JOÃO VITOR MEIRELLES DE MIRANDA

Leis de Escala em Cidades

São Paulo

2015

JOÃO VITOR MEIRELLES DE MIRANDA

Leis de Escala em Cidades

Versão Corrigida

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos.

Área de concentração: Sistemas Complexos

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 13 de Novembro de 2015. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Camilo Rodrigues Neto

São Paulo

2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO
(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Miranda, João Vitor Meirelles de
Leis de escala em cidades / João Vitor Meirelles de Miranda ; orientador, Camilo Rodrigues Neto. – São Paulo, 2015.
97 p : il

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Modelagem de Sistemas Complexos, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo.
Versão corrigida

1. Cidades - Estudo. 2. Urbanização. 3. Planejamento territorial urbano. 4. Sistemas dinâmicos. I. Rodrigues Neto, Camilo, orient. II. Título

CDD 22.ed. – 307.76

Dissertação de autoria de João Vitor Meirelles de Miranda, sob o título **"Leis de Escala em Cidades"**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de sistemas Complexos, na área de concentração Sistemas Complexos, aprovada em 13 de Nobembro de 2015 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Camilo Rodrigues Neto
Presidente
Instituição: EACH/USP

Prof. Dr. Fernando Fagundes Ferreira
Instituição: EACH/USP

Prof. Dr. Fabiano Lemes Ribeiro
Instituição: DFI/UFLA

Agradecimentos

Agradeço profundamente ao meu orientador, Camilo Neto, e aos demais professores e colegas do programa. Extendo meus agradecimentos aos outros professores a amigos que encontrei pelo caminho. Agradeço, ainda, ao CNPq pelo fomento e apoio à produção dessa pesquisa.

Agradeço, por fim, ao tempo e ao espaço, por tornarem esse momento, assim como todos os outros, possível.

"...as cenas se repetem de maneira diferente, se multiplicam, se conurbam, umas às outras, formam-se cidades dentro da cidade, enquanto a fala de um com o outro, se dá mediada pelo troco..."

(Acareação - Germano Dushá)

Resumo

MEIRELLES, Joao. **Leis de Escala em Cidades**. 2015. 96 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015.

Ao longo da história, diversas foram as tentativas da ciência em sistematizar o conhecimento sobre as cidades. Um conjunto de recentes descobertas empíricas deu início a uma nova ciência urbana, fundamentada nos sistemas complexos(BATTY, 2013). Uma das raízes empíricos dessa ciência está nas análises de escalamento entre diversas variáveis urbanas com a população das cidades. Como hipótese fundamental, propõe-se que por mais diferentes que sejam, parece existir um padrão muito claro de escalamento entre a população dos centros urbanos com variáveis de produção socioeconômica (superlinear) e variáveis infraestruturais (sublinear). Um modelo de campo médio proposto por Bettencourt(BETTENCOURT, 2013) parece apresentar uma prossível explicação para a origem dessas leis de escala nas interações sociais. Essa hipótese ainda carece, entretanto, de maiores comprovações empíricas e de maiores explorações do modelo proposto. O presente trabalho apresenta os resultados de uma exploração das leis de escala entre as cidades brasileiras e uma tradução do modelo de Bettencourt em uma estrutura de modelagem por agentes. Os padrões de escalamento das cidades brasileiras parecem seguir aqueles encontrados em outras cidades do mundo com certa robustez. Escalamentos de variáveis de arrecadação e de despesas foram estudados e as cidades brasileiras parecem otimizá-los a medida que crescem. O modelo se mostrou coerente com os fatos observados empiricamente e indicou que cidades muito desiguais tendem a ter menores produções socioeconômicas e que áreas de interações maiores e custos de transporte menores tendem a produzir mais interações socioeconômicas.

Palavras-chaves: Ciência das Cidades. Sistemas Complexos. Leis de Escala. Modelagem Baseada em Agentes

Abstract

MEIRELLES, Joao. **Scaling Laws in Cities**. 2015. 96 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2015.

Throughout history science has tried in many ways to sistematyze the knowledge about cities. Recent empirical discoveries started a new urban science, based on complex systems and data science(BATTY, 2013). One of the empirical foundation of this science is the scaling laws of diferent urban variables in relation to the urban population size. As a main hypothesis, it is sugested that as diferent as cities may be, there seems to be an evident scaling pattern between population size and socioeconomic production (superlinear) and infrastrufture variables (sublinear). A mean field model proposed by Bettencourt(BETTENCOURT, 2013) appears to present a plausible explanation for the origin of scaling laws in social interactions. However, this hypothesis still lacks more concluding empirical proof and further study of the model itself. This paper presents the results of an exploration of the scaling laws among Brazilian cities and a translation of Bettencourt's model in a ABM framework. The scaling patterns of Brazilian cities appear to follow those found in other cities in the world with a certain robustness. Tax revenues and costs scaling were studied and Brazilian cities seem to optimize them as they grow. The model proved to be consistent with the facts observed empirically and indicated that very unequal city tend to have lower socioeconomic productions and greater areas of interactions and lower transportation costs tends to imply greater socioeconomic productions.

Keywords: Science of Cities. Complex Systems. Scaling Laws. Agent Based Modelling

Lista de figuras

Figura 1 – Distribuição Normal
fonte:produção do próprio autor 18
Figura 2 – Distribuição de Lei de Potência
fonte:produção do próprio autor 19
Figura 3 – A média na distribuição normal e na distribuição de lei de potência
fonte:produção do próprio autor 20
Figura 4 – Escalamento do lado L (linha bege), da área A (linha roxa) e do volume V (linha azul) de um cubo em eixos normais e logarítmos
fonte:produção do próprio autor 23
Figura 5 – Escalamento metabólico de animais
fonte: (WEST; BROWN, 2000) 26
Figura 6 – Urbanização no mundo
fonte: (UN-HABITAT, 2014) 27
Figura 7 – Características básicas do escalamento urbano
fonte: (BETTENCOURT, 2013) 29
Figura 8 – Indicação de Universalidade Espacial. Expoentes de escalamento super-lineares e relativamente próximos são encontrados em distintos países
fonte: (BETTENCOURT, 2013) 31
Figura 9 – Indicação de Universalidade Temproal. O expoente de escalamento do PIB com a população das cidades americanas se mantém praticamente constante ao longo de um intrvalo de 40 anos
fonte: (BETTENCOURT, 2013) 32
Figura 10 – Panorama das Leis de Escalas Encontradas em Cidades
fonte: (BETTENCOURT, 2013) 33
Figura 11 – Índice G para as cidades americanas
fonte: (BETTENCOURT, 2013) 34
Figura 12 – Caracteristicas Basicas de Escalamento Urbano
fonte: (SCHLAPFER et al., 2014) 35
Figura 13 – Síntese das Leis de Escalas Encontradas em Cidades
fonte: (BETTENCOURT, 2013) 39
Figura 14 – Histogramas das variáveis estudadas. fonte:produção do próprio autor . 43

Figura 15 – Escalamento entre a População e o PIB das cidades estudadas. 44
Figura 16 – Escalamento entre a População e infraestrutura viária das cidades estudadas. 45
Figura 17 – Escalamento entre a População e á Área das cidades estudadas. 45
Figura 18 – Escalamento entre a População e o G das cidades estudadas fonte:produção do próprio autor 46
Figura 19 – Oriximiná-PA: relação entre área do município e seu centro urbano fonte:adaptado de orixi.wordpress.com 47
Figura 20 – Expoente de escalamento referente aos valores mínimos de corte para densidade e população da base de dados de Cidades (esquerda) e de Regiões Metropolitanas (direita) fonte:produção do próprio autor 48
Figura 21 – inclinação α da função de regressão linear encontrada para cada uma das variáveis básicas, referente a uma densidade mínima de corte nos bancos de dados estudados fonte:produção do próprio autor 50
Figura 22 – r^2 da função de regressão linear referente a uma densidade mínima de corte da base de dados de cidades encontrada para cada uma das variáveis básicas fonte:produção do próprio autor 51
Figura 23 – Escalamento entre a População e o PIB das cidades estudadas após a adoção do critério de corte fonte:produção do próprio autor 52
Figura 24 – Escalamento entre a População e a Área das cidades estudadas após a adoção do critério de corte fonte:produção do próprio autor 53
Figura 25 – Escalamento entre a População e infraestrutura viária das cidades estudadas após a adoção do critério de corte fonte:produção do próprio autor 53
Figura 26 – Escalamento entre a População e do G das cidades estudadas após a adoção do critério de corte fonte:produção do próprio autor 54

Figura 27 – Esquematização da visão geral do modelo proposto
fonte:produção do próprio autor 61
Figura 28 – Esquematização de um passo de tempo do modelo
fonte:produção do próprio autor 62
Figura 29 – Exemplo de condição inicial do modelo
fonte:produção do próprio autor 63
Figura 30 – Evolução temporal de uma cidade pouco densa
fonte:produção do próprio autor 64
Figura 31 – Evolução temporal de uma cidade muito densa
fonte:produção do próprio autor 65
Figura 32 – Taxa de morte referente àos valores iniciais de População e Área
na versão onde o custo de transporte é invariante com a densidade
fonte:produção do próprio autor 66
Figura 33 – Taxa de morte referente àos valores iniciais de População e Área
na versão onde o custo de transporte é uma função da densidade
fonte:produção do próprio autor 67
Figura 34 – expoentes entre Y e N para diversas parametrizações
fonte:produção do próprio autor 68
Figura 35 – expoentes entre Y e N na melhor parametrização
fonte:produção do próprio autor 68
Figura 36 – expoentes entre $Area$ e N na melhor parametrização
fonte:produção do próprio autor 69
Figura 37 – Y final da cidade com relação ao valor de a_0
fonte:produção do próprio autor 69
Figura 38 – Y final da cidade com relação ao valor de ϵ
fonte:produção do próprio autor 70
Figura 39 – Y final da cidade com relação ao valor de $\bar{g_{ag}}$
fonte:produção do próprio autor 71
Figura 40 – Y final da cidade com relação ao valor de σg_{ag}
fonte:produção do próprio autor 72
Figura 41 – Y final da cidade com relação ao valor de $\bar{y_{ag}}$
fonte:produção do próprio autor 72

Figura 42 – Y final da cidade com relação ao valor de σy_{ag}
fonte:produção do próprio autor 73
Figura 43 – β final do sistema urbano com relação ao valor de a_0
fonte:produção do próprio autor 74
Figura 44 – β final do sistema urbano com relação ao valor de ϵ
fonte:produção do próprio autor 75
Figura 45 – β final do sistema urbano com relação ao valor de $\bar{g_{ag}}$
fonte:produção do próprio autor 76
Figura 46 – β final do sistema urbano com relação ao valor de σg_{ag}
fonte:produção do próprio autor 77
Figura 47 – β final do sistema urbano com relação ao valor de $\bar{y_{ag}}$
fonte:produção do próprio autor 78
Figura 48 – β final do sistema urbano com relação ao valor de σy_{ag}
fonte: produção do próprio autor 79

