[49,50] For this study, we define three "high-risk communities" as those people identified with the top 0.1%, 0.5%, and 1.0% of risk to become infected. These correspond to populations with 138, 691, and 1,382 individuals from the largest connected component, respectively.

We compared the predictive abilities of three different models:

Demographics model: This model uses each person's demographic features and risk factors to predict who will be infected on a given day. We include all features available in our data, capturing many of the variables shown to be most critical in predicting gunshot victimization.[31] We label as infected all people who have been shot before that date and label all others as non-infected. We then perform a logistic regression over the entire population, using the formula

victim ~ sex + race + age + gang.member + gang.name + N.prior.arrests + neighborhood

(while additional features would surely have been useful, we unfortunately did not have access to any variables beyond these). The resulting probabilities correspond to the background rate of the Hawkes contagion model and identify each person's risk to be shot.

Contagion model: This model uses the social contagion element of the Hawkes model to identify who is at most risk to become infected on a given day. It accounts for the network structure and infection history, but ignores all demographic and environmental attributes. Based on the observed pattern of gunshots, we measure each person's exposure to violence at a given time.

Combined model: This model uses the results from the demographics and network models. We combine the risk scores from the other two models using a

weighted sum, generating a fully specified Hawkes contagion model for the spread of gunshot violence through the co-offending network.

For every day of the study period, we executed all three models to predict each person's likelihood to be shot on that day. We then identified (based on the data) the people who were actually shot on the current day of the trial and noted their relative risk in the population of co-offenders according to each model. For each model, we ended up with the rankings of all the victims on the day they were shot. We compared the different models by measuring how often they select victims when identifying the network's high-risk population. An ideal model would identify each day's *N* victims as the individuals with the *N* highest levels of risk.

eFigure 10 plots the cumulative distribution function for each model. The contagion model outperformed the demographics model for the high-risk quarter of the population (identifying more than half of the victims in this group), while the demographics model outperformed the contagion model for the rest of the population. The combined model reaped the benefits of both models, and performs best across the entire distribution. This shows that the contagion model is best equipped to predict future victims when focused on the portion of the population that faces the highest risk. Given that the goal of predicting victims is to provide law enforcement and social services with a small population for targeted interventions, the contagion and combined models are more effective than the traditional demographic model.

