Modeling the origin of urban-output scaling laws

V. Chuqiao Yang,[1,2,*] Andrew V. Papachristos,[3,4] and Daniel M. Abrams[1,5,6]
[1]*Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208, USA*
[2]*Santa Fe Institute, Santa Fe, New Mexico 87501, USA*
[3]*Department of Sociology, Northwestern University, Evanston, Illinois 60208, USA*
[4]*Institute for Policy Research, Northwestern University, Evanston, Illinois 60208, USA*
[5]*Northwestern Institute for Complex Systems, Northwestern University, Evanston, Illinois 60208, USA*
[6]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*

(Received 26 January 2018; revised manuscript received 8 April 2018; published 16 September 2019)

Urban outputs often scale superlinearly with city population. A difficulty in understanding the mechanism of this phenomenon is that different outputs differ considerably in their scaling behaviors. Here, we formulate a physics-based model for the origin of superlinear scaling in urban outputs by treating human interaction as a random process. Our model suggests that the increased likelihood of finding required collaborations in a larger population can explain this superlinear scaling, which our model predicts to be non-power-law. Moreover, the extent of superlinearity should be greater for activities that require more collaborators. We test this model using a novel dataset for seven crime types and find strong support.

DOI: 10.1103/PhysRevE.100.032306

I. INTRODUCTION

Physics-based models for human processes have had remarkable success in recent years, perhaps due to the increasing availability of relevant quantitative data (see, e.g., models of pedestrian synchrony, crowd dynamics, migration patterns, community formation, even changing religious affiliation [1–8]). Here, we explore the sociophysics of human productivity, focusing in particular on the origin of superlinear scaling laws that have been observed for a wide range of urban outputs (see Fig. 1(a) for several examples). Increases in these outputs can be mostly beneficial, as with GDP and patents, or mostly harmful, as with crime or contagious disease [9–13].

The scaling of serious crime was previously reported to be superlinear [10,14] with a power law exponent of approximately 1.16. When we break down the data and compare across the seven FBI crime report categories,[1] however, the scaling behavior varies significantly: some categories show approximately linear scaling, while others are strongly superlinear. These differences persist for all years since 1999, the earliest year for which data are available (see Fig. 2). One illustrative example is the comparison between robbery and rape, as shown in Fig. 1(b). Robbery scales superlinearly with city size, while rape scales close to linearly.[2]

It has remained unclear why some quantities are affected by city population more than others: previous efforts at understanding superlinear scaling in urban outputs have largely focused on the similarities rather than differences [14–18]. In addition, many models [14,16–19] rely on a power-law assumption for the scaling behaviors, which was recently challenged [20].

Explaining the variations in the scaling behavior and formulating a non-power-law framework are now two significant challenges in developing a scientific understanding of urban scaling. Here, we propose a novel model that explains and predicts the variation in scaling among different urban outputs, without relying on a power law hypothesis.

II. MATHEMATICAL MODEL

A. Overview of the mathematical model

Since most urban outputs, such as AIDS infection, patenting, and many types of crime have social components [21–26], we are motivated to incorporate existing knowledge about social processes into our model. Mark Granovetter's landmark work "The Strength of Weak Ties" [27] argued that weak ties play an important role in providing information novel to one's social network that fosters outputs such as finding a job or starting a business. Motivated by this and direct empirical evidence for the importance of weak ties in innovation and crime [28,29], we base our model on the assumption that finding the right collaboration is key to human productivity: one must meet all the necessary collaborators for an output in order to produce. Mathematically, this key concept is expressed as

$$y(N) \sim N P_n[u(N)], \tag{1}$$

where $y(N)$ denotes the volume of an urban output (such as the total number of robbery cases or the number of patents) for a city with population N. The parameter n is the number of partners needed for the output, $u(N)$ is the average number of unique contacts for a person living in the city, and $P_n[u(N)]$

*vcy@u.northwestern.edu

[1]FBI crime categories are murder, rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft.

[2]Note that, for convenience, we frequently use the simpler but more ambiguous term "city" to refer to a Metropolitan Statistical Area (MSA) throughout this manuscript.

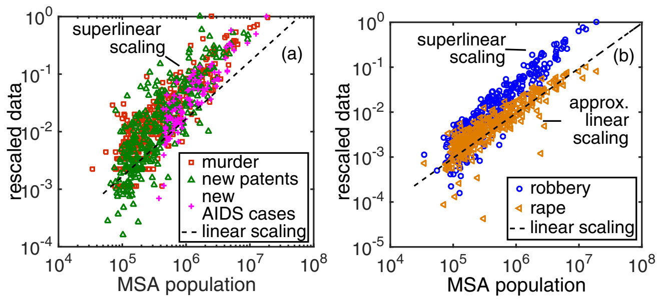

FIG. 1. *Urban outputs versus city size.* (a) Number of new patents, murder cases, and AIDS cases in U.S. Metropolitan Statistical Areas (MSA's) vs population. These three urban outputs exhibit superlinear scaling—as MSA population doubles, the amount of output more than doubles. (b) Number of robbery cases and rape cases vs MSA population. Robbery scales superlinearly, while rape scales close to linearly. All data are scaled to have maximum value 1.

is the probability of finding all required n collaborators among $u(N)$ contacts.

Here, we give an overview of the model's general conclusions, without functional form assumptions. In the next subsection, we will first show that P_n's dependency on u is in the form of

$$P_n \sim u^n(N). \tag{2}$$

Combining Eqs. (1) and (2), we have

$$y(N) \sim Nu^n(N).$$

Taking the logarithm of both sides and differentiating with respect to $\ln(N)$, we have

$$\beta \equiv \frac{d\ln(y)}{d\ln(N)} \sim 1 + n\frac{d\ln(u)}{d\ln(N)}. \tag{3}$$

Equation (3) leads to three predictions:

Superlinear scaling. $\beta \equiv d\ln(y)/d\ln(N)$ is often interpreted as the scaling exponent of y. Equation (3) predicts that if $d\ln(u)/d\ln(N) > 0$, meaning residents of more populated cities have more contacts (supported by empirical findings in [30]), and $n > 0$, meaning the activity typically requires more than one participant, then $\beta > 1$, giving rise to superlinear scaling.

Variation in scaling exponents. Equation (3) shows that β increases with n. For fixed N, β grows linearly with n. Thus Eq. (3) predicts that urban outputs requiring more participants should exhibit more pronounced superlinear scaling.

Possibility of non-power-law superlinear scaling. Individuals in bigger cities have the chance to meet more people, but cognitive limits (among other things) restrict them to only interacting with a small subset in a given period [31]. It is thus plausible for $d\ln(u)/d\ln(N)$ to decrease for large N. Since the scaling exponent β in Eq. (3) may depend on N, the result is superlinear scaling behavior that is not a power-law.

Considering typical patterns of collaboration can resolve the puzzle of why scaling behaviors vary across different urban outputs. The three predictions above hold for any general increasing function of $u(N)$. In the following sections, we first provide a derivation for $P_n(u) \sim u^n$. Then, in order to make quantitative predictions and compare the model with empirical data, we propose one general framework for estimating $u(N)$, derived from treating social interactions as a biased sampling process.

1. Derivation for $P_n(u)$

Among $u(N)$ unique contacts, only a small subset of those should result in partners for outputs such as crime and inventions. Whether a contact becomes a partner can depend on many factors, e.g., possession of a certain skill or establishment of a certain level of trust. We denote this probability by γ, which differs by the type of activity.

The need to find all required collaborators for an output to occur can be interpreted in two ways. The first is that the output requires n partners, each with a unique set of attributes (e.g., skill, relationship, etc.). The second is that the output requires n individuals each possessing the same set of attributes. We present calculations for both interpretations and show that the scaling relationship for P_n, the probability of finding all partners needed, is $P_n \sim u^n$ (to leading order) for both interpretations.

Finding n partners with unique attributes. The probability of finding at least one partner with a desired attribute out of the u people met is

$$q(u, \gamma_i) = 1 - (1 - \gamma_i)^u,$$

where $\gamma_i \ll 1$ is the probability of any given individual having the attribute.

After finding one partner among the u individuals met [with probability $q(u, \gamma_1)$], then, if $n > 1$, the searcher also needs to find another compatible partner among the $u - 1$ remaining contacts [with probability $q(u - 1, \gamma_2)$], and so on. The probability of finding all n partners can be expressed as

$$P_n = q(u, \gamma_1)q(u - 1, \gamma_2)\cdots q(u - n + 1, \gamma_n). \tag{4}$$

We expand Eq. (4) assuming $\gamma_i \ll 1$ for all i. Since the expansion of $q(u, \gamma_i)$ near $\gamma_i = 0$ is $q(u, \gamma_i) \approx u\gamma_i + O(\gamma_i^2)$, the leading order term for P_n is

$$P_n \sim u(u - 1)\cdots(u - n + 1)\prod_{i=1}^{n}\gamma_i.$$

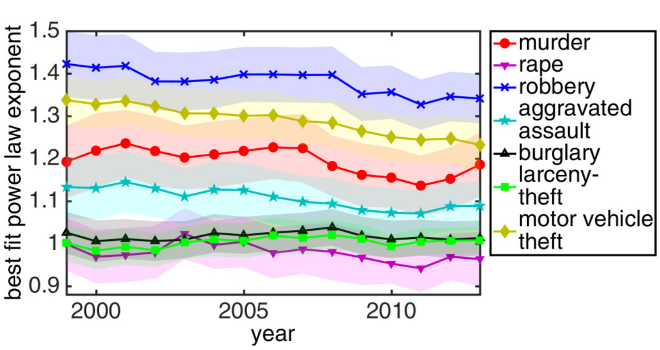

FIG. 2. *Superlinear scaling over time.* Best-fit exponents (measuring degree of superlinear scaling) for scaling laws, i.e., slopes of curves such as those in Fig. 1, for all seven FBI crime categories. The shaded regions show 95% confidence intervals. Exponents can differ considerably among crime categories, and are consistent over time.