Lista de tabelas

Tabela 1 – Estatística Descritiva das variáveis analisadas 43
Tabela 2 – Escalamento das variáveis com a população das cidades estudadas . . . 46
Tabela 3 – Padrões de escalamento após o corte 52
Tabela 4 – Escalamento das variáveis com a população das cidades estudadas após adoção de critério de corte 54
Tabela 5 – Variáveis estudadas e suas respectivas fontes 55
Tabela 6 – Escalamento das variáveis com a população das cidades estudadas antes da adoção de critério de corte 57
Tabela 7 – Escalamento das variáveis com a população das cidades estudadas após a adoção de critério de corte 58
Tabela 8 – Definição de Parâmetros da cidade 60
Tabela 9 – Definição de Parâmetros dos agentes 60
Tabela 10 – Parametrização capaz de reproduzir a hipótese de escalamento 68

Sumário

1 Introdução 14

2 Leis de Escala 17
2.1 História 17
2.2 Formalização Matemática 19
2.3 Aplicações 25

3 Escalamento Urbano 27
3.1 Evidências Empíricas 27
3.2 Origens das Leis de Escala em Cidades 35

4 Escalamento das Cidades Brasileiras 41
4.1 Metodologia 41
4.2 Resultados 43

5 Origem das Leis de Escala em Cidades com ABM 59
5.1 Metodologia 59
5.2 Resultados 63

6 Conclusões 80
6.1 Conclusões sobre as análises de dados das cidades brasileiras 80
6.2 Conclusões sobre o modelo de origem das leis de escala 82

Referências[1] 84

Apêndice A–Código da análise dos dados em R . 88

Apêndice B–Código do modelo em Python 93

[1] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

As cidades têm crescido e se perpetuado como a principal forma de organização espacial da humanidade. O rápido desenvolvimento urbano é muitas vezes acompanhado de pouco ou nenhum planejamento (UN-HABITAT, 2013). Mesmo quando o processo de planejamento está presente, muito pouco conhecimento científico sobre o funcionamento das cidades é empregado.

O rápido processo de urbanização não acompanhado por um planejamento urbano com fortes bases científicas gera grandes problemas sociais e ambientais no mundo contemporâneo. Ocupando apenas 2% da superfície terrestre, as cidades consomem 80% de todos os recursos naturais e geram 75% de todo o PIB do planeta (GIRARDET, 1996). É esperado, ainda, que as cidades continuem crescendo aceleradamente ao longo do século XXI (UN-HABITAT, 2014). É preciso entender melhor as implicações sociais e ambientais de um planeta predominantemente urbano se desejamos nos preparar para as décadas à frente.

De fato, as cidades são sistemas que integram de maneira complexa seus componentes econômicos, sociais, ambientais e infraestruturais. Talvez pela dificuldade de compreendê-las de maneira complexa e transdisciplinar, muitos de seus problemas sociais e ambientais são tratados de maneira isolada. Isso se reflete em políticas ineficientes e leva a consequências não esperadas e, muitas vezes, desastrosas. Esbarra-se na necessidade urgente de desenvolver uma compreensão unificada das cidades, que integre suas diversas facetas e consiga produzir políticas eficientes de gestão urbana. Nesse sentido, muitos pesquisadores têm se dedicado ao desenvolvimento de uma ciência das cidades (BATTY, 2013).

Nesse caminho, recentes estudos empíricos, encontraram regularidades estatísticas no padrão de escalamento de diversas variáveis urbanas com a população das cidades (BETTENCOURT, 2013). Esses estudos são chamados de leis de escala ou alometria urbana. De uma maneira geral, observa-se que variáveis materiais e de uso de espaço parecem escalar de maneira sublinear, ou seja, crescem mais lentamente do que a população das cidades. Em contrapartida, variáveis que derivam de interações sociais escalam de maneira superlinear, crescendo mais rapidamente do que a população das cidades. Alguns autores sugerem, inclusive, que esses dois padrões básicos de escalamento apresentariam universalidades espaço-temporais (BETTENCOURT et al., 2007; BATTY, 2013). Outros, porém, tem encontrado inconsistências nesses padrões de escalamento, principalmente no

que diz respeito à definição de cidade adotada, determinada a partir de limites minimos de população e densidade (ARCAUTE et al., 2015; COTTINEAU et al., 2015). Ainda se faz necessário, portanto, testar a hipótese de universalidade das leis de escala.

Além das evidências empíricas, esforços têm sido despendidos no sentido de entender as dinâmicas sociais por trás desses escalamentos. Bettencourt publicou, recentemente, um modelo teórico de campo médio que demonstra, a partir de alguns pressupostos básicos, como interações sociais locais resultam nas leis de alometria urbana (BETTENCOURT, 2013). O modelo se baseia em interações locais e num desenvolvimento de infraestrutura que segue uma lógica de preenchimento botton-up. Sendo capaz de explicar grande parte dos padrões de escalamento baseado em pressuposto de auto-organização, o modelo proposto por Bettencourt parece confirmar a seminal, porém ainda incomprovada tese proposta por Jacobs (JACOBS, 1961) de que a essência das cidades seria as interações que ela comporta.

Pela primeira vez em sua história, a ciência caminha em direção à uma teoria geral do funcionamento urbano. Nesse caminho ainda se faz necessário, porém, validar os achados empíricos para uma diversidade maior de cidades do mundo e para uma série mais completa e representativa de indicadores urbanos. Além disso, é de extrema importância explorar mais profundamente o modelo teórico proposto por Bettencourt, a fim de entender sua validade e buscar uma compreensão das dinâmicas sociais que operam por trás das leis de escalas das cidades. Só assim conseguiremos, de fato, transformar esse novo conhecimento sobre o funcionamento das cidades em políticas públicas que operem num sentido de melhorar nossos centros urbanos.

A presente dissertação pretende integrar as recentes pesquisas sobre leis de escala em cidades, produzindo análises empíricas sobre os padrões de escalamento das cidades brasileiras e reproduzindo o modelo teórico proposto por Bettencourt numa estrutura de modelagem por agentes, capaz de incluir agentes heterogêneos e parâmetros flexíveis e, assim, testar algumas hipóteses sobre as forças que governam a alometria urbana.

O primeiro capítulo apresenta uma visão histórica das leis de escala e sua formalização matemática. O segundo capítulo apresenta o estado atual das pesquisas em leis de escala de cidades, suas evidências empíricas e seu principal modelo de explicação. Nesse ponto, espera-se ter explorado satisfatoriamente o que diz respeito à teoria e às observações empíricas do que se produziu sobre leis de escala em cidades. O quarto capítulo apresenta a metodologia e os resultados das análises de padrões de escalamento nas cidades brasileiras. O quinto capítulo apresenta a metodologia e os resultados do modelo proposto

para explicar a origem das leis de escala nas cidades. O sexto e último capítulo traz uma discussão sobre os resultados e indica futuros caminhos a serem seguidos nessa direção de estudo.

2 Leis de Escala

Neste capítulo, apresenta-se uma visão histórica das leis de escala e sua formalização matemática.

2.1 História

A última e maior obra de Galileo Galilei foi intitulada *Two New Sciences* (GALILEI, 1638). Trata-se de um diálogo em quatro capítulos, sendo os dois últimos dedicados à histórica solução do problema de movimento de projeteis. Esse viria a ser o nascimento da física como a conhecemos hoje. Nos dois primeiros e menos conhecidos capítulos, Galileo discute o efeito do escalamento em problemas de construção e de materiais. Essa era a primeira das duas ciências apresentada por ele: a do escalamento (PETERSON, 2002).

Peterson (PETERSON, 2002) afirma que a ideia de escala é básica nas ciências físicas e normalmente é formalizada no início dos livros de introdução à disciplina sob a alcunha de análise dimensional. Ainda que não possuísse as técnicas de que dispomos hoje, Galileu já afirmava, em 1683, que a relação superfície-volume em um sólido pequeno era maior do que essa mesma relação em um sólido grande. Isso é facilmente observado uma vez que a superfície cresce como o quadrado de um do lado L do sólido ao passo que o volume cresce como cubo do lado L. Assim, à medida que se reduz o volume de um sólido, sua superfície é reduzida mais lentamente e as forças que atuam nessa superfície, como a viscosidade ou o atrito, se tornam relativamente mais importantes. Galileo usou esse argumento para explicar porque objetos com tamanhos diferentes levam tempos ligeiramente distintos para cair.

Das duas novas ciências propostas por Galileo, a segunda, do movimento dos corpos, foi profundamente explorada. A das escalas, porém, experienciou um desenvolvimento mais modesto. Talvez isso se explique pela predominância histórica do Teorema Central do Limite (TCL). Proposto em 1810 por Laplace, esse teorema afirma que a distribuição de grandes amostras com variância finita tende a uma forma normal (RICE, 2001). Nessas distribuições, todos os valores flutuam próximos a uma média, nunca indo muito longe dela. Essa média é uma escala típica da variável estudada e observações que fogem muito a essa escala são desprezíveis(figura 1). A distribuição de alturas numa população e de velocidades dos carros numa rodovia seguem formas normais (RICE, 2001).

Se grandes variações são desprezíveis, o papel das leis de escala também o são. Newman (NEWMAN, 2005) afirma que durante muito tempo, a ciência se dedicou a entender

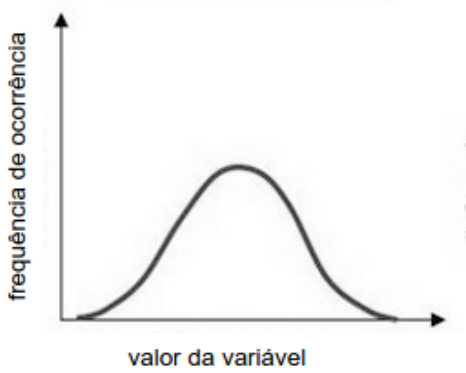


Figura 1 – Distribuição Normal
fonte:produção do próprio autor

os processos de formação das variáveis que apresentavam distribuições normais pois acreditava-se que elas representavam grande parte dos fenômenos conhecidos. Distribuições normais são encontradas em todos os casos em que os resultados podem ser tratados como a soma da contribuição de um grande número de fatores independentes com variância finita (RICE, 2001). Se os fatores envolvidos estão inter-relacionados (como é o caso de muitos sistemas sociais ou naturais) a distribuição pode fugir da normal, principalmente no alcance dos valores extremos (PISARENKO; RODKIN, 2010).

Desde muito cedo, alguns exemplos empíricos de distribuições que fugiam da normal foram encontrados, como as leis de Pareto e de Zipf (NEWMAN, 2005). Acreditava-se, porém, que essas distribuições eram exceções no mundo real ou que eram encontradas pela falta de dados, que as afastava do TCL (PISARENKO; RODKIN, 2010). O tempo, porém, acumulou evidências indicando que as distribuições diferentes das normais eram mais comuns do que se imaginava, principalmente no que tangia eventos naturais (chuvas, terremotos, cheias de rios).