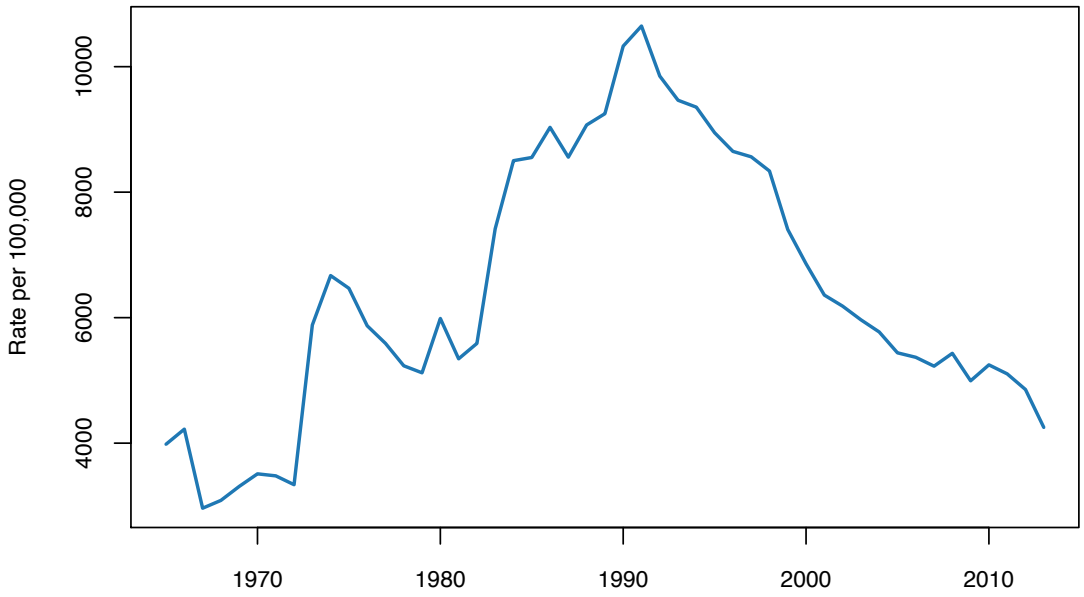
References

1. Papachristos AV, Braga AA, Hureau D. Social networks and the risk of gunshot injury. *Journal of Urban Health*. 2012;**89**(6):992-1003.
2. Papachristos AV, Braga AA, Piza E, Grossman L. The company you keep? The spillover effects of gang membership on individual gunshot victimization in a co-offending network. *Criminology*. 2015;**53**(4):624-49.
3. McGloin JM, Piquero AR. On the relationship between co-offending network redundancy and offending versatility. *Journal of Research in Crime and Delinquency*. 2009.
4. Schaefer DR. Youth co-offending networks: An investigation of social and spatial effects. *Social Networks*. 2012;**34**(1):141-9.
5. Morselli C. Inside Criminal Networks: Springer; 2009.
6. Blumstein A, Wallman J. The Crime Drop in America. New York: Cambridge University Press; 2000.
7. Zimring FE. The Great American Crime Decline. New York: Oxford University Press; 2006.
8. Crime in the United States [Internet]. U.S. Department of Justice. 2015.
9. McDowall D, Curtis KM. Seasonal variation in homicide and assault across large US cities. *Homicide Studies*. 2014.
10. McDowall D, Loftin C, Pate M. Seasonal cycles in crime, and their variability. *Journal of Quantitative Criminology*. 2012;**28**(3):389-410.
11. Papachristos AV. 48 years of crime in Chicago: A descriptive analysis of serious crime trends from 1965 to 2013. 2013 Contract No.: ISPS13-023.
12. Papachristos AV, Wildeman C, Roberto E. Tragic, but not random: The social contagion of nonfatal gunshot injuries. *Social Science & Medicine*. 2015;**125**(1).
13. Morenoff JD, Sampson RJ. Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces*. 1997;**76**(1):31-64.
14. Sampson RJ. Great American City: Chicago and the Enduring Neighborhood Effect: University of Chicago Press; 2012.
15. Block R, Block CR. Street gang crime in Chicago. In: Klein MW, Maxson CL, Miller J, editors. The Modern Gang Reader. Thousand Oaks, CA: Roxbury; 1995.
16. Papachristos AV, Kirk DS. Changing the street dynamic: Evaluating Chicago's Group Violence Reduction Strategy. *Criminology & Public Policy*. 2015;**14**(3):525-58.
17. Papachristos AV, Meares TL, Fagan J. Attention felons: Evaluating Project Safe Neighborhoods in Chicago. *Journal of Empirical Legal Studies*. 2007;**4**(2).
18. Papachristos AV, Wildeman C. Network exposure and homicide victimization in an African American community. *American Journal of Public Health*. 2014;**104**(1):143-50.
19. Kirk DS. Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *Journal of Quantitative Criminology*. 2006;**22**(2):107-29.
20. Thornberry TP, Krohn MD, editors. Comparison of self-report and official data for measuring crime. Measurement problems in criminal justice research: Workshop summary; 2002: The National Academic Press Washington, DC.

21. Berk RA. An introduction to sample selection bias in sociological data. *American Sociological Review*. 1983:386-98.
22. Warr M. Companions in Crime: The Social Aspects of Criminal Conduct. New York: Cambridge University Press; 2002.
23. Adamic LA. The small world web. Research and Advanced Technology for Digital Libraries: Springer Berlin Heidelberg; 1999. p. 443-52.
24. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*. 2002;**74**(1):47.
25. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;**393**(6684):440-2.
26. Aral S, Muchnika L, Sundararajana A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*. 2009;**106**(51):21544–9.
27. Cohen-Cole E, Fletcher JM. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *BMJ,*. 2008;**337**.
28. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*. 2011;**40**(2):211-39.
29. Anagnostopoulos A, Kumar R, Mahdian M, editors. Influence and Correlation in Social Networks. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008; Las Vegas, Nevada.
30. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*. 2007;**357**:370-9.
31. Tracy M, Braga AA, Papachristos AV. The Transmission of Gun and Other Weapon-Involved Violence Within Social Networks. *Epidemiologic Reviews*. 2016.
32. Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*. 1971;**58**(1):83-90.
33. Marsan D, Lengliné O. Extending earthquakes' reach through cascading. *Science*. 2008;**319**(5866):1076-9.
34. Farajtabar M, Du N, Gomez-Rodriguez M, Valera I, Zha H, Song L, editors. Shaping social activity by incentivizing users. Proceedings of The 28th Annual Conference on Neural Information Processing Systems; 2014; Montreal, Canada: Curran Associates, Inc.
35. Lininger T. Multivariate Hawkes Processes [Dissertation]. Zurich: Swiss Federal Institute of Technology; 2009.
36. Rasmussen JG. Temporal point processes: The conditional intensity function. 2011 January 24, 2011. Report No.
37. Daley DJ, Vere-Jones D. An Introduction to the Theory of Point Processes: Springer-Verlag New York; 2007.
38. Gomez-Rodriguez M, Leskovec J, Krause A. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data*. 2012;**5**(4).
39. Linderman S, Adams R, editors. Discovering Latent Network Structure in Point Process Data. Proceedings of the 31st International Conference on Machine Learning; 2014; Beijing, China: JMLR.
40. Grenfell B, Bjørnstad O, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. *Nature*. 2001;**414**(6865):716-23.