With $u \gg n$, we have

$$P_n \sim u^n \prod_{i=1}^{n} \gamma_i.$$

Because γ_i are constants and we are only interested in how P_n scales with u, we express the scaling relationship as

$$P_n \sim u^n \ .$$

Finding n partners with the same attributes. The probability of finding *at least n* compatible partners with the same attributes out of the u people met is

$$P_n(u) = \sum_{r=n}^{u} \binom{u}{r} \gamma^r (1-\gamma)^{u-r}. \tag{5}$$

where γ is the probability of any person being a suitable partner. This calculation assumes the probability of each person being a suitable partner is independent.

The leading order term in Eq. (5) is the first term ($r = n$):

$$P_n \sim \binom{u}{n} \gamma^n (1-\gamma)^{u-n} = \frac{u!\gamma^n(1-\gamma)^{u-n}}{n!(u-n)!} \ . \tag{6}$$

We can simplify the ratio of factorials by writing it in terms of Γ functions:

$$\frac{u!}{(u-n)!} = \frac{\Gamma(u+1)}{\Gamma(u-n+1)} \ , \tag{7}$$

which can then be approximated using Stirling's approximation [32]

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b}\left[1 + \frac{(a+b)(a+b-1)}{2z}\right]$$

for large z and bounded a, b. Setting $z = u$, $a = 1$, and $b = 1 - n$, Eq. (6) can be simplified as

$$P_n \sim \frac{\gamma^n(1-\gamma)^{u-n}(2u - n^2 + n)u^{n-1}}{2n!}. \tag{8}$$

Assuming $u \gg n$, and retaining only leading order terms (highest power in u and lowest power in γ), we get

$$P_n \sim \gamma^n u^n.$$

So the leading order scaling behavior of $P_n(u)$ with u is

$$P_n(u) \sim u^n.$$

2. Derivation for $u(N)$

In order to make quantitative predictions, we provide a framework to estimate the expression $u(N)$. Since an MSA is defined based on social and economic integration, we approximate an MSA of population N as a closed system with respect to social interactions: all people in the city have some probability of interacting with one another.

Spatial population distribution in cities is a complex problem on its own, and we wish to avoid assumptions about the spatial distribution of social interactions. Instead, we consider a "social space"—a mathematical convenience to simplify our analysis—using the following approach. For each individual under consideration, we map all other individuals in the city to a one-dimensional space in which they are uniformly distributed and ordered by social distance to the individual under consideration (for an illustration, see Fig. 3). A larger distance implies a smaller probability of interaction. The position of an individual in this social space corresponds to his or her rank based on social distance to (or probability of interaction with) the individual under consideration. We then treat each social interaction as a sampling process (independent and with replacement)—the person under consideration chooses a person in the social space with whom to interact with probability density function $\rho(x;N)$, where x is position in the social space. By definition, ρ is rank-probability distribution, and a non-increasing function of x.[3]

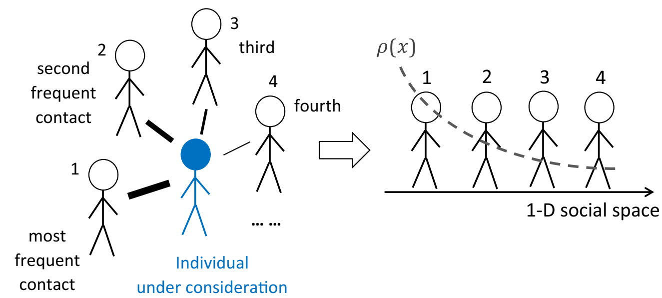

FIG. 3. *Mapping to social space.* Cartoon for individuals in a city mapped to a one-dimensional space, ordered by rank of probability of social interaction. The probability of interaction with a particular peer decreases with position in one-dimensional space x as $\rho(x)$.

To simplify calculations, we first consider sampling segments of social space rather than individuals embedded in it. We discretize the one-dimensional social space of total length L into M patches of size Δx. The ith patch (with center position x_i) is thus chosen with probability $\rho(x_i)\Delta x$. Taking n_s to be the total number of samplings (interactions) that have occurred, the expected total unique space sampled, L_u is

$$L_u = \Delta x \sum_{i=1}^{M} [1 - (1 - \rho(x_i)\Delta x)^{n_s}]. \tag{9}$$

We denote by $L_s = n_s \Delta x$ the length of total space sampled with repeated samples counted cumulatively. We can then rewrite Eq. (9) as

$$L_u = \Delta x \sum_{i=1}^{M} \left[1 - \left(1 - \frac{\rho(x_i)L_s}{n_s}\right)^{n_s}\right]. \tag{10}$$

The Laurent series expansion for $(1 - c/n)^n$ as $n \to \infty$ is

$$\left(1 - \frac{c}{n}\right)^n = e^{-c} - \frac{c^2 e^{-c}}{2n} + O\left(\frac{1}{n^2}\right).$$

Using this expansion, Eq. (10) can be rewritten as

$$L_u = \Delta x \sum_{i=1}^{M} \left[1 - e^{-\rho(x_i)L_s} + \frac{\rho(x_i)^2 L_s^2}{2n_s} e^{-\rho(x_i)L_s} + O\left(\frac{1}{n_s^2}\right)\right].$$

[3]In taking this "social space" approach, we leave for other work (see, e.g., [33] or [34]) the interesting questions of how geographical and social network structure may lead to particular scaling laws $\rho(x;N)$.

We take the continuum limit $\Delta x \to 0$ and $n_s \to \infty$, and neglect the $O(1/n_s)$ and higher order terms. Using $M\Delta x = L$, we have

$$L_u = \lim_{\Delta x\to 0} \Delta x \sum_{i=1}^{M}[1 - e^{-\rho(x_i)L_s}]$$
$$= M\Delta x - \lim_{\Delta x\to 0} \Delta x \sum_{i=1}^{M} e^{-\rho(x_i)L_s}. \quad (11)$$

The second term of Eq. (11) is a Riemann sum. Taking the continuum limit of Eq. (11) as $\Delta x \to 0$, the sum can be expressed in terms of an integral. We then have

$$L_u = L - \int_0^L e^{-\rho(x)L_s} dx. \quad (12)$$

Since the population distribution on the social space is uniform, the unique length covered by sampling (L_u) and the cumulative length sampled (L_s) directly correspond to the number of unique individuals met u and the cumulative number of samples s, respectively. The total length of the rank-space L by construction corresponds to the total population N. Changing notation in Eq. (12), we find an expression for the number of unique individuals met u:

$$u(N) = N - \int_1^N e^{-\rho(x)s} dx, \quad (13)$$

where s is the amount of sampling made by each individual in a certain period of time. Here, we assume s (reflecting casual contact interactions) does not change with city population.[4] We will estimate its value (assumed universal for simplicity) by fitting to all available datasets.

B. Closed-form expression for scaling of urban outputs

The closed-form expression for the total output y is thus

$$y(N) \sim N\left(N - \int_1^N e^{-\rho(x)s} dx\right)^n. \quad (14)$$

The parameter s is a measure of social capacity, or the amount of casual social interaction (including repeated interactions) in a characteristic time period. For simplicity we make the conservative approximation that s is universal for all individuals (see earlier footnote). The function $\rho(x)$ is a rank-probability distribution representing the probability of interacting with an individual at rank x in one's list of contacts sorted by contact frequency. Intuitively, $\rho(x)$ can be understood as representing a social interaction pattern: how often does one interact with one's most frequent contact vs one's second-most-frequent contact, third-most-frequent contact, etc. Importantly, it applies not just to close relationships, but also extends to casual contacts with whom one may not maintain any relationship; those are taken as seeds of new collaborations.

[4]We note that little evidence exists for this hypothesis: much work has been done on friendship and acquaintance networks (e.g., [31,35]), but little on scaling of casual contact or weak tie numbers. We expect our hypothesis of constant s to be conservative in the sense that, if there is some city size dependence, presumably $s(N)$ is an increasing function, which would result in even greater superlinearity than our model currently predicts.

Secondary correction

Other authors [36–38] have argued that the incentive to commit crime drops with city size; we incorporate this effect as a secondary correction to our model:

$$y(N) \sim Nu(N)^n N^{-0.12}$$
$$\sim N^{0.88}\left(N - \int_1^N e^{-\rho(x;N)s} dx\right)^n. \quad (15)$$

Note that this is simply a shifted version of the model in Eq. (14), and results reported below stay largely the same with either version (our theoretical predictions for power law fits are simply shifted globally by 0.12). See Supplemental Material [39] section 7 for details on this secondary correction.

III. RESULTS AND EMPIRICAL EVIDENCE

Even without assumption on the functional form of $\rho(x)$, Eq. (13) gives $du/dN \geqslant 0$ ([39], see section 8). This implies that individuals with identical social capacities and social interaction patterns will (on average) meet more unique individuals in more populated cities. This result is consistent with empirical findings from phone contact networks in a number of cities [30].