Dessa percepção, nasceu a teoria dos valores extremos (EVT), que trata do comportamento dos eventos extremos de uma distribuição e permite realizar inferências de eventos raros a partir de poucas observações (BURRY, 1975). Uma consequência muito importante desse teorema diz que a cauda de uma distribuição de cauda longa é invariante para grandes amostras, mesmo que a distribuição como um todo varie com a agregação (BURRY, 1975). Significa dizer, basicamente, que as caudas longas existem e, muitas vez, o que falta são observações empíricas suficientemente extensas para comprová-las.

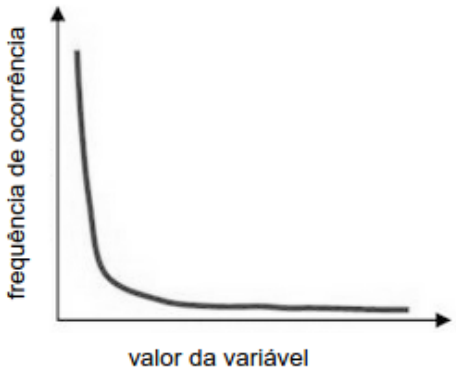


Figura 2 – Distribuição de Lei de Potência
fonte:produção do próprio autor

As leis de potência são um tipo de distribuição de cauda longa(figura 2) que ganharam grande visibilidade por sua recorrência em diversos fenômenos naturais e sociais e pela capacidade de gerar consequências físicas surpreendentes. A maioria dos sistemas complexos apresentam, em alguma dimensão funcional, distribuições de lei de potência (CLAUSET; SHALIZI; NEWMAN, 2009). Acredita-se que elas estejam intrinsecamente ligadas às estruturas hierárquicas e às dinâmicas estocásticas de muitos sistemas (BAK, 1996). De fato, Bak afirma que as leis de potência seriam uma assinatura geral dos sistemas complexos. A pesquisa nas origens das relações de leis de potência e seus esforços em validá-las empiricamente é um tópico muito ativo em diversos campos da ciência, como a física, computação, linguística, geofísica e economia (NICOLIS; NICOLIS, 2007). Recentemente, o estudo das leis de potências bivariadas, as chamadas Leis de Escala, trouxeram grandes avanços em biologia e nos estudos urbanos (MITCHELL, 2011).

2.2 Formalização Matemática

Na estatística, uma lei de potência é uma relação funcional entre duas quantidades onde uma varia como uma potência da outra (BAR-YAM, 2014). Nessas distribuições, grandes acontecimentos possuem baixa, porém não desprezível probabilidade de ocorrência ao passo que pequenos eventos acontecem com grande probabilidade, segundo a lei

$$f(x) = x^{-\alpha}, \text{ onde f(x) é a frequência de ocorrência de um evento e } \alpha \text{ ¿ 1.} \qquad (1)$$

De uma maneira geral, as distribuições normais apresentam grande probabilidade de ocorrência de um valor central, com baixíssima probabilidade de ocorrência de valores extremos. Já na distribuição de lei de potência, uma frequência altíssima de ocorrência de valores pequenos é observada e pouquíssimos mas não desprezíveis valores muito grandes acontecem. Medidas de tendência central fazem pouco sentido nessas distribuições (Figura 3), uma vez que nelas poucas ocorrências se encontram próximas à media. Como se estendem por diversas ordens de magnitude, as leis de potência implicam uma possibilidade de universalidade dos fenômenos observados (BAK, 1996).

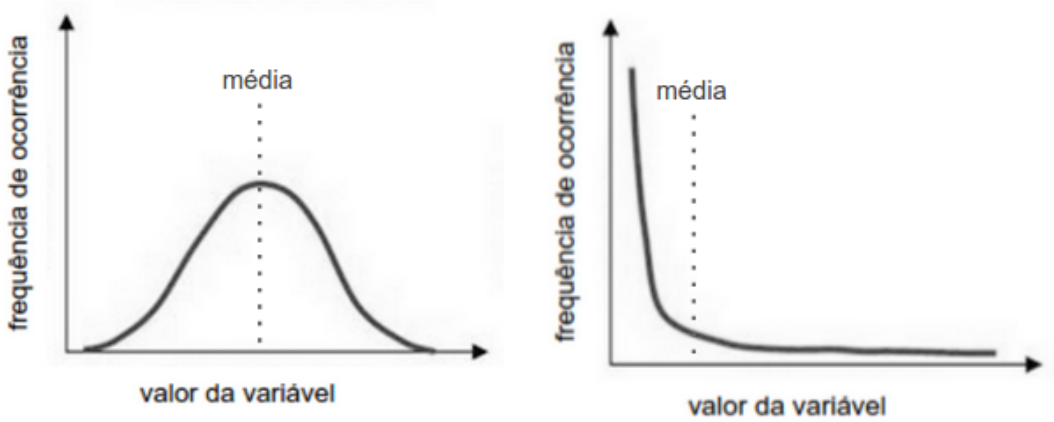


Figura 3 – A média na distribuição normal e na distribuição de lei de potência
fonte:produção do próprio autor

Uma maneira intuitiva de entender a impotência das medidas centrais perante as leis de potência é conceber que a distribuição da renda familiar numa sociedade segue uma de lei de potência (a lei de Pareto). Assim, muitas famílias têm uma renda baixa, enquanto pouquíssimas famílias têm uma renda extremamente alta, muitas ordens de magnitude maior do que a média. Uma descrição da distribuição de renda dessa sociedade através da média, variância e desvio padrão mostra-se ineficaz. Claramente grande parte das famílias estarão muito longe da média, e os valores de variância de desvio padrão encontrados serão enormes. Isso acontece porque as poucas famílias com altíssima renda enviesam a média para cima.

Por essa característica, análises estatísticas baseadas em variância e desvio padrão (como as regressões lineares) são inviabilizadas para esse tipo de distribuição. De fato, leis de potência só possuem médias bem definidas se o expoente α é maior do que 2 e só possuem variância finita para expoentes maiores do que 3. Muitas leis de potência na

natureza apresentam médias definidas, mas variância infinita (NEWMAN, 2005). A maneira de utilizar regressões lineares é adotar a escala logarítmica, linearizando uma distribuição de lei de potência.

As leis de potência são uma generalização para qualquer relação que segue uma função de potência. Quando se estuda relações bivariadas, costuma-se chamá-las de leis de escala (WEST; BROWN, 2000). Um exemplo básico seria a relação entre a área A e o lado L de um cubo. Dobrando-se o lado, multiplica-se a área do cubo por quatro. Pode-se dizer que a A e L possuem uma relação de lei de escala entre si. Mais do que isso, é possível afirmar que a área escala superlinearmente com relação ao lado do cubo, uma vez que ela cresce mais rapidamente.

O grande determinante das leis de escala é a dimensionalidade em que cada uma se constrói. Ainda no exemplo do cubo, a dimensionalidade fica clara quando olhamos para a relação de escalamento entre o lado L , a área A e o volume V. A dimensão do lado é 1 (m), da área é 2 (m^2) e do volume é 3 (m^3). Assim, quando se multiplica o lado L por 2, a área cresce como $2.2 = 2^2$ e o volume cresce como $2.2.2 = 2^3$. Quanto maior for o fator multiplicador do lado, maior será o escalamento da área e, maior ainda, do volume. O exemplo do cubo parece trivial por possuir uma representação geométrica intuitiva, mas alguns sistemas apresentam, conforme veremos mais à frente, relações de escalamento entre suas variáveis funcionais.

A dimensionalidade de uma lei de escala pode adotar a forma de números positivos, negativos e fracionárias. Yaneer Bar-Yam (BAR-YAM, 2014) explica que das dimensionalidades fracionárias em leis de potência nasceu o conceito de fractal (MANDELBROT, 1982). Nesses casos, um crescimento no lado L de um cubo, por exemplo, levaria a um crescimento fractal da área (dimensão entre 1 e 2).

As leis de escala já são objeto de estudo de físicos e matemáticos a bastante tempo, mas só recentemente ganharam importância, quando passaram a ser utilizadas para compreender alguns fenômenos dos sistemas complexos (NICOLIS; NICOLIS, 2007). Trata-se de um conceito matemático muito claro que nos ajuda a entender alguns processos que parecem desconexos, colocando-os dentro de um mesmo paradigma, ou lei.

Alguns escalamentos revelam regularidades subjacentes nas propriedades dos sistemas estudados. Yaneer Bar-Yam (BAR-YAM, 2014) afirma que é comum encontrar, em sistemas complexos, características semelhante nas mais diversas escalas, gerando regularidades surpreendentes. Isso acontece porque o fator de escalamento não depende

do tamanho do objeto estudando, ele se mantém constante em qualquer escala. Essa característica é conhecida como auto-semelhança e é inerente aos sistemas que apresentam leis de potência (NICOLIS; NICOLIS, 2007).

De acordo com Benjamin Widom (WIDOM, 2014), Leis de escala são expressões de princípios físicos através da linguagem matemática das funções homogêneas. Uma função $f(x)$ é dita homogênea de grau n na variável x se, para qualquer λ

$$f(\lambda x) = \lambda^n f(x). \tag{2}$$

Escalando a função por uma constante λ qualquer, simplesmente multiplica-se a função inicial por λ elevado a uma potência n. Isso indica que essa função é invariante com a escala. Ou seja, independente de aumentar ou diminuir a escala dessa função, ela se comporta da mesma maneira com relação às variáveis. Essa característica não é inerente a todos os sistemas físicos. Muitas vezes, quando se muda a escala, muda-se, também, as leis que governam o sistema.

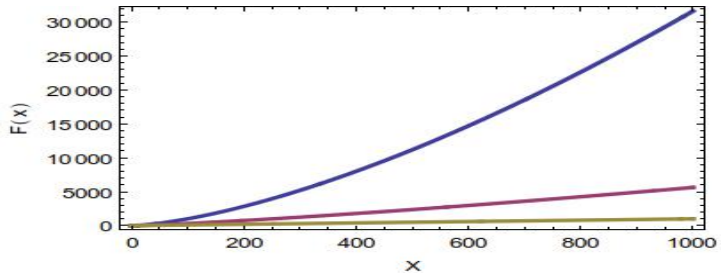
Conforme apresentado anteriormente, uma lei de escala é representada, matematicamente, pela expressão:

$$f(x) = x^{\alpha}. \tag{3}$$

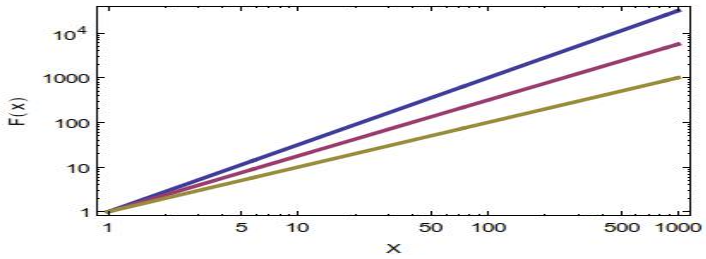
Em uma lei de escala, as variáveis possuem uma relação exponencial, o que significa que $f(x)$ e x divergem para intervalos grandes de x. Na representação gráfica dessa função, duas curvas com pequenas variações no valor do expoente divergem rapidamente, como pode ser observado na Figura 4.