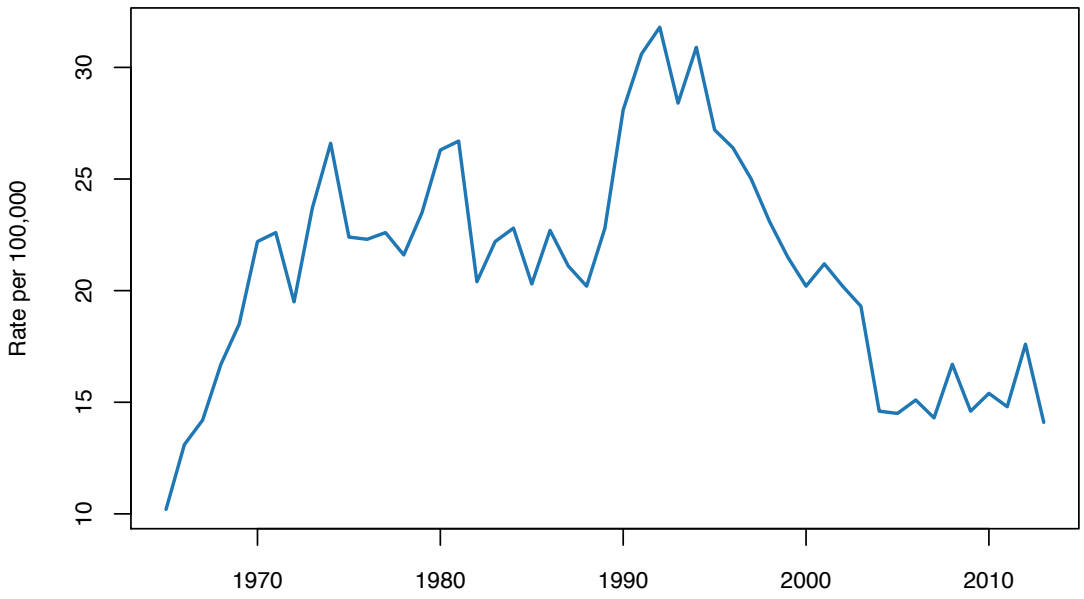
41. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*. 2006;**312**(5772):447-51.
42. Stewart JQ. Demographic gravitation: evidence and applications. *Sociometry*. 1948:31-58.
43. Sen A, Smith T. Gravity Models of Spatial Interaction Behavior: Springer Science & Business Media; 2012.
44. Møller J, Rasmussen JG. Perfect simulation of Hawkes processes. *Advances in Applied Probability*. 2005:629-46.
45. Leskovec J, McGlohon M, Faloutsos C, Glance NS, Hurst M, editors. Patterns of cascading behavior in large blog graphs. Proceedings of the 2007 SIAM International Conference on Data Mining; 2007; Minneapolis: SIAM.
46. Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J, editors. Can cascades be predicted? Proceedings of the 23rd International Conference on World Wide Web; 2014; Seoul, Korea. 2567997: ACM.
47. Pearl J. Causality: Cambridge University Press; 2009.
48. Hall N. Two concepts of causation. *Causation and Counterfactuals*. 2004:225-76.
49. Chandler D, Levitt SD, List JA. Predicting and preventing shootings among at-risk youth. *The American Economic Review*. 2011;**101**(3):288-92.
50. Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*. 2011;**106**(493):100-8.
51. Block CR, Block RL, Illinois Criminal Justice Information Authority. Homicides in Chicago, 1965-1995. Inter-university Consortium for Political and Social Research (ICPSR); 2005.

eFigure 1. Index Crime in Chicago (Rate per 100,000), 1965 to 2013



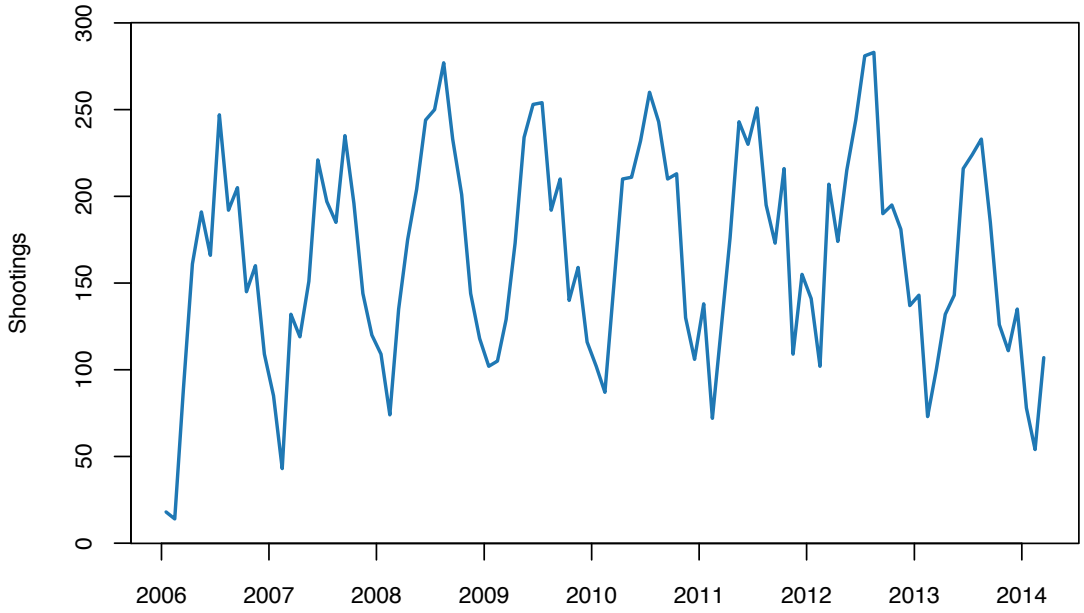
Index crimes include all murders, criminal sexual assaults, aggravated assaults/batteries, burglaries, thefts, robberies, arsons, and motor vehicle thefts. Crime rose throughout the 1970s and 1980s, peaking in 1991 with a rate of 10,647.9 per 100,000 people. Crime in Chicago has since declined steadily, with a rate of 4251.2 per 100,000 in 2013. Data come from the FBI Unified Crime Reports.[8]