In order to get quantitative predictions, we need to make an assumption about the form of $\rho(x;N)$. Motivated by observations of Zipf's law scaling in a variety of rank distributions (e.g., word frequency, city population, earthquake magnitudes [40]), and direct evidence supporting the hypothesis that communication networks (such as emails, phone calls, and face-to-face interactions) have power-law like degree distribution [41–43], we assume $\rho(x)$ to have the following form:

$$\rho(x;N) = m(N)x^{-\alpha}, \quad (16)$$

where $m(N)$ is a normalization factor such that $m(N)\int_1^N x^{-\alpha}dx = 1$. We will fit the parameter α when validating with urban scaling data, and also use an independent dataset (the communication patterns in the Enron email corpus) to check that the parameter found is in a reasonable range. Note that we also consider other options for the algebraic form of $\rho(x)$ and find similar results (see [39], section 3).

The integral in Eq. (13), after plugging in Eq. (16) can be approximated as an incomplete Γ function $(sm)^{1/\alpha}\alpha^{-1}\Gamma(-\alpha^{-1}, mN^{-\alpha}s)$ ([39], see section 8). Combining that with Eq. (15), we reach a closed-form estimate for the scaling behavior of social output in a city of population N:

$$y(N) \sim N^{0.88}\left[N - \frac{(s\,m)^{1/\alpha}}{\alpha}\Gamma\left(-\frac{1}{\alpha}, m\,N^{-\alpha}s\right)\right]^n, \quad (17)$$

where $m = m(N) = (\alpha - 1)/(1 - N^{1-\alpha})$ if $\alpha \neq 1$; $m = 1/\ln(N)$ if $\alpha = 1$. The parameter n is the typical number of partners needed for an output. We input this parameter's value from data on average co-offending group size in the National Incident-Based Reporting System (NIBRS).

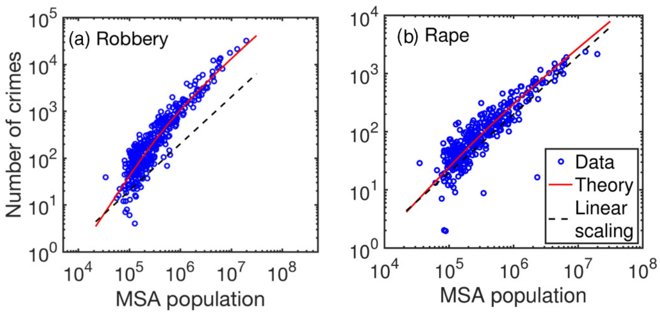

FIG. 4. *Model predictions with data.* (a) Number of robbery cases versus MSA population. (b) Number of rapes cases versus MSA population. Both for US in 2012. Red solid curves shows model predictions (not power-law), blue dots show data points, and the black dashed curves are reference lines of unit slope. Our model predicts superlinear scaling for robbery, and approximately linear scaling for rape. The difference between the model predictions in (a) and (b) is a result of the differing average co-offending group sizes. This figure uses co-offending group size calculated from the NIBRS dataset. The average group sizes are 1.74 for robbery and 1.29 for rape (group size is $n+1$). The parameters used in the model's prediction are $s = 2.63 \times 10^6$ and $\alpha = 0.93$ (global fit to all types of crimes).

Support from empirical data

Our model agrees well with US FBI data on all seven crime categories across 14 years; typical comparisons are shown in Fig. 4 (see ([39], section 5) for all comparisons) and a summary is shown in Fig. 5. The model explains not only the observed superlinear scaling for some urban outputs (e.g., robbery in Fig. 4(a)), but also close-to-linear scaling in others (e.g., rape in Fig. 4(b)). In Fig. 5 we show the relationship between the average co-offending group size and the degree of superlinearity (quantified for ease of comparison by the best fitting power-law exponent to the scaling relation minus one—though note that our model *does not* predict power-law scaling). The two panels in Fig. 5 use two independent sources of data for the co-offending group size—NIBRS in Fig. 5(a) and Chicago Police Department in Fig. 5(b). The Pearson correlation between average co-offending group size and superlinearity in the data is 0.764 (p-value 0.046) in Fig. 5(a), and 0.761 (p-value 0.047) in Fig. 5(b). The Spearman's rank correlation is 0.821 (p-value 0.034) in Fig. 5(a) and 0.786 (p-value 0.048) Fig. 5(b).

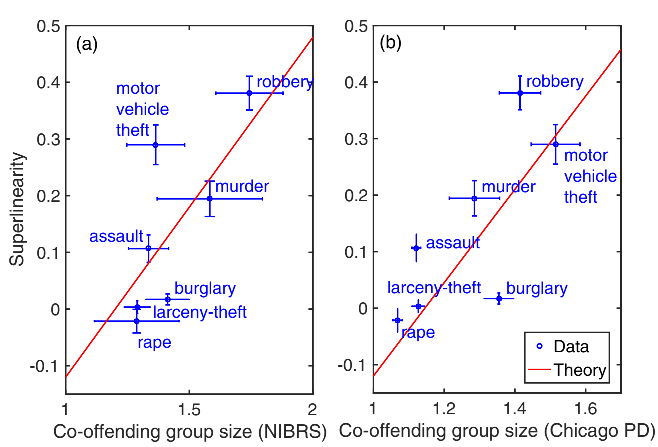

FIG. 5. *Superlinearity as a function of average group size for crimes.* Superlinearity is quantified as the exponent of the best-fit power law minus 1; this is for convenience—our model does not predict power-law scaling. Red solid line shows model prediction. Vertical error bars in both panels are standard deviation of year-to-year variation over 1999–2012. In (a), the horizontal error bars are state-to-state standard deviation in mean group sizes. In (b), the horizontal error bars are standard deviation of year-to-year variation over 1999–2012. Sources: [44,45].

1. Background on empirical data

Average co-offending group sizes for the seven types of crimes come from crime reports of two independent sources: the National Incident-Based Reporting System in 2014 and the Chicago Police Department arrest records in 1999–2012 (see [39], section 1) for more detail on those sources). Both sources report incident-level records for a variety of crime types. The co-offending group size of each incident is defined as the number of unique offenders reported in that incident. We average over all incidents of each type to reach the average group size. The parameter n, the number of partners is calculated by average group size minus 1. Note that average group sizes vary only over a small range for our crime data. This is primarily due to inherent limitations in offender reporting: for many crimes committed by groups, only a subset of the co-offenders are arrested or listed in crime reports (a single arrestee is the most common case). Despite this limitation, the data do show significant and consistent variation in group sizes, but the average values we use should be interpreted as correlates of, rather than direct estimates of, true co-offending group sizes.[5] To ensure the robustness of our results, we use both sources for model validation and find support of our prediction from both data sources.

2. Parameter fitting, model selection, and robustness

When comparing our model to data, we use a total of 98 crime scaling datasets (7 types × 14 years). Two parameters are assumed to be the same across all datasets, α and s, describing a social interaction pattern and "social capacity," respectively. Each dataset has another proportionality constant that is fitted. We find the values of the two global parameters (s and α) by minimizing the sum of the two-norm error across all 98 datasets. The parameter n is input from the average co-offending group size data for each type of crime. The best fitting global parameter pair was $s = 2.6 \times 10^6$, $\alpha = 0.93$ for the NIBRS dataset, and $s = 4.0 \times 10^6$, $\alpha = 0.69$ for the Chicago dataset. Our model performs better than power-law models (as measured by AIC and BIC [39, see section 6]) and has fewer fitted parameters,[6] suggesting that our

[5] As long as data-derived co-offending group sizes are monotonically increasing functions of the true co-offending group sizes, we expect correlations between model predictions and data to be preserved.

[6] We compare our model with the power law assumption ($y = aN^b$) for each data set. For k data sets, the power-law model requires fitting $2k$ parameters, while ours only requires $k+2$.

framework may be valuable for understanding this and similar phenomena.

We note that our model predictions are generally quite robust to the choice of $\rho(x)$. Any decreasing function $\rho(x)$ will imply $du/dN \geqslant 0$, leading to the prediction of superlinearity. The predictions are not particularly sensitive to the details of the particular form of that function—in addition to power laws, we also tested truncated log normals and a piecewise-constant function (motivated by the "circle of acquaintanceship" concept [46]), with nearly equivalent results ([39], see section 3 for details). To avoid overfitting, we do not attempt to explore the space of all possible (or all plausible) functions $\rho(x)$, we choose what we see as the simplest, the power-law.

IV. DISCUSSION

Some prior work has attributed superlinear scaling to the hypothesis of hierarchy in infrastructure and social networks [14,16], or differences in population density [15]. Our model suggests that a simple "finite-size effect," i.e., limited population to sample from in small to midsize cities, could be the key underlying mechanism. This may at first appear surprising, since even medium-sized cities in the U.S. include hundreds of thousands of unique individuals. At a plausible high rate of 100 "sampling events" per day, however, an individual would have nearly 1.5×10^6 samples after 40 years, more than the population of all but the largest U.S. cities (though of course the number of *unique* individuals met will be far less).

In our model, the finite-size effect reduces as city population becomes large: Eq. (1) implies that dy/dN decreases as N increases, and as $N \to \infty$, $dy/dN \to 1$. Data such as that shown in Fig. 1 display a reduced slope for the largest cities. This is consistent with the disappearance of this finite-size effect at the upper limits of U.S. city size. This suggests that, with limited resources, populating smaller cities would have a bigger impact on overall urban productivity than populating already big ones.