Como a ideia nesses estudos é trabalhar com um intervalo amplo de ocorrências, gráficos de escala normal não são os mais indicados, uma vez que impossibilitam uma visualização clara da relação entre as variáveis. Um gráfico de eixos logaritmos (log x log) lineariza a relação de escalamento. Esses gráficos permitem a observação da ocorrência do fenômeno para um grande intervalo de escalas e neles, a relação entre dois padrões de escalamentos distintos se reduz à comparação da inclinação das retas. A figura 4 representa o escalamento do lado L (linha bege), da área A (linha roxa) e do volume V (linha azul) de um cubo em eixos normais e logaritmos.

A demonstração matemática para essa transformação se segue:



(a) Escala Normal



(b) Escala LogLog

Figura 4 – Escalamento do lado L (linha bege), da área A (linha roxa) e do volume V (linha azul) de um cubo em eixos normais e logarítmos fonte:produção do próprio autor

$$f(x) = x^{\alpha}. \tag{4}$$

Aplicando o logarítmo nos dois lados da equação,

$$logf(x) = logx^{\alpha}, \tag{5}$$

chega-se em:

$$logf(x) = \alpha logx. \tag{6}$$

É fácil observar que, aplicando o logaritmo nos dois lados da igualdade, se chega a uma equação de reta, sendo o expoente α a inclucação da reta.

Em dados empíricos, a observação de linhas retas é uma condição necessária mas não suficiente para afirmar que o fenômeno corresponde a uma lei de escala pois existem diversas maneiras de gerar grupos finitos de dados que seguem esse padrão de distribuição para algumas ordens de magnitude mas nos limites assintóticos não são verdadeiras leis de potência (NEWMAN, 2005). O comportamento da distribuição no limite assintótico diferencia a distribuição de lei de potência das distribuições log-normais.

De fato, a caracterização das leis de potência em dados empíricos é uma tarefa complicada pela grande variação que ocorre na cauda das distribuições (NEWMAN, 2005).

Métodos usualmente aplicados para a caracterização de distribuições de cauda longa, como o método dos mínimos quadrados, podem gerar estimativas imprecisas para os parâmetros e não indicam em qual intervalo os dados seguem uma lei de potência. Clauset (CLAUSET; SHALIZI; NEWMAN, 2009) propôs uma estrutura estatística para a detecção e quantificação de leis de potência univariada em dados empíricos que combina estimativa por máxima verossimilhança com o teste de Komolgorov-Smirnov.

Para leis de escalas bivariadas, Gomez-Lievano (GOMEZLEVIANO; YOUN; BETTENCOURT, 2012), enquanto investigavam a relação entre o tamanho da população e o número de homicídios das cidades brasileiras, propuseram uma estrutura estatística capaz de derivar o expoente da lei de escala entre duas variáveis a partir dos expoentes das leis de potência de cada uma delas. Gomez-Lievano ainda provou, analiticamente, que todas as características de um mesmo sistema que se distribuem em uma lei de potência também apresentarão uma lei de escala bivariada entre si.

Leis de potência em um sistema podem ser originadas de diversas maneiras. Em muitas delas, o escalamento é fruto de processos dinâmicos inerentes ao sistema. Na física, por exemplo, transições de fases em sistemas termodinâmicos estão associadas com a emergência de leis de potência (SOLE, 2012). Diversos sistemas apresentam comportamento escalar à medida que se aproximam da criticalidade e, mais do que isso, alguns sistemas se auto-organizam a atingir a criticalidade (BAK, 1996). Bak afirma que as leis de potência indicariam um atrator em sistemas com criticalidade auto-organizada. De fato, muitas leis de potência que apresentam o mesmo comportamento de escalamento, compartilham das mesmas dinâmicas fundamentais (NEWMAN, 2005).

Alguns processos estocásticos também tem a capacidade de gerar leis de potência. O passeio aleatório com saltos, onde o passeante pode caminhar longas distâncias é um dos exemplos (BARABASI; ALBERT, 1999). De fato, o Teorema Central do Limite, quando destituído da limitação da variância leva à leis de potência (LAM et al., 2011). Estruturas hierárquicas em sistemas também geram distribuições de leis de potência, como é o caso da geometria fractal (SCHROEDER, 2009). Em redes complexas, Albert e Barabasi desenvolveram um método de geração de leis de potência pela dinâmica de ligação preferencial (BARABASI; ALBERT, 1999).

2.3 Aplicações

A distribuição de muitos fenômenos naturais, biológicos e humanos seguem leis de potência. Alguns exemplos incluem a magnitude de terremotos, o tamanho das crateras da lua, a intensidade das explosões solares, a frequência de palavras em um texto, a população de cidades e a renda de sociedades (CLAUSET; SHALIZI; NEWMAN, 2009). Muitas dessas leis de potência encontradas empiricamente se estendem por uma quantidade limitada de ordens de grandeza. É comum encontrar desvios do regime da lei de potência para valores extremos (GOMEZLEVIANO; YOUN; BETTENCOURT, 2012).

A simplicidade e a universalidade das leis de escala parece contrastar fortemente com a diversidade e complexidade das formas de vida. Apesar disso, o escalamento como uma manifestação de dinâmicas e geometrias subjacentes é familiar à física. Ao longo do tempo, essa abordagem trouxe uma percepção mais profunda sobre problemas que atravessam todo o espectro das ciências e das tecnologias pois, tipicamente, indicam princípios físicos e características comuns (WEST; BROWN, 2000).

O interesse científico pelas leis de potência pode ser explicada pela facilidade com que elas são derivadas por algumas classes gerais de mecanismos. Se trata de uma questão de universalidade. Uma vez demonstradas empiricamente, leis de escala apontam para mecanismos específicos de funcionamento das estruturas geradoras e podem indicar relações com outros sistemas aparentemente muito diversos (NICOLIS; NICOLIS, 2007).

Em muitos campos, as Leis de Escala tem sido aplicadas para compreender como distintos aspectos de um sistema variam enquanto o sistema cresce. Diversos sistemas complexos apresentam escalamentos surpreendentes e contraintuitivos, que acabam por nos dizer muito sobre suas dinâmicas estruturais.

Leis de escala em biologia, por exemplo, foram verdadeiros divisores de água nos estudos de morfogênese e evolução (WEST; BROWN, 2000). Desde uma visão superficial da seleção natural, alguém que pensasse sobre a relação da taxa metabólica com a massa de diversos organismos poderia imaginar que encontraria baixa correlação e uma variação enorme, uma vez que cada organismo evoluiu a partir de sua história própria e sua relação com o entorno. O que se observa a partir dos dados, porém, não poderia ser mais diferente do esperado (Figura 5).

Os dados indicam que existe uma sistemática economia de escala, representada pelo expoente β da reta de escalamento menor do que 1. Quanto maior o mamífero, menos

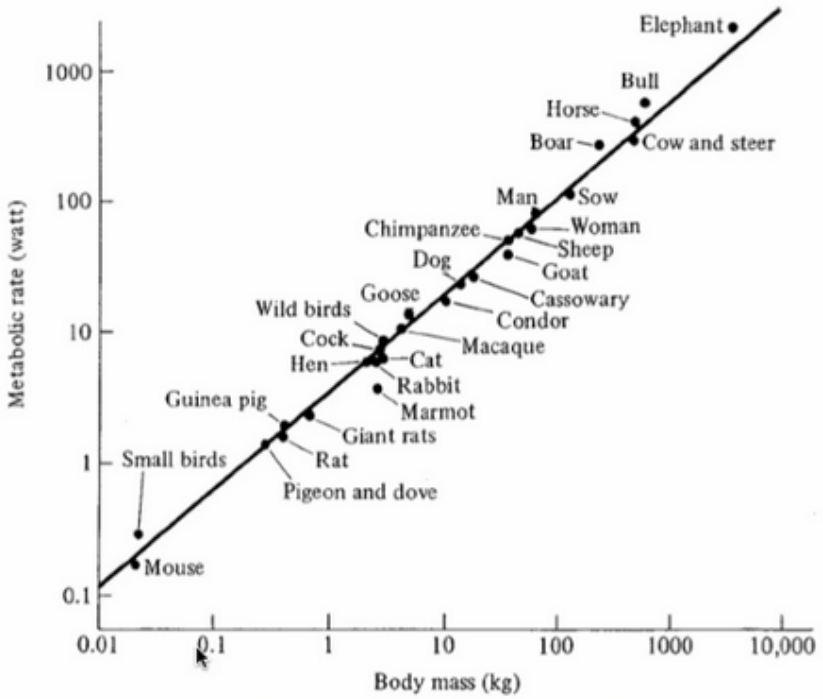


Figura 5 – Escalamento metabólico de animais
fonte: (WEST; BROWN, 2000)

energia por unidade de massa ele precisa para sobreviver. A explicação dada por West e Brown a essa regularidade reside na dimensionalidade fractal das redes de distribuição dos organismos (WEST; BROWN, 2000).

Escalamentos similares são encontrados em diversos processos fisiológicos e nos eventos da vida através de todo seu espectro (WEST; BROWN, 2000). As plantas, por exemplo, ainda que tenham um 'projeto de engenharia' muito diferente dos animais, apresentam leis de escala semelhantes, uma vez que possuem os mesmos princípios de preenchimento de espaço e otimização em suas estruturas de distribuição em rede (WEST; BROWN, 2005). A ideia é que todas essas leis de escala podem ser derivadas das propriedades genéricas do tipo de rede hierárquica subjacentes a esses processos.

Além da biologia, as leis de escala ajudaram a iluminar importantes princípios e estruturas universais em áreas tão diversas como os estudos de transição de fase, a teoria do caos, forças fundamentais da natureza e a descoberta dos quarks. Recentemente, regularidades surpreendentes nas leis de escala de diversas variáveis urbanas com a população das cidades têm sido descobertas. O próximo capítulo se dedica a apresentar esses achados empíricos e o modelo teórico proposto por Luis Bettencourt para explicar sua emergência (BETTENCOURT, 2013).

3 Escalamento Urbano

Neste capítulo, apresenta-se o estado atual das pesquisas em leis de escala de cidades, suas evidências empíricas e um modelo de campo médio com uma possível explicação para sua emergência.

3.1 Evidências Empíricas

Desde a primeira revolução industrial, as cidades são vistas como aglomerados não naturais, arruinados por doenças, poluição, agressões e custos de vida exorbitantes (THOMAS, 1996). As taxas de urbanização, porém, não param de crescer e a humanidade parece caminhar para um futuro predominantemente urbano. De fato, em 2007, pela primeira vez na história da espécia humana, mais da metade de toda a população mundial passou a viver em centros urbanos(Figura 6). Daqui até 2050, é esperado que semanalmente 1 milhão de pessoas troquem o campo pela cidade e cheguemos a 70% de população urbana no mundo (UN-HABITAT, 2014). Se as cidades carregam essa má fama a tanto tempo, porque as pessoas insistem em se mudar para os centros urbanos?