eFigure 2. Homicide in Chicago (Rate per 100,000), 1965 to 2013



Homicide rates between 1965 and 2013 follow a similar pattern as index crime rates, peaking in the early 1990s (with a rate of 31.8 per 100,000 in 1992) and declining steadily since then. The homicide rate in 2013 was 14.1 per 100,000 people, the lowest since 1966. Homicide data from 1965 to 1994 were provided by Carolyn Rebecca Block and Richard L. Block through the National Archive of Criminal Justice Data.[51] Detailed data on homicides from 1995 to 2010 were provided by the Chicago Police Department.

eFigure 3. Monthly Counts of Fatal and Nonfatal Gunshot Injuries in Chicago, 2006 to 2014

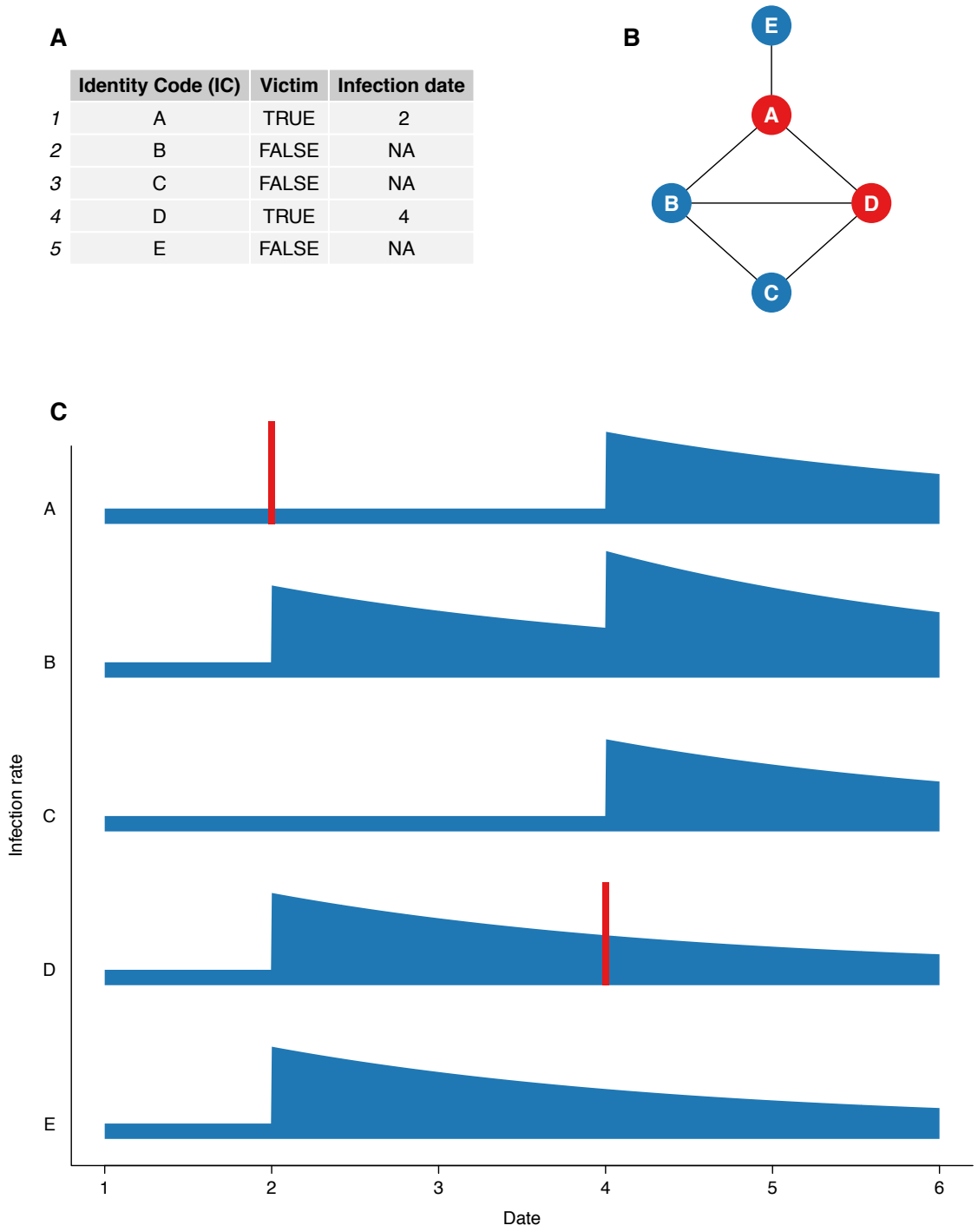


The number of shootings per month varies widely depending on the time of year: violence peaks in the summer and declines in the winter. In 2008, for example, the number of shootings per month varied from 74 in February to 277 in August.