Some authors have taken alternative approaches to scaling of crime in cities, such as using a Bayesian framework [47] or looking at empirical connections between crime and other urban indicators [48]. Others have discussed how urban outputs such as crime may display long-term memory [49], how temporal clustering relates to crime scaling [13], or how crimes cluster geographically in cities [50]. Our model is not mutually exclusive with these others. However, most efforts continue to operate under the assumption of power-law scaling. We hope our work will encourage the study of crime outside of the power-law framework.

In [51,52], Clauset *et al.* argue that commonly used statistical approaches to fitting and testing power-laws can be problematic. Leitao *et al.* [53] recently examined a variety of power-law models in this context, showing that data were often inconsistent with models, and that many estimated exponents were not statistically distinguishable from 1. Our model was partially motivated by the (perhaps) over-dependence on power-law assumptions in the literature. The scaling we predict could explain poor fits of data to power-laws: inferred exponents would vary with the range of city size.

Depersin and Barthelemy [54], motivated by a longitudinal dataset on traffic congestion in cities, argue that scaling laws depend not only on population but also on growth history. Our model could also be generalized to incorporate such dependency—factors such as group size or social interaction patterns could show memory effects.

Our model, in agreement with previous models [14,15], implies that the dual aspects of cities are not separable: both positive and negative urban outputs (e.g., inventions and crimes) share common driving mechanisms rooted in social interaction. Future research—especially in the study of crime, law, deviance, and other sources of urban inequality—would do well to consider how scaling models such as ours might be further calibrated to capture differences within cities, especially across neighborhoods or communities.

This paper relies heavily on crime as an example because of the abundance and quality of data we were able to compile. However, the principle of the model can generalize to other urban outputs driven by forming collaborations, such as inventions and starting new businesses. We applied our model (with the same α and s parameters found by fitting to the crime data) to patent scaling laws, while extracting empirical average group sizes from patent co-authorship. We find good agreement between our model and patent scaling behavior (see [39, see section 9] for details). We did not include the patent group size in the comparison in Fig. 5 because the rate of under-reporting of group sizes likely differs between patents and crime.

V. CONCLUSIONS

The good agreement between our simple model and data indicates that differences in scaling relationships can indeed result from differences in the typical number of participants for an urban output: those outputs that are more "social" in nature are more strongly affected by city population. In agreement with previous models [14,15,19], we find that a fundamental driving mechanism of scaling in urban productivity is social interaction.

ACKNOWLEDGMENTS

The authors would like to thank the Chicago police department for making data available, and Sara Bastomski and Jennifer Wu for help with the co-offending dataset. This research was partially supported by the James S. McDonnell Foundation through Award No. 220020230.

[1] S. H. Strogatz, D. M. Abrams, A. McRobie, B. Eckhardt, and E. Ott, Nature **438**, 43 (2005).

[2] B. Eckhardt, E. Ott, S. H. Strogatz, D. M. Abrams, and A. McRobie, Phys. Rev. E **75**, 021110 (2007).

[3] J. L. Silverberg, M. Bierbaum, J. P. Sethna, and I. Cohen, Phys. Rev. Lett. **110**, 228701 (2013).

[4] I. Karamouzas, B. Skinner, and S. J. Guy, Phys. Rev. Lett. **113**, 238701 (2014).

[5] S. H. Lee, R. Ffrancon, D. M. Abrams, B. J. Kim, and M. A. Porter, Phys. Rev. X **4**, 041009 (2014).

[6] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[7] D. M. Abrams, H. A. Yaple, and R. J. Wiener, Phys. Rev. Lett. **107**, 088701 (2011).
[8] M. McCartney and D. H. Glass, Phys. A **419**, 145 (2015).
[9] J. Jacobs, *The Death and Life of Great American Cities* (Random House, New York, 1961).
[10] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, Proc. Natl. Acad. Sci. **104**, 7301 (2007).
[11] L. M. A. Bettencourt and J. Lobo, J. R. Soc., Interface **13**, 20160005 (2016).
[12] L. E. C. Rocha, A. E. Thorson, and R. Lambiotte, J. Urban Health **92**, 785 (2015).
[13] Q. S. Hanley, D. Lewis, and H. V. Ribeiro, PloS ONE **11**, e0149546 (2016).
[14] L. M. A. Bettencourt, Science **340**, 1438 (2013).
[15] W. Pan, G. Ghoshal, C. Krumme, M. Cebrian, and A. Pentland, Nat. Commun. **5**, 1961 (2013).
[16] S. Arbesman, J. M. Kleinberg, and S. H. Strogatz, Phys. Rev. E **79**, 016115 (2009).
[17] G. Duranton and D. Puga, Handb. Reg. Urban Econ. **4**, 2063 (2004).
[18] S. Arbesman and N. A. Christakis, Phys. A **390**, 2155 (2011).
[19] A. Gomez-Lievano, O. Patterson-Lomba, and R. Hausmann, Nat. Hum. Behav. **390**, 2155 (2017).
[20] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, J. R. Soc., Interface **12**, 20140745 (2015).
[21] *HIV Surveillance Reports, 2017* (US Centers for Disease Control and Prevention, Atlanta, GA, 2018), Vol. 29, http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html.
[22] K. B. LeFevre, *Invention as a Social Act* (SIU Press, Carbondale, Illinois, USA, 1987).
[23] E. Glaeser, *Triumph of the City: How Our Greatest Invention Makes US Richer, Smarter, Greener, Healthier and Happier* (Pan Macmillan, London, UK, 2011).
[24] M. Warr, *Companions in Crime: The Social Aspects of Criminal Conduct* (Cambridge University Press, New York, 2002).
[25] A. J. Reiss, Crime Justice **10**, 117 (1988).
[26] L. Stolzenberg and S. J. D'Alessio, J. Res. Crime Delinquency **45**, 65 (2008).
[27] M. S. Granovetter, Am. J. Sociol. **78**, 1360 (1973).
[28] C. Hauser, G. Tappeiner, and J. Walde, Reg. Stud. **41**, 75 (2007).
[29] E. Patacchini and Y. Zenou, Eur. Econ. Rev. **52**, 209 (2008).
[30] M. Schläpfer *et al.*, J. R. Soc., Interface **11**, 20130789 (2014).
[31] S. Milgram, Science **167**, 1461 (1970).
[32] F. Tricomi and A. Erdélyi, Pac. J. Math. **1**, 133 (1951).
[33] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature **484**, 96 (2012).
[34] P. Grindrod and D. J. Higham, Sci. Rep. **8**, 9737 (2018).
[35] R. A. Hill and R. I. Dunbar, Hum. Nat. **14**, 53 (2003).
[36] E. D. Gould, B. A. Weinberg, and D. B. Mustard, Rev. Econ. Stat. **84**, 45 (2002).
[37] J. Grogger, J. Labor Econ. **16**, 756 (1998).
[38] S. Machin and C. Meghir, J. Hum. Resour. **39**, 958 (1997).
[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.100.032306 for additional derivations and discussions.
[40] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).
[41] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103(R) (2002).
[42] W. Aiello, F. Chung, and L. Lu, Proceedings of the 32nd Annual ACM Symposium on the Theory of Computing (ACM, 2000), p. 171.
[43] C. Cattuto *et al.*, PloS ONE **5**, e11596 (2010).
[44] US Federal Bureau of Investigation, Crime in the United States (1999–2013).
[45] US Department of Justice. Federal Bureau of Investigation, Uniform crime reporting program data: National incident-based reporting system (2014), data retrieved from Ann Arbor, MI: Inter-university Consortium for Political and Social Research in March 2017, http://doi.org/10.3886/ICPSR36398.v1.
[46] R. I. Dunbar, Group Dyn.: Theor. Res. Pract. **12**, 7 (2008).
[47] A. Gomez-Lievano, H. Youn, and L. M. Bettencourt, PloS ONE **7**, e40393 (2012).
[48] L. G. Alves, H. V. Ribeiro, E. K. Lenzi, and R. S. Mendes, Phys. A **409**, 175 (2014).
[49] L. M. Bettencourt, J. Lobo, D. Strumsky, and G. B. West, PloS ONE **5**, e13541 (2010).
[50] M. Oliveira, C. Bastos-Filho, and R. Menezes, PloS ONE **12**, e0183110 (2017).
[51] A. Clauset, C. R. Shalizi, and M. E. Newman, SIAM Rev. **51**, 661 (2009).
[52] Y. Virkar and A. Clauset, Ann. Appl. Stat. **8**, 89 (2014).
[53] J. C. Leitao, J. M. Miotto, M. Gerlach, and E. G. Altmann, R. Soc. Open Sci. **3**, 150649 (2016).
[54] J. Depersin and M. Barthelemy, Proc. Natl. Acad. Sci. **115**, 2317 (2018).

Supplemental Material for
“Modeling the origin of urban output scaling laws”

V. Chuqiao Yang, Andrew V. Papachristos, and Daniel M. Abrams

Contents

1 Data Sources

US crime statistics by MSA

US crime statistics by MSA are obtained from Table 6 of the Federal Bureau of Investigation (FBI) publication *Crime in the United States* for years 1999–2012. The data can be accessed online at `https://ucr.fbi.gov/ucr-publications` (as of January 17, 2017). For each year, population, as well as crime rates for seven types of crimes (murder, forcible rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft) are reported for between 260 and 360 MSA's. We used a total of 98 data sets (14 years $\times$ 7 crime types).