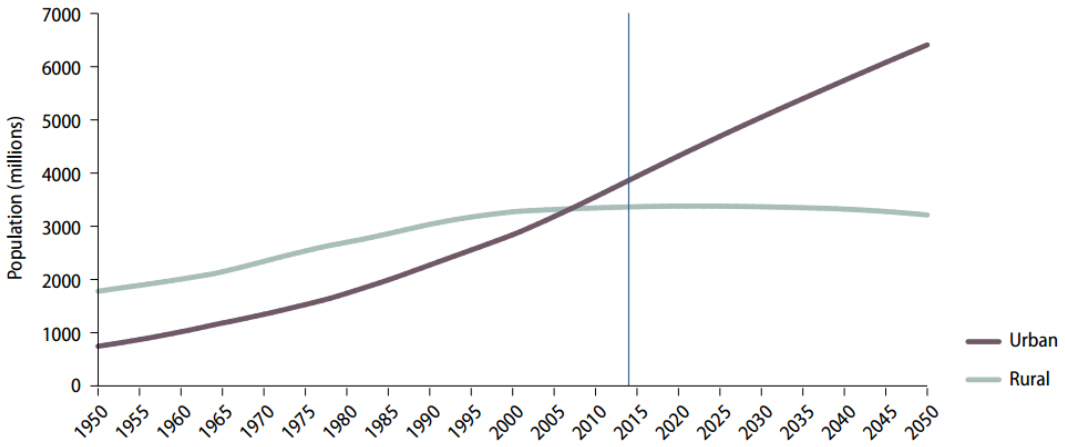


Figura 6 – Urbanização no mundo
fonte: (UN-HABITAT, 2014)

A resposta a essa pergunta pode ser vislumbrada na própria história do nascimento das cidades. Segundo Benevolo, os centros urbanos nasceram primeiramente para juntar indivíduos que desejavam trocar excedentes de produção. Esses indivíduos, até então separados no espaço, passaram a coordenar suas produções, se especializar e trocar os excedentes de maneira muito mais constante e eficiente. Assim, a cidade como organização social da espécie humana se propagou e ainda se propaga (BELEVOLO, 1997). Por isso,

Batty afirma que a função principal e a essência da cidade seria a de gerar interações no espaço e no tempo entre indivíduos diferentes (BATTY, 2013).

Essa noção da cidade como integradora de indivíduos não é nada nova e teve sua primeira proposição no trabalho seminal de Jacobs, *The Life and Death of the Great American City* (JACOBS, 1961). Nessa obra, Jacobs abriu o caminho para estudar as cidades como entidades complexas, organizadas de maneira botton-up a partir do comportamento de seus habitantes.

No capítulo de conclusão do livro, Jacobs define a cidade como um problema de complexidade organizada, baseada no então recém-publicado artigo de Warren Weaver, *Science and Complexity* (WEAVER, 1948). As ideias sobre complexidade eram muito novas naquele momento, e Jacobs rapidamente percebeu a importância de entender as cidades a partir dessa estrutura conceitual. Apesar disso, os meios técnicos ainda não estavam prontos para caracterizar, formalmente, as propriedades urbanas que asseguravam o termo complexo às cidades.

As aglomerações humanas potencializam as interações por juntar, espacialmente, seus habitantes. Marshall definiu os efeitos dessa característica como efeitos de aglomeração mais de um século atrás (MARSHALL, 1890). Até recentemente creditava-se os efeitos de aglomeração à especialização da mão de obra (KRUGMAN; FUJITA; VENABLES, 1999). Entretanto, a hipótese levantada por Bettencourt (BETTENCOURT, 2015) propõe uma nova explicação baseada, fundamentalmente, num efeito de rede: o simples fato de indivíduos estarem espacialmente próximos aumenta o número de encontros potenciais entre eles. Teoricamente, a probabilidade de encontros e interação cresce como a população e produz reduções nos custos de transações e comunicação da sociedade, além de outras vantagens competitivas locacionais (BATTY, 2013).

De fato, muitas vantagens co-locacionais são comumente creditadas às cidades (UN-HABITAT, 2013). Essas vantagens permitiriam, através de economias de escala, uma exploração mais eficiente do espaço urbano (KRUGMAN; FUJITA; VENABLES, 1999). A implementação de sistemas de transporte de alta capacidade (metrô, trens) é um exemplo dessas vantagens.

Apenas recentemente, porém, foi possível confirmar empiricamente essas proposições. As confirmações têm se tornado possível graças à crescente disponibilidade de informações urbanas mundo afora, de estatísticas oficiais às novas maneiras de medir atividades sociais (BATTY, 2013). Se valendo desse grande volume de dados, que cobrem milhares de cidades

do mundo, pesquisadores têm encontrado regularidades estatísticas que indicam como concentrações de pessoas afetam as atividades econômicas, retornam em investimentos em infraestrutura e vitalidade social (BETTENCOURT; WEST, 2010).

Os principais achados desse novo campo de estudos que integra urbanismo, geografia e a física dos sistemas complexos pode ser resumido através da hipótese de que todas as cidades produzem algum tipo de economia de escala espacial à medida que crescem e, ao mesmo tempo, alcançam algum nível de ganho em sua produtividade socioeconômica[1] (Figura 7).

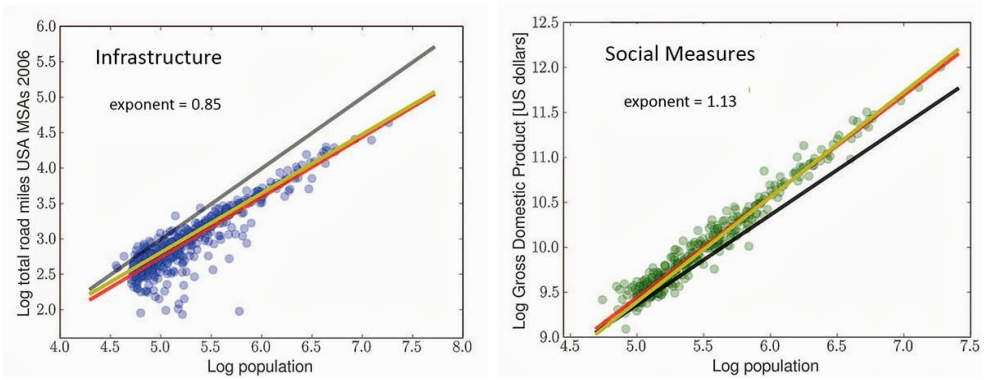


Figura 7 – Características básicas do escalamento urbano
fonte: (BETTENCOURT, 2013)

De uma maneira geral, quando se compara cidades distintas dentro de um mesmo sistema urbano (um país), encontra-se que a maior cidade é mais densa do que a menor cidade e, por isso, o volume total de infraestrutura por habitante é menor na primeira. Ao mesmo tempo, nota-se, via de regra, que a cidade maior é mais rica, mais cara e mais produtiva cultural e tecnologicamente quando se emprega métricas por habitante.

Esses estudos empíricos parecem indicar que centros urbanos produzem mais invenções e criam mais oportunidades de crescimento econômico para seus habitantes (BETTENCOURT, 2013; GLAESER, 2011; BETTENCOURT; WEST, 2010). Além disso, os estudos parecem indicar, também, que pessoas vivendo de maneira mais densa têm menor pegada energética, demandam menos infraestrutura e, por isso, consomem menos recursos naturais per capita (BETTENCOURT; WEST, 2010). Essa hipótese indica que de uma maneira geral, se comparadas com pequenas zonas urbanas, cidades grandes fazem mais com menos.

[1] Define-se produtividade socioeconômica como a taxa de produção dos fatores que derivam de interação social como produção econômica, inovação cultural e tecnológica, crimes e doenças socialmente transmissíveis. Para realizar as análises, é preciso traduzir essas taxas em métricas e indicadores mensuráveis.

Quanto maior uma cidade é, mais produtiva e eficiente ela parece se tornar. Batty (BATTY, 2013) chama essa hipótese de lei de Brand, por conta do trabalho de Stewart Brand (BRAND, 2010), que defendia a necessidade de adensamento dos aglomerados urbanos para que pudéssemos manter os níveis de qualidade de vida que buscamos através de soluções de transporte e construção mais eficientes. É importante esclarecer as limitações dessa hipótese, que só considera impactos gerados dentro da fronteira política da cidade. Se considerados os impactos de todo o ciclo de vida dos produtos consumidos em um centro urbano, é provável que veriámos esse argumento sendo desfeito.

Alguns estudos, entretanto, não encontraram os mesmos padrões de escalamento. Arcaute e colegas (ARCAUTE et al., 2015) publicaram um recente trabalho indicando que, no sistema urbano da Inglaterra, a maior parte dos indicadores urbanos escala de maneira linear com a população das cidades. Os pesquisadores encontraram, ainda, que mesmo quando expoentes de escalamento não linear emergem, eles flutuam consideravelmente com a definição de cidade adotada. Isso indicaria, segundo os autores, uma fragilidade na hipótese das leis de escala. Esses resultados foram reiterados um pouco mais tarde por Cottineau e colegas (COTTINEAU et al., 2015), a partir de uma avaliação de sensibilidade dos expoentes de escalamento no sistema urbano Frances. O estudo de Cottineau apresenta evidência de que, dependendo do que se considera uma cidade -normalmente as definições se baseiam em valores mínimos de população e densidade- é possível encontrar expoentes muito distintos.

Outras incoerências foram encontradas no estudo do escalamento da emissão de CO2 (FRAGKIAS et al., 2013; LOUF; BARTHELEMY, 2014; OLIVEIRA; ANDRADE; MAKSE, 2014), que também parece ser muito sensível à definição de cidade adotada. Louf e Barthelemy defendem a tese de que frente aos embates, pouco pode ser concluído a partir das análises dos dados destituídas de uma compreensão mecanística do fenômeno (LOUF; BERTHELEMY, 2014).

Entretanto, para os autores que defendem a hipótese da lei de Brand os resultados empíricos sugerem que, apesar da aparente complexidade, as cidades parecem escalar de maneira bem simples já que a média de suas propriedades pode ser deduzida com significativa certeza a partir da população da cidade (BETTENCOURT, 2013).

As propriedades socioeconômicas de um centro urbano parecem crescer mais rápido do que uma relação direta (ou linear) que poderia ser prevista a partir da população. Esse efeito é chamado de escalamento superlinear (BETTENCOURT, 2013). Os dados parecem

indicar, também, que o uso de recursos de uma cidade segue uma lei similar, mas invertida. Em média, quanto maior uma cidade, mais eficiente ela é no uso de sua infraestrutura, levando à importantes economias de matéria-prima, energia e emissões. Quando o tamanho de uma cidade duplica, sua infraestrutura material cresce menos do que isso, desde o número de postos de gasolina ao comprimento total dos tubos, ruas e cabos elétricos (BETTENCOURT et al., 2007). Essas quantidades materiais parecem crescer mais lentamente que o tamanho da população. Esse padrão têm sido chamado de escalamento sublinear.