eFigure 4. Hawkes Model Dynamics Over an Example Network

A

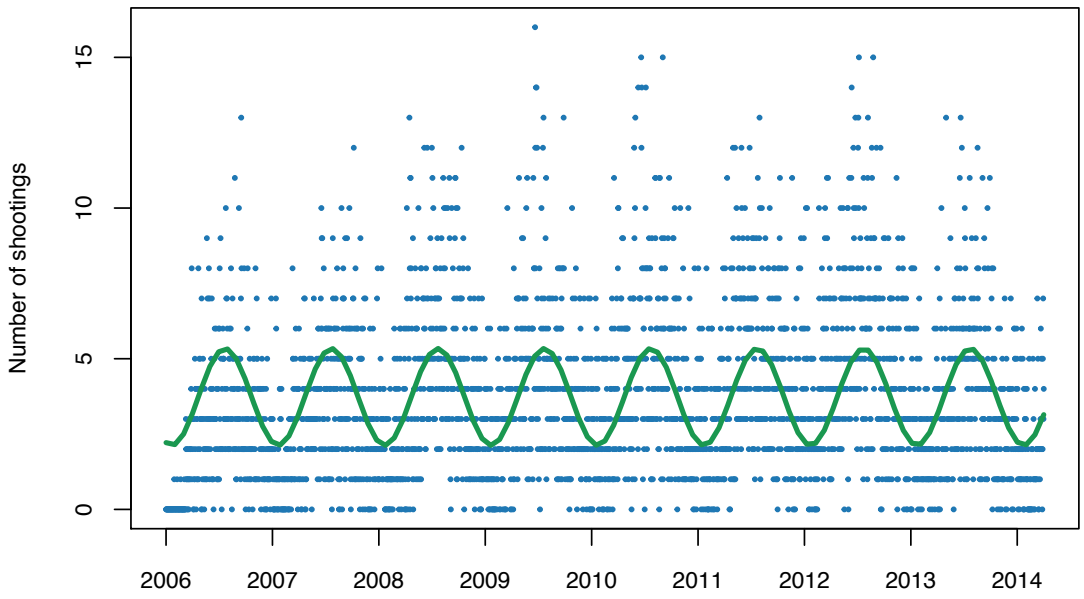
	Identity Code (IC)	Victim	Infection date
1	A	TRUE	2
2	B	FALSE	NA
3	C	FALSE	NA
4	D	TRUE	4
5	E	FALSE	NA



(A) A table of identities and whether each individual was a gunshot victim. The infection time for each victim is also recorded.
(B) The co-offending network of individuals in (A), with victims marked in red and non-victims in blue.

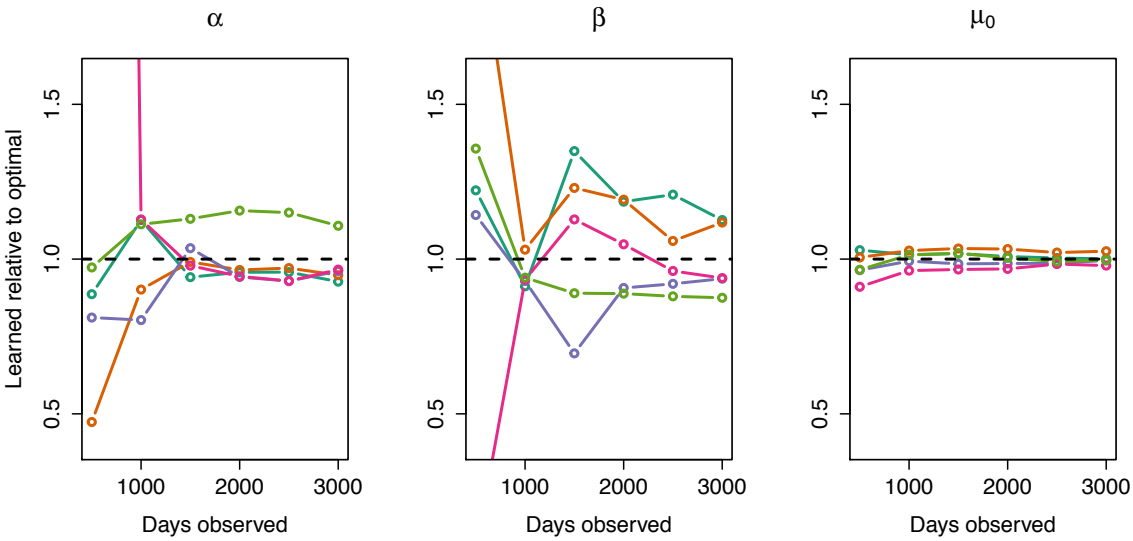
(C) The infection rates of each person in the network over a five-day period. Each individual is initially susceptible to infection only due to a small background rate, based on exogenous features, that is constant across the population. When individual A is infected on day 2 (marked with a red line), it causes a spike in the infection rate of its three neighbors: B, D, and E. The impact of this infection decays over time. Because a node cannot generate further infections in itself, A's infection rate does not change when it is infected. Node D is infected on day 4 (marked with another red line), causing the infection rates of A, B, and C to spike. Because the effects of peer contagion are additive and B is connected to both infected nodes, B has the highest infection rate after D is infected.