Co-offending data from the Chicago Police Department

We estimate the group size of crime by compiling a new data set of arrest records for 352,705 crime incidents from the city of Chicago, IL from 1999 to 2012. The data were provided to one of the authors through a memorandum of understanding with the Chicago Police Department. Data are recorded at the incident level and include detailed information on each arrest, including the charge (e.g., motor vehicle theft, assault, robbery, etc.) as well as individual information on the offender(s). Co-offending is defined as two or more individuals being charged for the same offence as co-perpetrators, such as when two individuals steal a car together, sell drugs together, or rob someone together. We define "group size" as the number of offenders participating in a single crime, and without regard to any indication of a formal criminal group, such as a street gang. For example, if three people were involved in a motor vehicle theft, the group would have a size of three. Data used in the present analyses are derived from a cross-tabulation of offence type by group size for all offences from 1999 to 2012. Some summary of the dataset can be found in Table S1. A plot of mean group sizes by year and bootstrapped estimation of the 95% confidence intervals is shown in Figure S1.

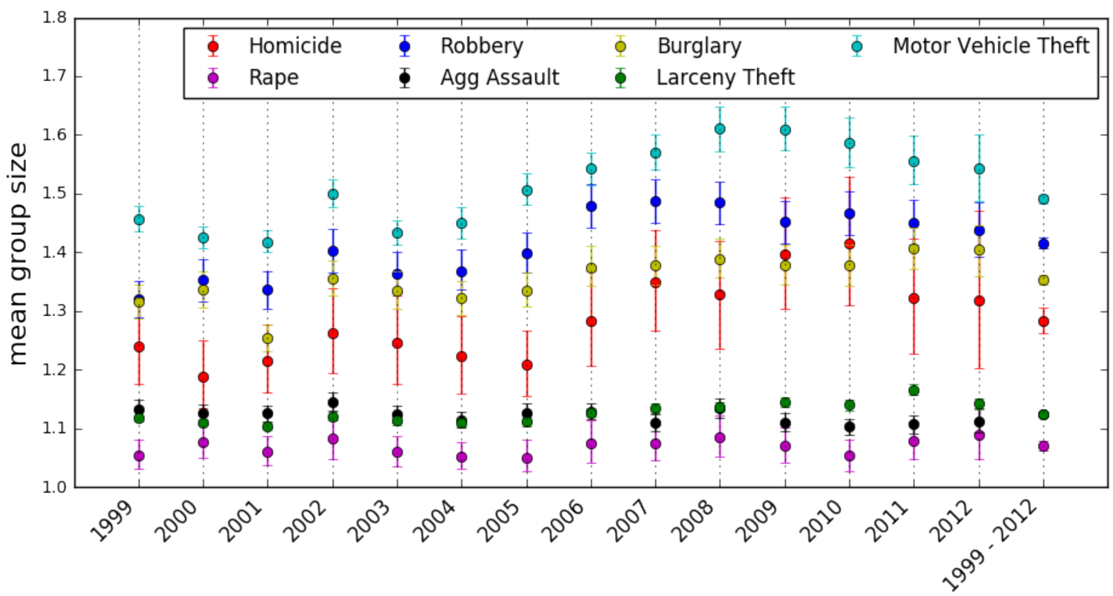

Figure S1: Mean group sizes (with 95% confidence interval as vertical error bars) for seven types of crimes over time extracted from the Chicago Police Department arrest records.

Co-offending group size from the National Incident-Based Reporting System

A dataset we use to extract co-offending group size is the Uniform Crime Reporting Program Data: National Incident-Based Reporting System (NIBRS) 2014 through the National Archive of Criminal Justice (`https://www.icpsr.umich.edu/icpsrweb/NACJD/studies/36398`), accessed in March 2017 [1]. We join the "offense" and "offender" tables of the data on incident number and state code. Based on descriptions in the codebook, we identify crime types of each instance by the UCR offence code using the following matching—murder: 09A, robbery: 120, rape: 11A, aggravated assault: 13A, burglary: 220, larceny-theft: 23A - 23H, and motor-vehicle theft: 240.

The co-offending group size for an offence is defined by the number of unique offenders reported in this offence. A total of 968,962 instances were available from this dataset for the categories of crimes of interest. A comparison of the values extracted from the NIBRS dataset with the Chicago dataset is shown in Table S1. Please note that the methodology of this dataset is different from that of the Chicago one: the Chicago dataset reports only arrested offenders, while the NIBRS includes voluntary instance reports from police departments, regardless of whether the offenders were arrested. Thus the values of co-offending group size are expected to have small systematic differences between these two datasets. Additionally, it is important to note that the Chicago Police Department does not participate in NIBRS, so the two datasets are mutually exclusive.

	Chicago PD arrest record	NIBRS
Homicide	1.28	1.58
Rape	1.07	1.29
Robbery	1.41	1.74
Agg. assault	1.12	1.33
Burglary	1.35	1.41
Larceny-theft	1.12	1.29
Motor vehicle theft	1.49	1.36
Year	1999 - 2012	2014
Number of instances	352,705	968,962

Table S1: Average co-offending group size for various types of crime in the Chicago PD arrest record and the NIBRS dataset.

The Enron corpus

We used email communication records from a corpus derived from Enron Corporation to check the plausibility of the parameter α in our rank-frequency distribution, ρ. The Enron corpus contains 517,431 emails sent by 151 employees of the Enron Corporation. This dataset was downloaded from `http://www.cs.cmu.edu/~enron/` in November 2014 [2].

2 Differences among scaling relationships

Table S2 summarizes the differences in some examples of urban outputs, including crimes broken down by category.

3 Parameter fitting and discussions

Our model has three parameters that must be fit from data: s, α, and a multiplicative prefactor. The first two parameters are assumed universal across all data sets, and the third

Urban output	Exponent	95% CI	Year
Larceny theft [3]	0.98	[0.95, 1.02]	2002
Rape [3]	0.98	[0.92, 1.04]	2002
Burglary [3]	1.01	[0.96 , 1.05]	2002
Total wages [4]	1.12	[1.09, 1.13]	2002
Aggravated assault [3]	1.13	[1.06 , 1.20]	2002
Murder [3]	1.22	[1.14, 1.30]	2002
New AIDS cases [4]	1.23	[1.18, 1.29]	2002–2003
New patents [4]	1.27	[1.25, 1.29]	2001
Motor vehicle theft [3]	1.32	[1.27 , 1.38]	2002
Robbery [3]	1.38	[1.31 , 1.45]	2002

Table S2: Best fitting power law exponent for scaling relations for various urban outputs. The differences among them can be pronounced.

is unique for each. For N sets of data, we thus require $N + 2$ parameters to be fitted. We find the values of the two global parameters (s and α) by minimizing the sum of the 2-norm error across all 98 datasets. The model parameter n is set by the average number of partners calculated from the average co-offending group sizes in data. Fig. S2 shows the landscape of 2-norm error in the neighbourhood of the optimal parameter pair (using NIBRS as the group size input). Interestingly, we observe a "valley" in the error landscape, showing that there may be an effective parameter that is a nonlinear combination of α and s. One potential explanation is that one can meet more distinct individuals by simply sampling more people (increasing s), or by change one's social interaction pattern to interact with more "weak ties" (decreasing α). We have also checked the robustness of the parameter fit by minimizing the 1-norm, and we found similar best fitting parameter ($s = 2.8 \times 10^6$, $\alpha = 0.93$), and we observe a similar error landscape.

Validating fitted α with Enron data

To check if the fitted parameters are plausible, we compare the power law exponent α for contact frequency that results from our parameter fit to an empirical estimate using the Enron email corpus. From the emails in the "sent_item" folder of each user, we extract sender and receivers' email addresses, as well as the length of each email in characters, excluding white spaces and quoted text in forwarded messages and replies. Some examples of the length of communication to contact vs rank of the contact are shown in Fig. S3-left.

We excluded those senders with too few contacts (< 100). The communication length vs rank relation may be approximated as a power law, with the exception of the low-frequency contacts at larger ranks. We used the communication length vs. rank relation (up to rank 100) to fit the power law exponent $-\alpha$ for each sender. Some examples of the relations are shown in Fig. S3-right. The distribution of those α is shown in Fig. S4. The best fitting α from fitting the model to crime scaling data, while using the Chicago data as input group size is 0.69. The best fitting α while using the NIBRS group size as input, is 0.93, both are plausibly consistent with the distribution.

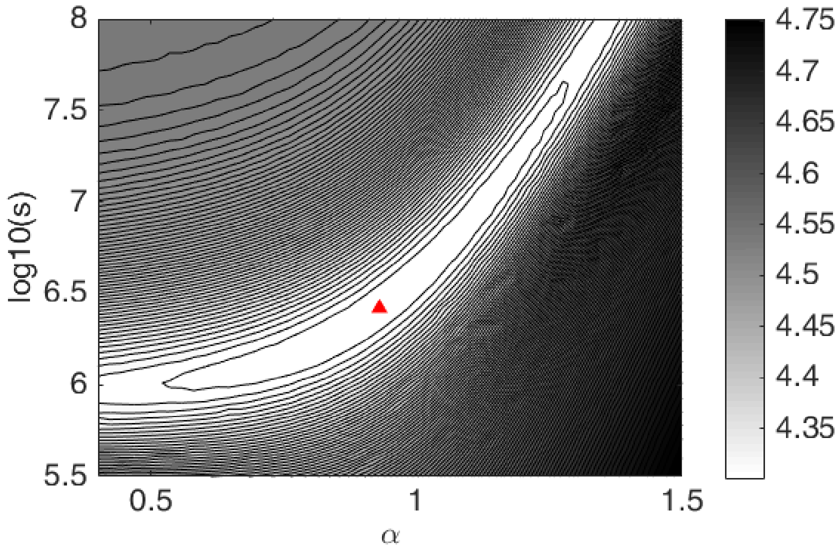

Figure S2: Error landscape for the global parameter optimization problem. Contours show lines of constant 2-norm for the difference between theory and data over all datasets. The red triangle shows optimized parameters giving the best fit.