Apesar da grande diversidade global entre as regiões metropolitanas, os mesmos padrões de escalamento parecem ser encontrados na forma como características socio-econômicas e infraestruturais crescem com a população de uma cidade. Esses autores apresentam resultados empíricos que indicam que a produção crescente e custos decrescentes se confirmam em nações com níveis de desenvolvimento, tecnologia e riqueza muito distintos (Figura 8). Ainda que se tenha muito mais informações para cidade em partes mais ricas do planeta, dados de países em desenvolvimento começam a aparecer, e eles parecem seguir os mesmos padrões. O produto interno bruto de cidades no Japão, China e na Alemanha, por exemplo, segue muito perto a mesma curva superlinear Europeia ou da América do Norte. Esses estudos empíricos ainda carecem de validação para países em desenvolvimento.

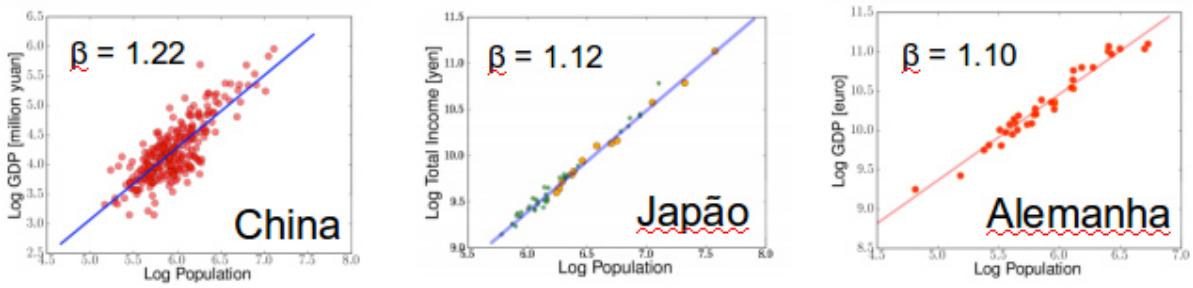


Figura 8 – Indicação de Universalidade Espacial. Expoentes de escalamento superlineares e relativamente próximos são encontrados em distintos países
fonte: (BETTENCOURT, 2013)

Além disso, alguns autores indicam que essas regularidades parecem apresentar consistência temporal (Figura 9). É possível observar o escalamento do PIB das cidades americanas em três momentos distintos da história: 1969, 1989, 2008. Nesses três anos, o expoente

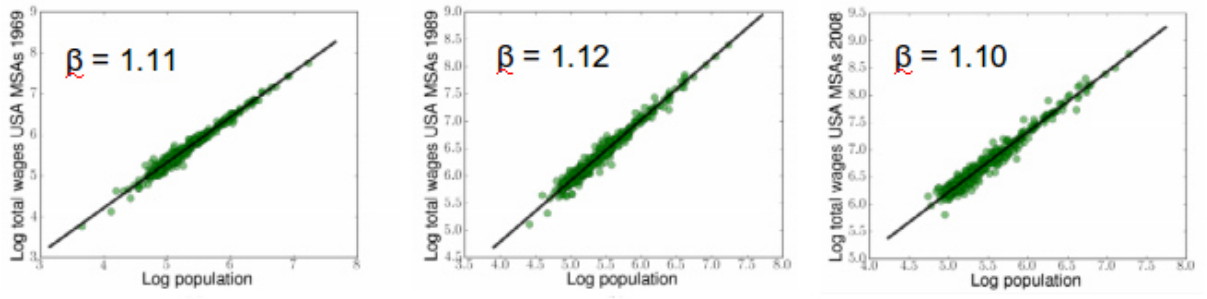
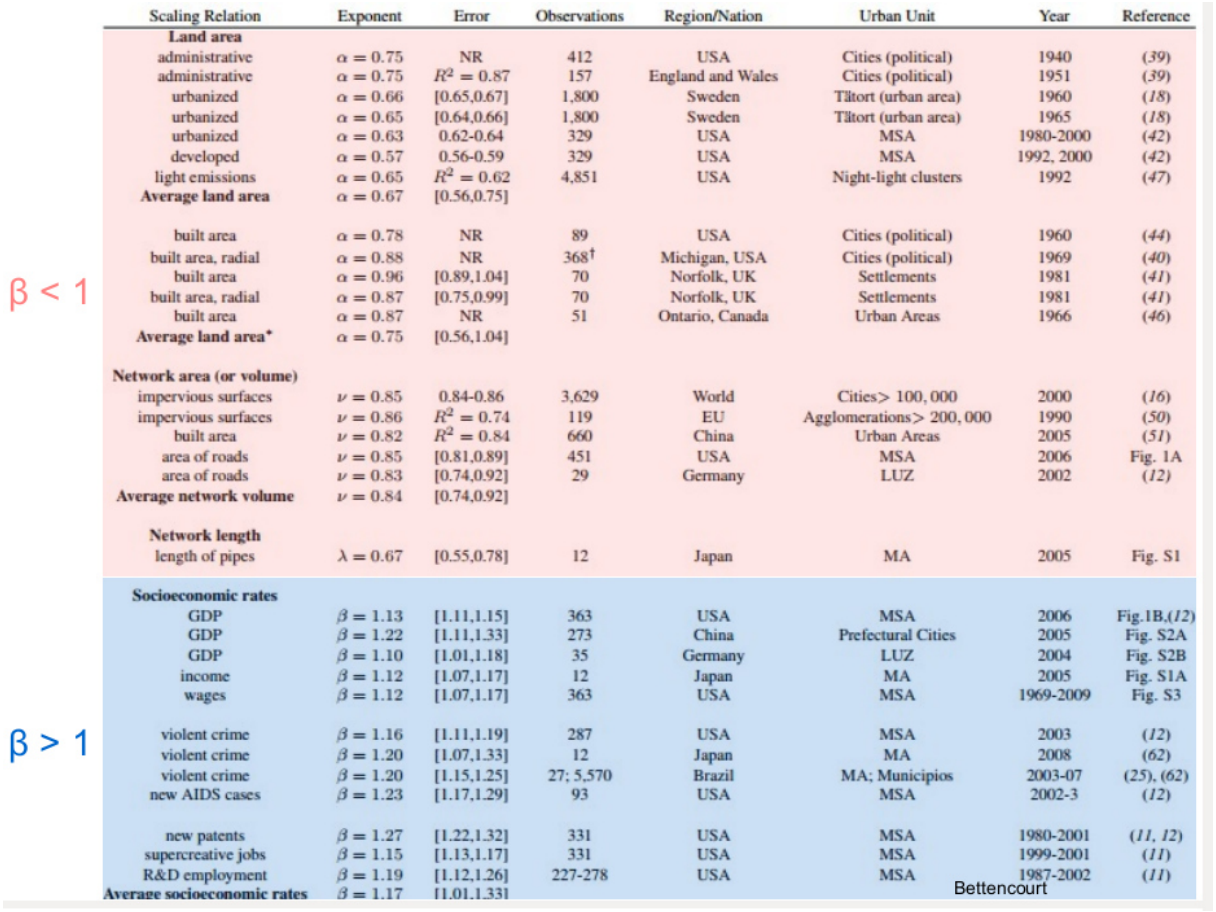


Figura 9 – Indicação de Universalidade Temproal. O expoente de escalamento do PIB com a população das cidades americanas se mantém praticamente constante ao longo de um intrvalo de 40 anos
fonte: (BETTENCOURT, 2013)

se manteve praticamente constante. A única variação significativa no período se relaciona ao Y_0, o PIB base. De fato, um trabalho publicado em parceria com arqueologistas, Bettencourt encontra as mesmas regularidades de escalamento sublinear com a área para cidades pré-colombianas no México (ORTMAN et al., 2014). Tais regularidades se estendendo, sistematicamente, por intervalos espaçotemporais tremendamente heterogêneos remetem à uma universalidade e começam a ser chamadas de leis (BATTY, 2013). Essas regularidades são sintetizadas pela Figura 10.

Para o sistema urbano brasileiro, Alvez (ALVES et al., 2013) e colegas encontraram um expoente de escalamento do PIB com relação à população das cidades que parece concordar com aqueles encontrados por Bettencourt (1.21). Além disso, esse e outros trabalhos exploraram o escalamento de outras métricas menos ortodoxas, como o total de homicídios, de suicídios e taxas de analfabetismo (ANTONIO et al., 2014) (MELO et al., 2014).

Outro importante achado empírico diz respeito ao índice G, proposto por Bettencourt (BETTENCOURT, 2013), que é composto pela multiplicação da produtividade socioeconômica com a infraestrutura material da cidade, ambos per capita. É observado, empiricamente, que o valor de G se mantém, em média, invariante com a população do sistema urbano estudado e sua variação é limitada (figura 11). Como a produtividade socioeconômica escala superlinearmente e a infraestrutura material escala sublinearmente com a população, a invariância do parâmetro G com a população não passa de uma relação



Scaling Relation	Exponent	Error	Observations	Region/Nation	Urban Unit	Year	Reference
Land area							
administrative	$\alpha = 0.75$	NR	412	USA	Cities (political)	1940	(*39*)
administrative	$\alpha = 0.75$	$R^2 = 0.87$	157	England and Wales	Cities (political)	1951	(*39*)
urbanized	$\alpha = 0.66$	[0.65,0.67]	1,800	Sweden	Tätort (urban area)	1960	(*18*)
urbanized	$\alpha = 0.65$	[0.64,0.66]	1,800	Sweden	Tätort (urban area)	1965	(*18*)
urbanized	$\alpha = 0.63$	0.62-0.64	329	USA	MSA	1980-2000	(*42*)
developed	$\alpha = 0.57$	0.56-0.59	329	USA	MSA	1992, 2000	(*42*)
light emissions	$\alpha = 0.65$	$R^2 = 0.62$	4,851	USA	Night-light clusters	1992	(*47*)
Average land area	$\alpha = 0.67$	[0.56,0.75]					
built area	$\alpha = 0.78$	NR	89	USA	Cities (political)	1960	(*44*)
built area, radial	$\alpha = 0.88$	NR	368†	Michigan, USA	Cities (political)	1969	(*40*)
built area	$\alpha = 0.96$	[0.89,1.04]	70	Norfolk, UK	Settlements	1981	(*41*)
built area, radial	$\alpha = 0.87$	[0.75,0.99]	70	Norfolk, UK	Settlements	1981	(*41*)
built area	$\alpha = 0.87$	NR	51	Ontario, Canada	Urban Areas	1966	(*46*)
Average land area*	$\alpha = 0.75$	[0.56,1.04]					
Network area (or volume)							
impervious surfaces	$\nu = 0.85$	0.84-0.86	3,629	World	Cities> 100, 000	2000	(*16*)
impervious surfaces	$\nu = 0.86$	$R^2 = 0.74$	119	EU	Agglomerations> 200, 000	1990	(*50*)
built area	$\nu = 0.82$	$R^2 = 0.84$	660	China	Urban Areas	2005	(*51*)
area of roads	$\nu = 0.85$	[0.81,0.89]	451	USA	MSA	2006	Fig. 1A
area of roads	$\nu = 0.83$	[0.74,0.92]	29	Germany	LUZ	2002	(*12*)
Average network volume	$\nu = 0.84$	[0.74,0.92]					
Network length							
length of pipes	$\lambda = 0.67$	[0.55,0.78]	12	Japan	MA	2005	Fig. S1
Socioeconomic rates							
GDP	$\beta = 1.13$	[1.11,1.15]	363	USA	MSA	2006	Fig.1B,(*12*)
GDP	$\beta = 1.22$	[1.11,1.33]	273	China	Prefectural Cities	2005	Fig. S2A
GDP	$\beta = 1.10$	[1.01,1.18]	35	Germany	LUZ	2004	Fig. S2B
income	$\beta = 1.12$	[1.07,1.17]	12	Japan	MA	2005	Fig. S1A
wages	$\beta = 1.12$	[1.07,1.17]	363	USA	MSA	1969-2009	Fig. S3
violent crime	$\beta = 1.16$	[1.11,1.19]	287	USA	MSA	2003	(*12*)
violent crime	$\beta = 1.20$	[1.07,1.33]	12	Japan	MA	2008	(*62*)
violent crime	$\beta = 1.20$	[1.15,1.25]	27; 5,570	Brazil	MA; Municipios	2003-07	(*25*), (*62*)
new AIDS cases	$\beta = 1.23$	[1.17,1.29]	93	USA	MSA	2002-3	(*12*)
new patents	$\beta = 1.27$	[1.22,1.32]	331	USA	MSA	1980-2001	(*11, 12*)
supercreative jobs	$\beta = 1.15$	[1.13,1.17]	331	USA	MSA	1999-2001	(*11*)
R&D employment	$\beta = 1.19$	[1.12,1.26]	227-278	USA	MSA	1987-2002	(*11*)
Average socioeconomic rates	$\beta = 1.17$	[1.01,1.33]					