eFigure 5. Shootings per Day and Best-Fit Curve During the Study Period



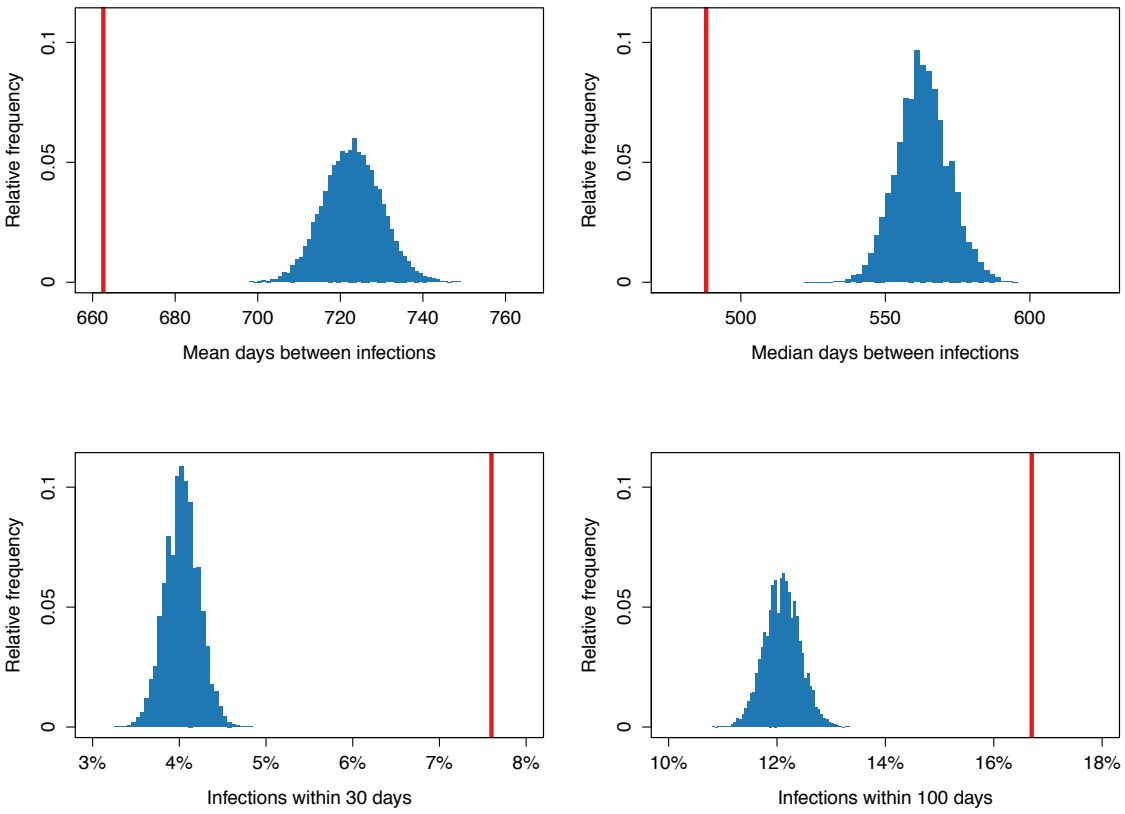
Each blue dot represents the number of shootings (fatal and non-fatal) on a single day. Values ranged from 0 (N=280, 9.3%) to 16 (N=1), with a mean of 3.7 and median of 3. In order to model how violence rates vary over time, we fit a sinusoidal curve to this data (in green).

eFigure 6. Learned Parameters Relative to Optimal From Five Simulations of the Hawkes Contagion Process