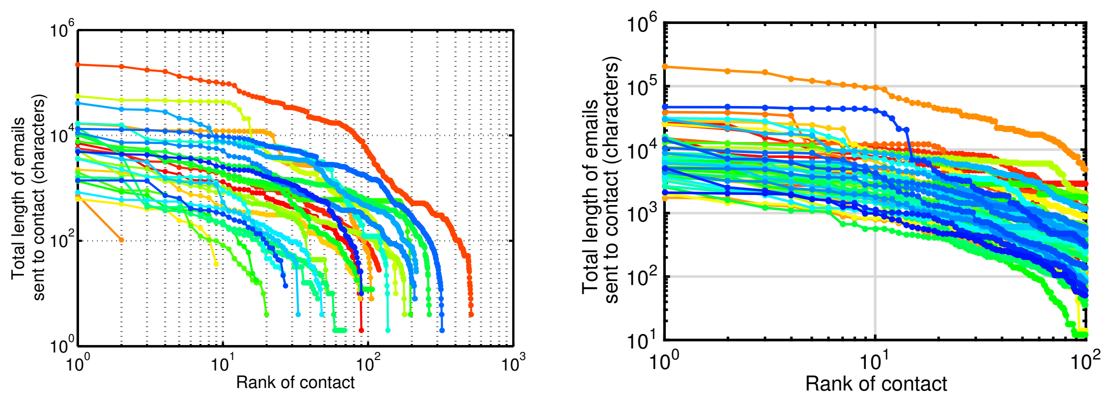

Figure S3: Left: Examples of the amount of contact vs. rank for 30 randomly selected individuals in Enron database. Right: The amount of contact vs. rank of contact distribution for the individuals in the dataset with more than 100 contacts. The curves for small ranks can be roughly approximated as power laws, though not perfectly.

Discussion of the functional form of $\rho(x)$

When comparing our model with data, we assume $\rho(x) \sim x^{-\alpha}$. The motivation is to avoid assumptions about the structure of social networks, since that is a complex question on its own and is not central to this paper's discussion. Although some network structures can result in $\rho(x) \sim x^{-\alpha}$, it is not a necessary condition to reach our major conclusions. It is important to note that relaxing this assumption to general non-increasing $\rho(x)$ (which is by definition true for a rank-probability distribution) does not affect the fundamental predictions of this work—people in larger cities meet more unique individuals, and the more participants an act requires, the greater the superlinearity. Other choices produce similar

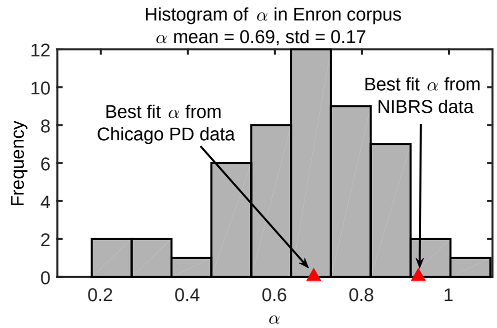

Figure S4: Histogram of fitted α for all individuals with more than 100 contacts in the Enron email database, compared with α found by fitting our model to crime scaling data while using two independent datasets as group size input. Both *alpha* values found are plausibly consistent with the distribution.

qualitative results. Figure S5 compares predictions using some alternative functional forms of ρ with data. These alternative functional forms are also able to predict both the superlinear scaling (left column of Figure S5) and nearly linear scaling (right column of Figure S5).

Figure S5-A and B assume ρ to be a decaying log normal function,

$$\rho(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right].$$

To ensure that $\rho(x)$ is a non-increasing function, we set $\rho(x)$'s maximum to be at $x = 1$, the starting point of the 1D rank space, which constrains the parameters to be $\mu = \sigma^2$. Figure S5-A and B shows scaling predictions for parameters $\mu = \sigma^2 = 10$, and $s = 2.63 \times 10^6$ (same s as used in the main text).

In Figure S5-C and D, we assume ρ to be a piecewise constant function representing the "circles of acquaintanceship" as proposed by Dunbar [5], with four discrete jumps in interaction probability. We use Dunbar's estimation and set the sizes of the circles to be 5, 15, 50 and 150 people. We assume the interaction frequency decreases by a factor of 10 when moving from one circle to the next:

$$\rho(x) = m(N) \times \begin{cases} 10^4 & 1 \leq x \leq 5 \\ 10^3 & 5 < x \leq 15 \\ 10^2 & 15 < x \leq 50 \\ 10^1 & 50 < x \leq 150 \\ 10^0 & 150 < x \leq N \end{cases},$$

where $m(N)$ is a multiplicative normalization factor so that $\int_1^N \rho(x)dx = 1$. Figure S6 visualizes these alternative ρ functions and compares with the power-law assumption in the main text. Figure S5-C and D also uses the same s parameter as in the main text.

Generalizing beyond these numerical simulations, Section 8 demonstrates that $u(N)$ must be a non-decreasing function of N even without any assumption regarding the functional form for ρ. Thus the qualitative conclusions of the main text hold for general ρ.

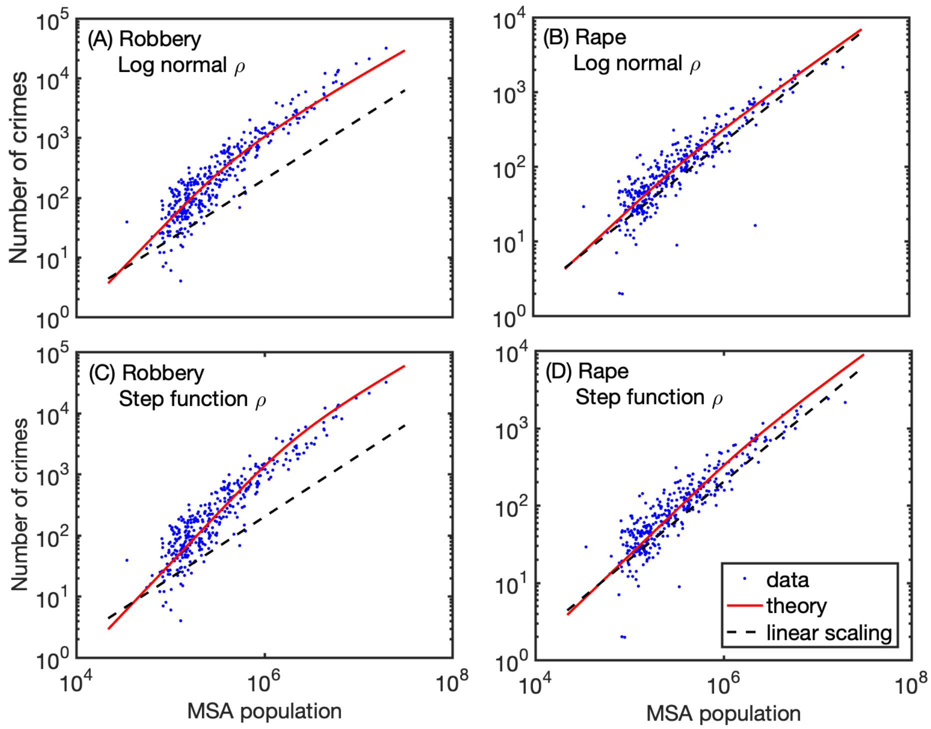

Figure S5: Example results for an alternative functional forms of ρ, where ρ takes the form of (A, B) a log normal function, or (C, D) a piecewise constant function denoting circles of acquaintanceship with decaying likelihood of interaction in each circle.

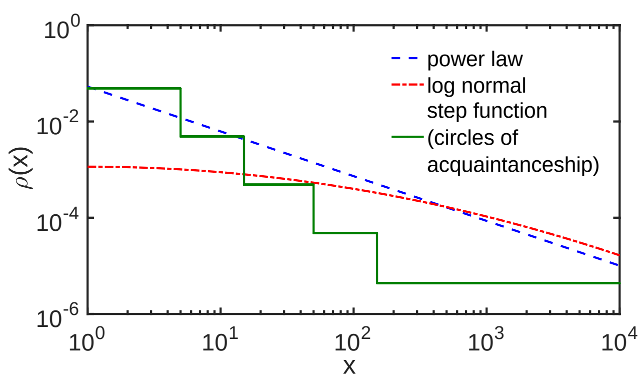

Figure S6: Comparing the power-law function ρ with alternative functional forms that generate results as shown in Figure S5. Parameter values used here are the same as those used to generate the scaling predictions.

4 Calculation of best-fitting power law exponent

We quantify the degree of superlinear scaling using the best fitting power-law exponent in the scaling relationship of MSA total output vs. MSA population, consistent with Bettencourt et. al [4]. The exponent is the slope of the linear fit $\log(y) = a_1 \log(x) + a_2$, where x and y are the MSA population (horizontal axis) and amount of output (vertical axis) respectively.

	AIC	BIC
power law model	5.34e+05	5.36e+05
our model	5.16e+05	5.17e+05

Table S3: AIC and BIC of the two types of models with best fitting parameters

The parameters a_1 and a_2 are to be fitted from the data; a_1 is the fitted power law exponent, and thus $a_1 - 1$ is the superlinearity shown in Fig. 4 of the main text.