Figura 10 – Panorama das Leis de Escalas Encontradas em Cidades
fonte: (BETTENCOURT, 2013)

geométrica básica. Esse parâmetro pode indicar as cidades mais e menos eficientes, de acordo com sua produtividade com relação á infraestrutura utilizada.

De uma maneira geral, então, as propriedades mais notáveis da hipótese de escalamento urbano parecem ser i) a concentração de pessoas no espaço e no tempo e ii) uso mais intenso da infraestrutura material urbana. Juntos, i e ii promovem encontros e coordenação social, aumentando as taxas de produção de variáveis sociais como riqueza, inovação e crime e possibilitam economias em ruas, cabos e outras infraestruturas à medida que uma cidade cresce (BETTENCOURT, 2013).

Essas regularidades estatísticas parecem indicar uma certa universalidade de variações não-lineares, geralmente com relações invariantes com a escala, das quantidades urbanas Y com a população N.

$$Y(N) = Y_0 N^{\beta}, \tag{7}$$

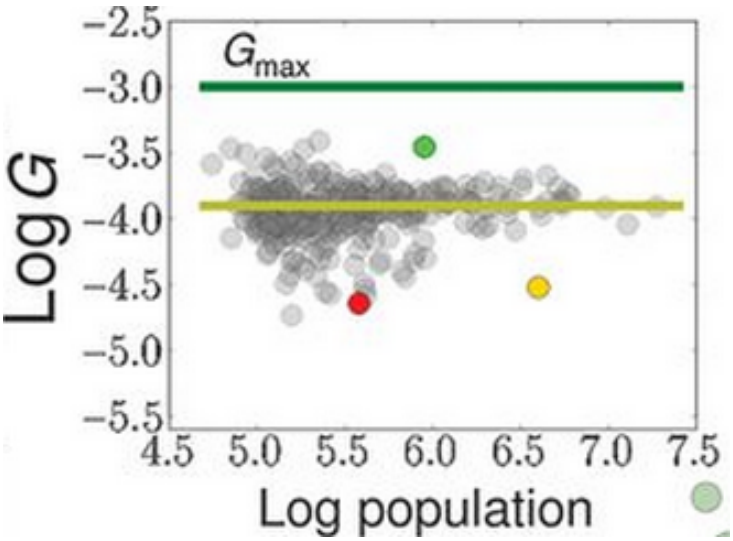


Figura 11 – Índice G para as cidades americanas
fonte: (BETTENCOURT, 2013)

onde Y_0 e β são constantes em N .

Acredita-se que o as regularidades emergem porque os mesmos processos sociais e econômicos atuam em todas as cidades. Da mesma maneira que na biologia, essas regularidades estariam relacionadas à característica de otimização e preenchimento do espaço da rede de distribuição de bens, pessoas e informação da cidade.

A hipótese que tem se criado para explicar a recorrência desses padrões reside nas relações sociais (SCHLAPFER et al., 2014). Um estudo do tamanho da rede média de ligações de celulares em Portugal e no Reino Unido indicaram que o total de números distintos que se liga em uma cidade escala de maneira semelhante a todas as variáveis socioeconômicas: superlinearmente com a população(Figura 12). O valor do expoente de escalamento β se aproxima muito do observado para as outras variáveis - 1.12.

O escalamento superlinear do tamanho da rede de interações na telefonia móvel seria uma proxy para as interações totais de uma cidade. Essa hipótese corrobora algo já sugerido por diversas gerações de urbanistas: a cidade tem sua essência nas relações sociais.

A hipótese das leis de escala carece de comprovação por falta de dados sobre interações totais na cidade. Além disso ainda se faz necessário validar os achados empíricos em mais sistemas urbanos e contra uma quantidade maior de indicadores. Os achados atuais se concentram prioritariamente em sistemas urbanos de países desenvolvidos, sendo

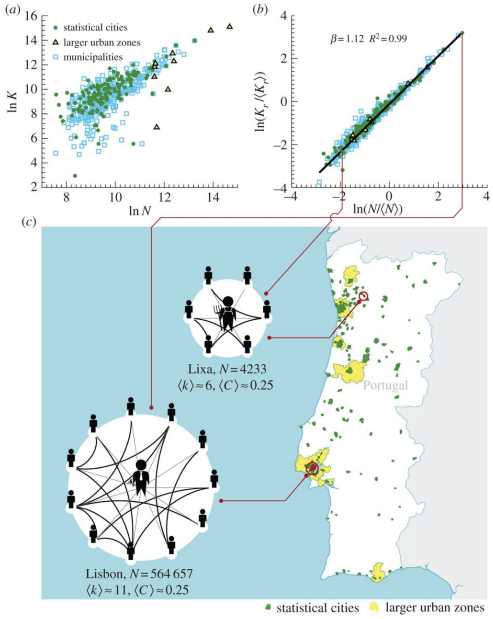


Figura 12 – Caracteristicas Basicas de Escalamento Urbano
fonte: (SCHLAPFER et al., 2014)

importante validar os padrões de escalamento para cidades de países em desenvolvimento, como o Brasil.

3.2 Origens das Leis de Escala em Cidades

O mapeamento empírico das leis de escala é de extrema importância na fundação da Ciência das Cidades, mas é necessário ir além do empirismo e entender de onde emergem esses escalamentos. Bethelemy afirma que é perigoso interpretar resultados empíricos sem uma visão mecanicista do problema (LOUF; BERTHELEMY, 2014). Segundo ele, até que se tenha uma compreensão satisfatória dos mecanismos responsáveis pelos comportamentos observados, conclusões baseadas apenas em dados podem levar à consequências desastrosas.

Nesse sentido, Luis Bettencourt, físico teórico português e professor do Santa Fé Institute publicou, na Science de junho de 2013, um modelo capaz de explicar a formação dessas leis de escala, baseado na ideia de cidades como redes de infraestrutura e sociais co-localizadas no espaço e no tempo (BETTENCOURT, 2013).

A ideia central do modelo é que as cidades são, principalmente, grandes redes sociais. Nesse sentido, Bettencourt afirma que cidades não são apenas grandes aglomerados de pessoas mas são, na verdade, aglomerados de relações entre pessoas. O espaço, o tempo e a infraestrutura desempenhariam um papel fundamental em permitir a formação

e permanência das interações sociais. A seguir, reproduzimos o modelo de Bettencourt (BETTENCOURT, 2013). Isso nos permitirá desenvolver nosso modelo baseado em agentes no próximo capítulo.

Assim, o modelo parte de 4 pressupostos básicos:

I) Cidades são redes de mistura social onde cada morador, com um custo aceitável, pode se conectar com qualquer outro morador. Isso explicaria a necessidade de densidades crescentes com o tamanho da cidade. Por serem maiores, os custos de transporte também são maiores, obrigando a população a se adensar no espaço para garantir o cumprimento desse pressuposto.

II) A rede de infraestrutura é descentralizada. O que significa dizer que ela é construída de maneira gradual, à medida que a cidade cresce.

III) O esforço individual para se manter socialmente conectado não depende do tamanho da cidade. Esse pressuposto é possível graças à I e II já que, por se tornarem mais densas à medida que crescem, mesmo se a distância percorrida por um indivíduo se mantiver constante, mais pessoas e instituições podem ser alcançadas por um morador de uma cidade maior se comparado com um morador de uma cidade menor.

IV) A produtividade socioeconômica das cidades é proporcional às interações sociais realizadas por unidade de tempo. Dessa maneira, a produtividade seria determinada pela densidade de pessoas nas redes públicas. Por produtividade socieconômica, Bettencourt entende toda e qualquer característica que derive de interações sociais, como a produção econômica, tecnológica, crimes e doenças sexualmente transmissíveis.

Matematicamente, a estrutura teórica proposta por Bettencourt pode ser transcrita da seguinte maneira: Adotando uma cidade hipotética, com população N, área $A = L^2$ e densidade N/A. Bettencourt descreve a interação do tipo K entre a pessoa i e a pessoa j em termos de uma rede social F_K^{ij} e assume que as interações sociais são locais, acontecendo numa área de interação do indivíduo a_0 atravéz de um comprimento médio l_0 e possuem uma intensidade g, que pode ser positivo ou negativo.

Assim, é possível descrever a média de interações sociais locais por habitante como o produto da área abrangida por seu movimento $v_0 = a_0.l_0$ com a densidade populacional N/A, onde l é o comprimento típico percorrido pelas pessoas, bens e informações em uma cidade.

Dessa maneira, a produção socioeconômica de um indivíduo pode ser computada como a multiplicação do número total de interações desse indivíduo pela produção média por interação g, obtendo-se

$$y = g.a_0.l_0.\frac{N}{A} = G\frac{N}{A}, \tag{8}$$

onde o parâmetro $G = ga_0l_0$, assim como a media empírica, mede o produto da produtividade socioeconômica pela área, ambos *per capita*. Um pressupostos adotados por Bettencourt em seu modelo é o de que o esforço humano é limitado (pressuposto III), ou seja, a produção média por encontro não varia em cidades maiores. G não depende da população.

Nessa equação, outro pressuposto adotado por Bettencourt é o do que a produção socioeconômica é proporcional ao número de encontros de uma cidade (pressuposto IV). Cidades mais densas seriam mais propícias a gerar encontros.