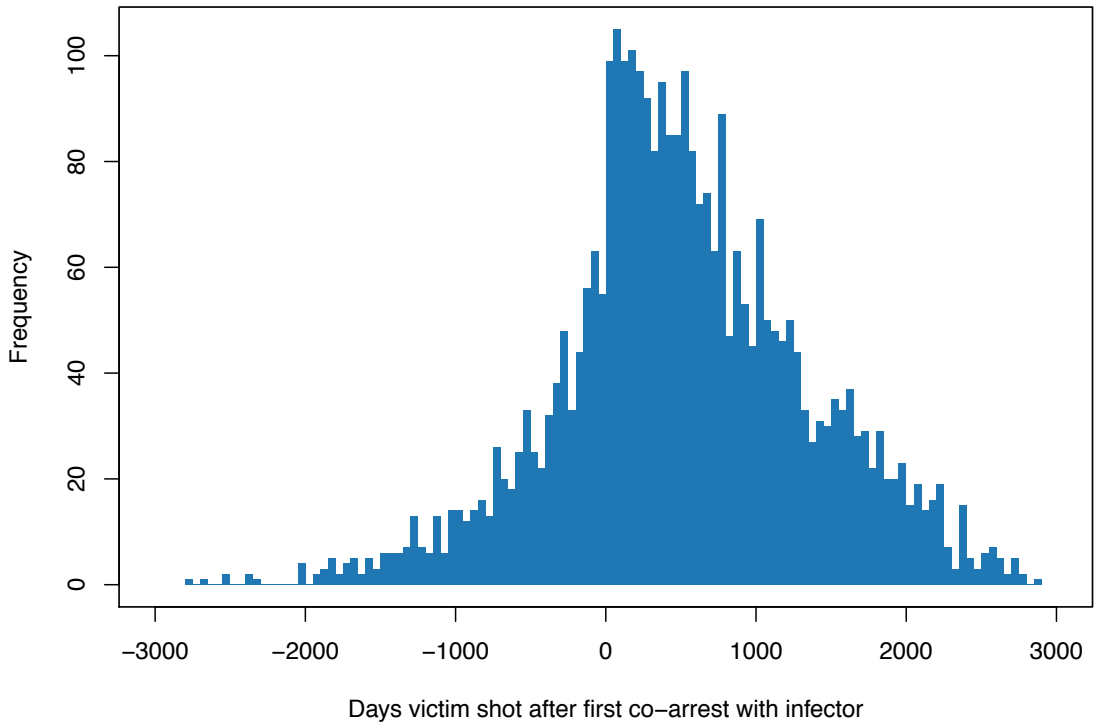
We simulated five Hawkes contagion processes over the LCC using the parameters found in Equation (13). Using the method in Section 4.2, we learned the parameters that best describe the simulated data and compare these to the actual value. The dashed black lines indicate that the optimal result is for the learned parameters to be identical to the parameters actually used. Colored lines show the learned parameters relative to optimal for each simulated dataset as we observe a different number of days. We observe that the inferred parameters for α and β vary notably for short study periods but quickly converge toward the optimal value as the study period increases. The learned parameters for μ_0 are close to optimal even for short study periods. The means of the learned parameters from the five trials at 3,000 days are all within 2% of the optimal value, indicating that our parameter inference method is able to determine the parameters of a Hawkes model over the actual study period.

eFigure 7. Results From 10,000 Monte Carlo Simulations of the Study Period Without Any Social Contagion



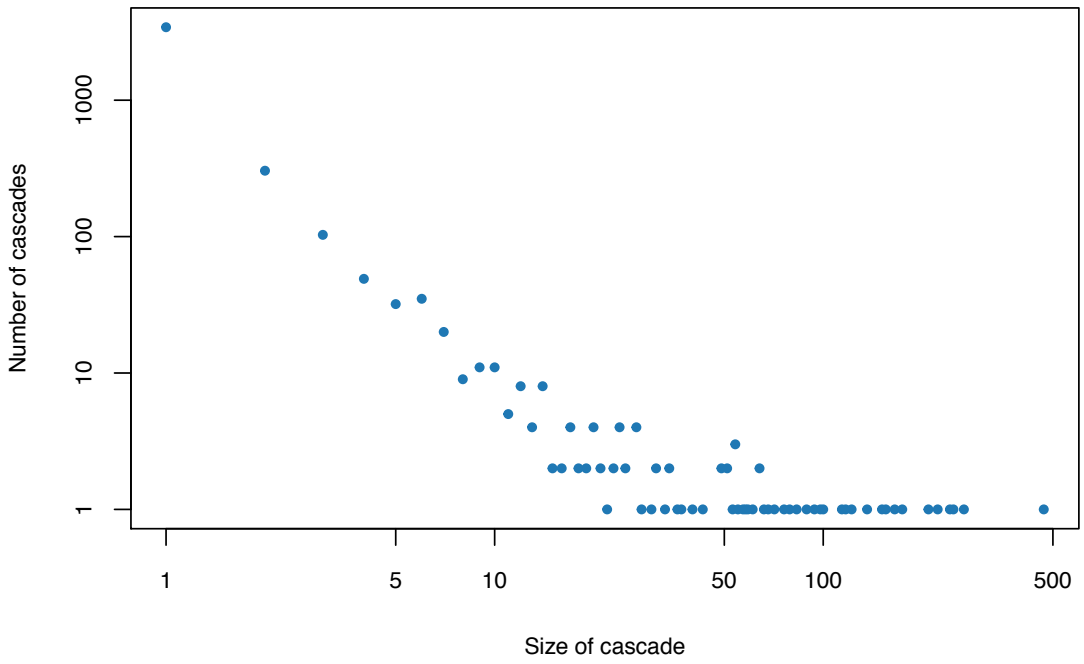
These plots display the temporal relationships between infections for all pairs of first-degree neighbors where both people were gunshot victims during the study period. Vertical red lines represent the observed values from the data. Simulations based on homophily underestimate by a large margin how many pairs will be infected close together in time, and can explain only 52.6% of infections that occur within 30 days of each other. The mean time between infections is 60 days shorter in the data than in the simulations.