The superlinearity for the theory lines in Fig. 4 is calculated as follows. Since the theory does not predict a power law relationship, we first calculate the theory's prediction of total output for each MSA population in the data. We then fit those predictions to a power law in the same way we fit the data to arrive at the estimate of superlinearity.

The authors are aware of the criticisms for linear fit on the log scale [6], and are motivated by them to develop the non-power-law model described in the main text. The purpose of using the linear fit is not to claim a power-law relationship. Instead, we use it to arrive at an indicator to assist us comparing the steepness of the scaling relationship in the data and in the theory, and to visualize the relationship between group size and the steepness of the scaling relationship.

5 Model's fits to all seven types of crimes

Fig. S7 shows our theory's comparison with the 7 categories of crime in year 2012, the most recent year of data available when the research was performed, with the group size input from the NIBRS dataset. In addition, Fig. S8 shows the fit to patent data (with sources and group size input described in Sec. 9, showing year 2000, the most recent year available), suggesting that the model may be generalized to urban outputs beyond crime. Please note that for all the data displayed, the theoretical curves use the same universal set of parameters α and s. The only parameter fitted separately to each data set is a multiplicative scaling factor (see Sec. 3).

6 Comparison with power-law models

A recently published study [7] also hypothesizes that variation may result from the need for a number of complementary factors to come together, but assumes the scaling take on a power-law form. Our work relaxes the power-law scaling assumption, and we present empirical evidence that validates our hypotheses.

We compare our model with the power law assumption ($y = aN^b$) for each data set. The power law model uses $2N$ parameters for N data sets. Our model uses $N + 2$ parameters for N datasets. We use the Akaike and Bayesian Information Criteria (AIC and BIC) to measure the goodness of fit. The result is shown in Table S3. Our model has lower AIC and BIC values than the power law model, thus reaches better fit with data, accounting for the parameters used (evidence ratio based on AIC is $\exp(\frac{1}{2}\Delta\text{AIC}) \approx \exp(9000) \sim 10^{3900}$ [8], strongly supporting our simpler model).

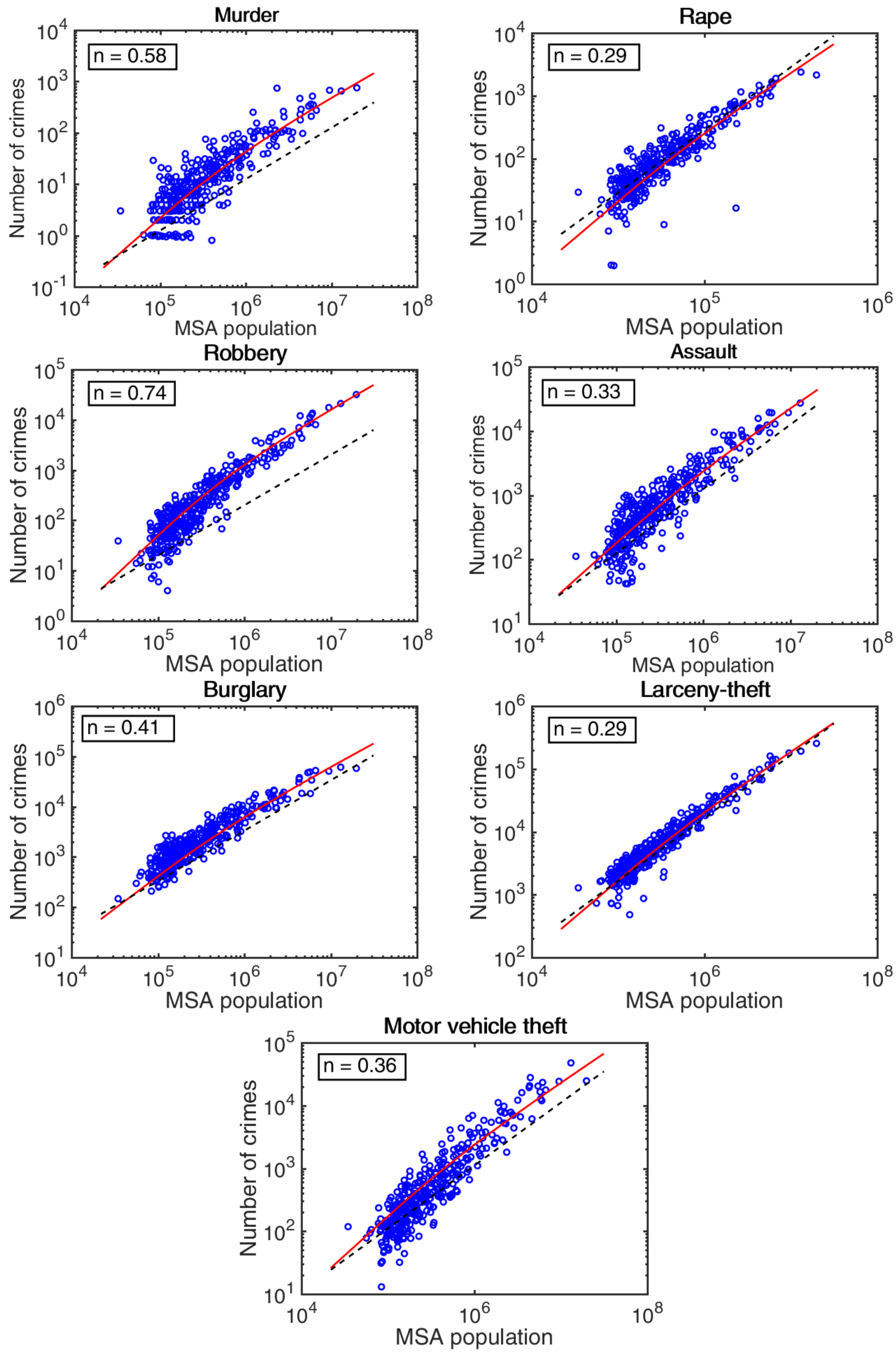

Figure S7: The fits of the model to all seven types of crimes and to patent data. Fits for other years are similar. In all panels, the blue dots are data points, red curves are model predictions, and the black dashed lines show linear scaling. The average number of partners n is indicated on each panel.

7 Secondary correction

Previous studies support the idea that higher wages can lead to a lower crime rate [9, 10, 11]. One theory regarding crime incentives suggests that if the gain from committing a crime exceeds the wage one would otherwise earn with the same time and effort, one would be inclined to commit a crime. Historical data support that at least some young men's behaviour is responsive to this type of crime incentive [10]. Motivated by these findings, we include a secondary correction in our prediction:

$$y(N) \propto N \cdot u^n(N) f(w(N))$$

where $w(N)$ is the average wage one would earn in a city of population N. Empirical data suggest that $w(N) \sim N^{0.12}$ [12]. The propensity to commit a crime decreases with increasing mean wage, so as a simple approximation we take f to be the function $f(w) = 1/w$. Then

$$\begin{aligned} y(N) &\sim N \cdot u^n(N) N^{-0.12} \\ &\sim N^{0.88} \left(N - \int_1^N e^{-\rho(x;N)s} dx \right)^n . \end{aligned} \tag{1}$$

We use this model equation for comparison with the crime data sets.

8 Additional mathematical derivations

Integral Approximation

Here, we find an asymptotic approximation of the integral in the expression of $u(N)$ of the main text to simplify computation of the model. We use this asymptotic approximation when validating the model with empirical data.

In main text Eq. (3), we have

$$u(N) = N - \int_1^N e^{-\rho(x)s} dx . \tag{2}$$

We would like to approximate the integral

$$I_1(N) = \int_1^N e^{-\rho(x)s} dx = \int_1^N e^{-mx^{-\alpha}s} dx ,$$

where $m = m(N)$ is a normalization constant with $m = (\alpha - 1)/(1 - N^{1-\alpha})$ if $\alpha \neq 1$; $m = 1/\ln(N)$ if $\alpha = 1$.

First, we make a change of variable, $t = x^{-\alpha}s$. Then we have

$$I_1(N) = \frac{1}{\alpha} s^{1/\alpha} \int_{N^{-\alpha}s}^{s} e^{-mt} t^{-\frac{\alpha+1}{\alpha}} dt . \tag{3}$$

Now we focus on the integral in (3),

$$I_2(N) = \int_{N^{-\alpha}s}^{s} e^{-mt} t^{-\frac{\alpha+1}{\alpha}} dt . \tag{4}$$

Then we have the relation $I_1 = s^{1/\alpha} I_2/\alpha$. Substitute $k = (\alpha+1)/\alpha$, $(k > 1)$ in (4), we have

$$I_2(N) = \int_{N^{-1/(k-1)}s}^{s} e^{-mt} t^{-k} dt \ . \tag{5}$$

Substitute $\epsilon = mN^{-1/(k-1)}s$ and $\tau = mt$, into (5), we have

$$I_2(\epsilon) = m^{k-1} \int_{\epsilon}^{ms} e^{-\tau} \tau^{-k} d\tau \ .$$

Note that the integrand $e^{-\tau}\tau^{-k}$ diverges at $\tau \to 0$ for all $k > 0$, and more importantly, it diverges with a heavy head for all $k > 1$. So the neighborhood of ϵ dominates the integral, and we can replace the upper bound by infinity with error that's only exponentially small:

$$I_2(\epsilon) \approx m^{k-1} \int_{\epsilon}^{\infty} e^{-\tau} \tau^{-k} d\tau \ . \tag{6}$$

The integral in (6) is the upper incomplete gamma function:

$$I_2(\epsilon) = m^{k-1} \Gamma(1-k, \epsilon) \ .$$

The upper incomplete gamma function, $\Gamma(a, z)$ is defined as,

$$\text{if } a > 0 : \quad \Gamma(a, z) = \int_z^{\infty} e^{-t} t^{a-1} dt \ ,$$
$$\text{if } a < 0 : \quad \Gamma(a, z) = \Gamma(a+1, z) - \frac{z^a}{a} e^{-x} \ .$$

The $a < 0$ recurrence relation is found using integration by parts.