eFigure 8. Temporal Difference Between Co-Offending and Gunshot Infection



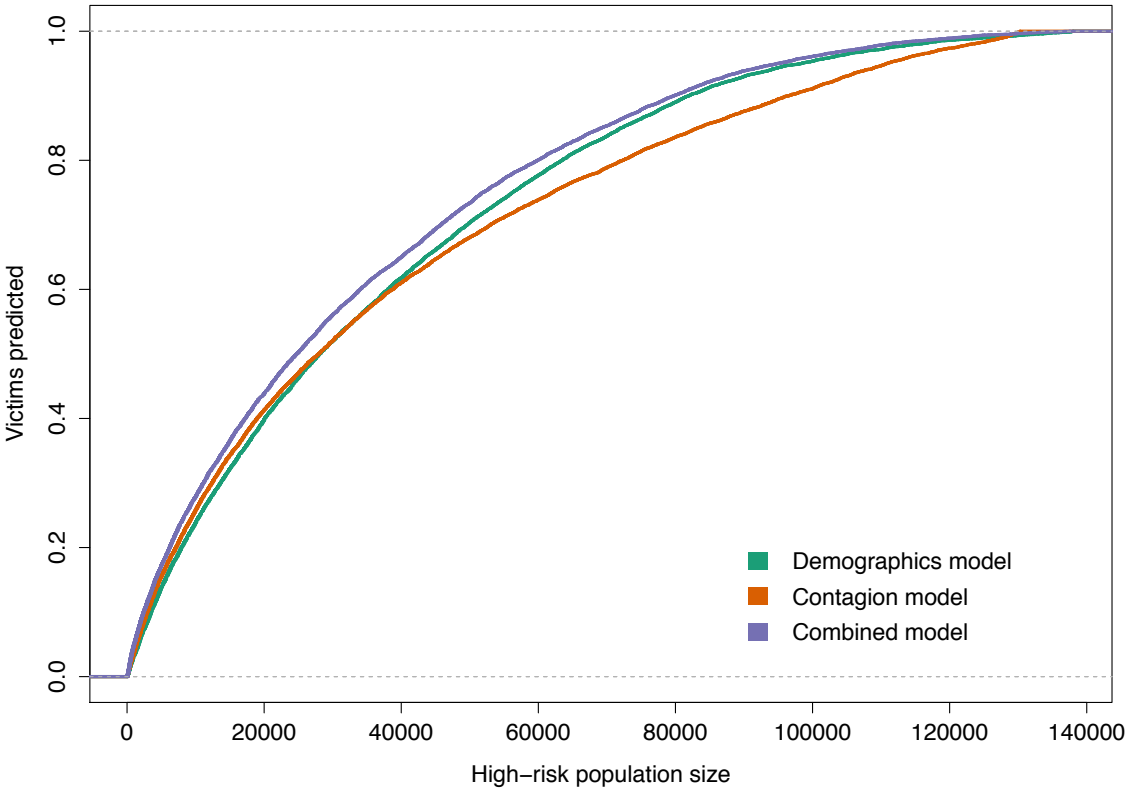
The values here indicate the number of days between co-offending and being victimized, among cases where our model determined that a victim was infected by a first-degree neighbor. Positive values indicate that the victim was infected *after* having previously co-offending with the infector. As is clear from the figure, the majority of infections (77.1%) that we detected between first-degree neighbors occurred after the infector and victim had already been arrested together. Among the victims who were shot before co-offending with their infector, 47.5% co-offended with the infector within a year of being infected. These results, combined with previous research on the enduring nature of co-offending relationships,[22] confirm the validity of modeling the contagion process over a static network.

eFigure 9. Distribution of Cascade Sizes Found in the Network



Cascade sizes ranged from 1 (N=3,427, 83.4%) to 469 (N=1), with a mean size of 2.7 people. The distribution follows a power law with scaling exponent 1.8.

eFigure 10. Cumulative Distribution Function for the Demographics, Contagion, and Combined Prediction Models



The x-axis represents a population size and the y-axis reports what fraction of victims was within the high-risk population of that size. Among the highest-risk 20,000 people, for example, the demographics model identifies 39.9% of victims, the contagion model identifies 41.3%, and the combined model identifies 43.9%. Overall, the contagion model outperforms the demographics model for high-risk quarter of the population (identifying more than half of the victims in this group), while the demographics model outperforms the contagion model for the rest of the population. The combined model reaps the benefits of both models, and performs best across the entire distribution.