Undoing the variable transformations, we have

$$I_1(N) = \frac{(s\ m)^{1/\alpha}}{\alpha} \Gamma(-\frac{1}{\alpha}, m\ N^{-\alpha} s) \ ,$$

so

$$u(N) = N - \frac{(s\ m)^{1/\alpha}}{\alpha} \Gamma(-\frac{1}{\alpha}, m\ N^{-\alpha} s) \ . \tag{7}$$

Show $du/dN \geq 0$

Here we show that $du/dN \geq 0$ is implied even without the assumption of power law ρ. Let ρ take on any separable form, $\rho(x; N) = m(N) f(x)$, where $f(x) \geq 0$ represents a communication pattern that is universal across cities, and $m(N)$ is a normalization factor. This assumption considers the case where individuals of identical social interaction pattern reside in cities of different sizes.

Note that the population in the 1-D space is ordered by decreasing probability of interaction. Thus, by definition, $f'(x) \leq 0$.

From the main text Eq. (6), we have

$$u(N) = N - \int_1^N e^{-\rho(x;N)s} dx \ . \tag{8}$$

Since the integral is dominated by small ρs values, we approximate (8) with a Taylor expansion near $\rho s = 0$:

$$\begin{aligned} u(N) &\approx N - \int_1^N \left(1 - \rho(x;N)s + \frac{1}{2}(\rho(x;N)s)^2\right) dx \\ &= 1 + s\int_1^N \rho(x;N)dx - \frac{s^2}{2}\int_1^N \rho^2(x;N)dx \\ &= 1 + s - \frac{s^2}{2}\int_1^N \rho^2(x;N)dx \ . \end{aligned} \tag{9}$$

Then

$$\frac{du}{dN} = -\frac{s^2}{2}\frac{d}{dN}\int_1^N \rho^2(x;N)dx \ . \tag{10}$$

Using Leibniz's rule for differentiation under the integral sign, we have

$$\frac{du}{dN} = -\frac{s^2}{2}\left[\int_1^N \frac{d}{dN}\rho^2(x;N)dx + \rho^2(N,N)\right] \ . \tag{11}$$

Let $\rho = m(N)f(x)$, denote $Q = \int_1^N \frac{d}{dN}\rho^2(x;N)dx + \rho^2(N,N)$. Then

$$\frac{d}{dN}\rho^2(x;N) = 2f^2(x)m(N)\frac{dm(N)}{dN} \ ,$$

and

$$Q = 2m\frac{dm}{dN}\int_1^N f^2(x)dx + f^2(N)m^2(N) \ . \tag{12}$$

Since the normalization factor m satisfies

$$m(N) = \frac{1}{\int_1^N f(x)dx} \ ,$$

then

$$\frac{dm}{dN} = -\frac{f(N)}{(\int_1^N f(x)dx)^2}.$$

Substituting into (12), we have

$$\begin{aligned} Q &= -2\frac{f(N)}{(\int_1^N f(x)dx)^3}\int_1^N f(x)^2dx + \frac{f^2(N)}{(\int_1^N f(x)dx)^2} \\ &= \frac{f(N)\left[f(N)\int_1^N f(x)dx - 2\int_1^N f^2(x)dx\right]}{(\int_1^N f(x)dx)^3} \ . \end{aligned}$$

Define $S(N) = f(N)\int_1^N f(x)dx - 2\int_1^N f^2(x)dx$ (the bracketed expression above). Since $f(x) \geq 0$ and ρ normalizable, the denominator of Q must be positive. In order to show

$du/dN = -s^2Q/2 \geq 0$, we need to show $S \leq 0$. It's clear that $S(1) = 0$. Now it suffices to show $dS/dN \leq 0$:

$$\begin{aligned} \frac{dS}{dN} &= f'(N)\int_1^N f(x)dx + f^2(N) - 2f^2(N) \\ &= f'(N)\int_1^N f(x)dx - f^2(N) \ . \end{aligned} \tag{13}$$

By definition, $f(x)$ is a non-increasing function, i.e., $f'(N) \leq 0$. Examining (13) with this in mind, we observe that

$$\frac{dS}{dN} \leq 0.$$

So $S(N) \leq 0$ for all $N > 1$ and therefore $Q \leq 0$ for all $N > 1$. Thus

$$\frac{du}{dN} = -\frac{s^2}{2}Q \geq 0 \ .$$

9 Model's application to patent scaling

In the main text, we discussed possible extensions of our model to urban outputs beyond crime. In this section, we show results comparing our model (with parameters found by fitting to the crime data) to the scaling of patents, and find good agreement.

We first compute the average group size for patents from the National Bureau of Economic Research (NBER) database [13]. The data were downloaded from http://www.nber.org/patents in July 2016. The data comprise detailed information on almost 3 million U.S. patents granted between January 1963 and December 1999. Here the group size is defined as the number of authors on a patent. Note that this dataset does not aggregate patent information by MSA. We also extract the patent scaling data (number of patents by MSA) from Bettencourt et al. [14] Fig. 1 using the GetData Graph Digitizer software.

We use the average group size found in the NBER database as the input of group size $(n + 1)$ in our model. We also use the best fitting global parameters found in the crime data set ($s = 2.8 \times 10^6$, $\alpha = 0.93$) when fitting to the patent dataset. The comparison of our theory with the empirical patent data is shown in Fig. S8. Our model predicts similar superlinear scaling behavior shown in the data, suggesting that our model can be generalized to urban outputs beyond crime.

10 Code and data availability

Code used in parameter fitting and generating the scaling laws fit figures, as well as the data input needed, can be accessed from the repository: `https://github.com/vc-yang/urban_productivity_scaling_laws`.

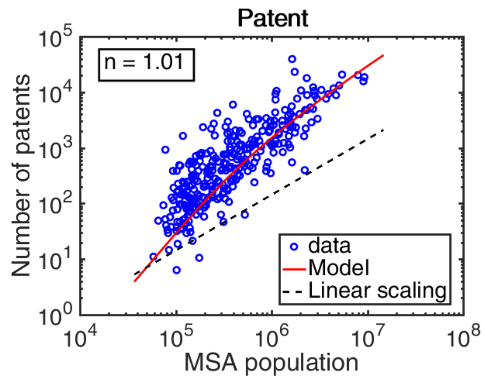

Figure S8: The fit of the model to patent scaling data in year 2000, using the best fitting α and s parameters found by fitting to crime scaling and co-offending group sizes in the NIBRS dataset.

References

[1] US Department of Justice. Federal Bureau of Investigation, "Uniform crime reporting program data: National incident-based reporting system," 2014. Data retrieved from Ann Arbor, MI: Inter-university Consortium for Political and Social Research in March 2017. http://doi.org/10.3886/ICPSR36398.v1.

[2] W. W. Cohen, "Enron email dataset," 2014. Data retrieved from `http://www.cs.cmu.edu/~enron/`.

[3] US Federal Bureau of Investigation, "Crime in the United States, 1999 – 2013," 1999 – 2013.

[4] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences*, vol. 104, no. 17, pp. 7301–7306, 2007.

[5] R. I. Dunbar, "Cognitive constraints on the structure and dynamics of social networks.," *Group Dynamics: Theory, Research, and Practice*, vol. 12, no. 1, p. 7, 2008.

[6] J. C. Leitao, J. M. Miotto, M. Gerlach, and E. G. Altmann, "Is this scaling nonlinear?," *Royal Society Open Science*, vol. 3, no. 7, p. 150649, 2016.

[7] A. Gomez-Lievano, O. Patterson-Lomba, and R. Hausmann, "Explaining the prevalence, scaling and variance of urban phenomena," *Nature Human Behavior*, vol. 390, no. 11, pp. 2155–2159, 2017.

[8] E.-J. Wagenmakers and S. Farrell, "AIC model selection using Akaike weights," *Psychonomic Bulletin & Review*, vol. 11, no. 1, pp. 192–196, 2004.

[9] E. D. Gould, B. A. Weinberg, and D. B. Mustard, "Crime rates and local labor market opportunities in the united states: 1979–1997," *Review of Economics and Statistics*, vol. 84, pp. 45–61, 2002.

[10] J. Grogger, "Market wages and youth crime, no. w5983," *National Bureau of Economic Research*, 1997.

[11] S. Machin and C. Meghir, "Crime and economic incentives," *Journal of Human Resources*, vol. 39, no. 4, pp. 958–979, 2004.

[12] L. M. A. Bettencourt, "The origins of scaling in cities," *Science*, vol. 340, no. 6139, pp. 1438–1441, 2013.

[13] B. H. Hall, A. B. Jaffe, and M. Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," tech. rep., National Bureau of Economic Research, 2001.

[14] L. M. A. Bettencourt, J. Lobo, and D. Strumsky, "Invention in the city: Increasing returns to scale in metropolitan patenting," *Los Alamos National Laboratory technical Report LAUR-04-8798*, 2004.