Para encontrar a produção socioeconômica da cidade inteira Y, basta multiplicar a produção per capita (8) pela população:

$$Y = G\frac{N^2}{A}, \tag{9}$$

onde

$$dG/dN = 0. \tag{10}$$

Um terceiro pressuposto importante do modelo, é o da cidade como uma rede de mistura social (pressuposto I). Ele visa garantir que ainda que os habitantes de uma cidade ocupem espaços diferentes em momentos diferentes, todos possam, em princípio encontrar qualquer outro indivíduo (Pressuposto I). Esse pressuposto é importado dos modelos de biologia populacional e converge com a definição de Área Estatística Metropolitana (MSA) adotada pelo Escritório de Censo Norte-Americano (MANAGEMENT; BUDGET, 2010). Esse pressuposto, assegura uma unidade territorial à cidade. Matematicamente, ele é garantido ao computar o custo de transporte por indivíduo como sendo proporcional à dimensão transversal L de uma cidade. Assim, o custo total de transporte para manter um indivíduo socialmente misturada é

$$\tau = \epsilon L = \epsilon A^{1/2}, \tag{11}$$

onde ϵ é o custo de transporte por unidade de comprimento. Da mesma maneira que se faz com a produção socioeconômica, é possível extrapolar o custo de transporte da cidade inteira multiplicando o custo por indivíduo (10) pela população da cidade,

$$T = \epsilon A^{1/2} N. \tag{12}$$

Garante-se o pressuposto de mistura social igualando y e t. Assim, obtemos:

$$y = t, \tag{13}$$

substituíndo por (8) e (10),

$$G\frac{N}{A} = \epsilon A^{1/2}, \tag{14}$$

isolando A, chegamos à

$$A = a_0 N^{\alpha}, \tag{15}$$

onde $a_0 = (\frac{G}{\epsilon})^{2/3}$ e $\alpha = 2/3$

Como os resultados empíricos nos mostram que é possível derivar a produção socioeconômica Y de uma cidade a partir de sua população, substituímos em (8) para encontrar.

$$Y = G\frac{N}{A}N = G\frac{N^2}{aN^{2/3}} = \frac{G}{a}\frac{N^2}{N^{2/3}}, \tag{16}$$

$$Y = Y_0 N^{\beta}, \tag{17}$$

onde $Y_0 = \frac{G}{a}$ e $\beta = 4/3$

Bettencourt afirma que essa abordagem nos dá apenas um limite máximo de β, uma vez que se considera a área da cidade inteira como área de interação potencial. Na verdade, à medida que uma cidade cresce, seu espaço é ocupado e os fluxos de pessoas, bens e informações são canalizados em suas redes de infraestrutura (pressuposto II). As interações sociais utilizam essa rede como meio, sendo fácil prever que adotar a área da cidade inteira como uma proxy para a área de interação superestimaria β. A área da rede de infraestrutura de uma cidade, nesse sentido, seria uma aproximação muito mais razoável. Por isso, Bettencourt propõe adotar uma nova área de interação, baseada na área da rede de infraestrutura A_n, calculada a partir da distância média por habitante $d = \frac{A}{N}^{1/2}$:

$$A_n(N) \approx Nd = N\frac{A}{N}^{1/2} = \frac{N}{N^{1/2}}A^{1/2} = N^{1/2}A^{1/2}. \tag{18}$$

É possível colocar A_n em função de N apenas, substituindo a equação acima em (15), obtendo

$$A_n = (aN^{2/3})^{1/2} N^{1/2} = a^{1/2} N^{5/6}. \tag{19}$$

Se subistutuírmos essa nova área em (8), obtemos

$$Y = G\frac{N}{An}N = G\frac{N^2}{a^{1/2}N^{5/6}}, \tag{20}$$

$$Y = Y_0 N^{7/6}. \tag{21}$$

Assim, chega-se á um resultado satisfatório, como apresentado na Figura 13. É possível observar a melhor regressão dos dados (vermelho) se comportando de maneira muito semelhante às previsões do modelo (amarelo), tanto para o escalamento superlinear da variável socioeconômica e quanto para o escalamento sublinear da variável infraestrutural.

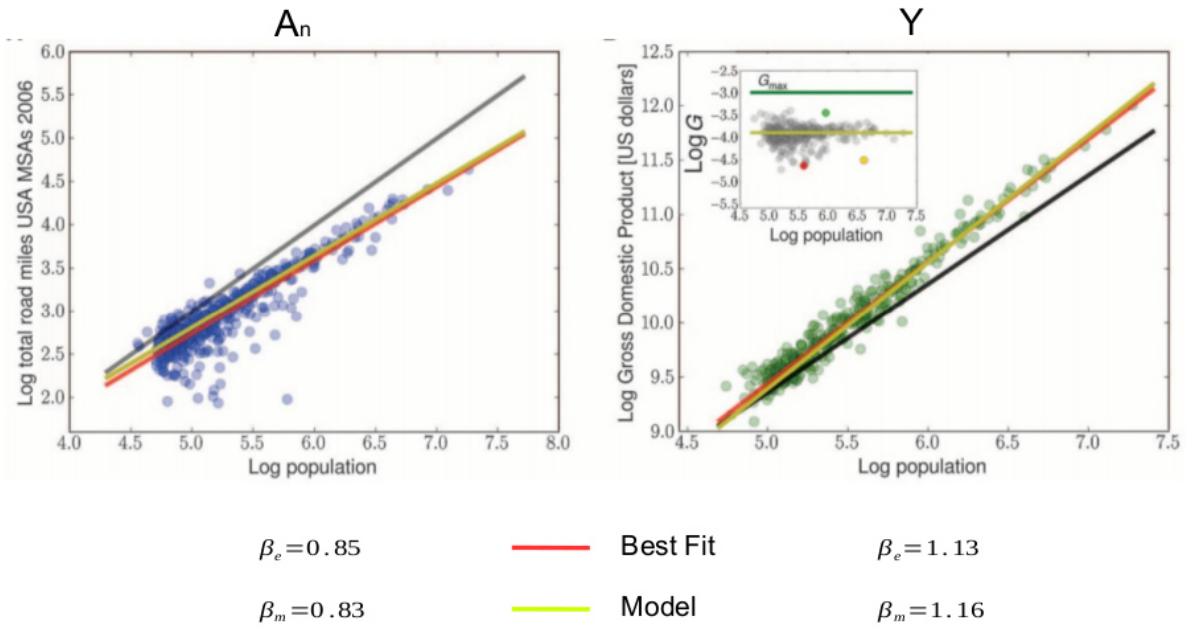


Figura 13 – Síntese das Leis de Escalas Encontradas em Cidades
fonte: (BETTENCOURT, 2013)

A estrutura conceitual proposta por Bettencourt consegue traduzir a origem das leis de escala em poucas linhas. Os dados empíricos, porém, apresentam certa variação. As regressões médias são feitas a partir de uma nuvem de dados e é de extremo interesse dos planejadores urbanos explorar as variações internas dessa nuvem. A universalidade das leis de escala não nos permite entender o que pode fazer uma cidade escalar sua produção socioeconômica no limite inferior ou superior do intervalo encontrado empiricamente. Nem

tampouco o modelo de Bettencourt, por se tratar de uma aproximação de campo médio. Para isso, é preciso representar esse fenômeno dentro de um ambiente de modelagem por agentes. Esse método de modelagem comporta heterogeneidades dos agentes, da rede social e da infraestrutura, permitindo uma análise exploratória dos parâmetros.

4 Escalamento das Cidades Brasileiras

Neste capítulo, apresenta-se os resultados encontrados a partir da exploração dos dados de cidades brasileiras.

4.1 Metodologia

Com o intuito de validar empiricamente as leis de escala apresentadas na literatura (BETTENCOURT, 2013), optou-se por analisar dados de cidades brasileiras. Assim, obteve-se os dados do censo 2010 do IBGE (ftp.ibge.gov.br) referentes à população, área e renda de cada uma das 5561 cidades do país. Dados referentes à infraestrutura de ruas foram obtidos no site de mapeamento colaborativo Open Streen Maps - OSM (openstreetmaps.com). Os demais indicadores, foram levantados junto ao portal de dados abertos do IPEA, mas são referentes a outros órgãos do governo (DataSUS e Ministério da Fazenda). Fez-se necessário preprocessar os dados do OSM, usando ferramentas de GIS (QGIS), para computar o total de quilômetros de ruas por cidade.

Conforme a hipótese proposta por Bettencourt em seu modelo, as leis de escala das cidades derivariam de relações sociais entre indivíduos. É preciso adotar uma definição de cidade que seja coerente com esse pressuposto e que traduza as áreas onde as interações acontecem. Uma região metropolitana (RM), a princípio, desempenha, funcionalmente, o papel de uma única cidade já que uma parcela significativa da população que habita a região metropolitana interage intensamente. No Brasil, entretanto, ainda que o IBGE tenha desenvolvido no fim da década de 60 uma definição nacional para Região Metropolitana, a partir da constituição de 1988, os estados passaram a ser responsáveis pela delimitação das RMs (IBGE, 2015). Disso, incorre um certo grau de incerteza quanto à consistência das definições de Região Metropolitana pelo país.

Dessa forma, optou-se por testar os padrões de escalamento utilizando dois bancos de dados diferentes. O primeiro, constituído pelas cidades brasileiras de acordo com suas divisões políticas. No segundo, substituiu-se os municípios que constituem uma região metropolitana por uma única entidade espacial, a fim de representar o que seriam as cidades funcionais. Para isso, somou-se as variáveis das cidades que formam uma região metropolitana segundo a definição oficial do IBGE (IBGE, 2015) e essas cidades foram substituídas, na base de dados, por sua região metropolitana representativa. Ao total, 632 cidades foram substituídas por 42 regiões metropolitanas. Só a Região Metropolitana de São Paulo, por exemplo, representa 39 municípios.

Ainda foram excluídas da base os municípios que apresentavam o total de quilômetros de ruas igual a 0 de acordo com o Open Street Maps. Esse procedimento excluiu mais 41 municípios. Dessa maneira, a base de dados final das regiões metropolitanas que foi utilizada nas análises continha 4930 cidades funcionais.

Todas as análises que se seguem foram realizadas paralelamente para o banco de dados das cidades e o banco de dados das regiões metropolitanas. Escreveu-se um script na linguagem de programação estatística R que fizesse a análise dos dados de maneira parametrizada a fim de automatizar o processo (apêndice A).

A análise de dados consistiu, primeiramente, numa caracterização do banco, com uma análise descritiva das variáveis estudadas. Nesse momento, histogramas foram produzidos para entender a distribuição dos dados. Um boxplot normalizado com z-valores também foi plotado para cada variável.

Em seguida, uma regressão simples pelo método dos mínimos quadrados foi realizada para o log de cada variável em relação ao log da população total da cidade. Os valores dos parâmetros, do coeficiente de determinação (r^2) e o nível descritivo ($valor-p$) encontrados na regressão foram adicionados à uma tabela para cada uma das duplas de variáveis analisadas. Num primeiro momento, a nuvem de pontos das cidades brasileiras pareceu conter uma grande variabilidade. É possível que isso se dê por conta da definição de município do IBGE, que pode divergir profundamente da noção de cidade como unidade funcional. Os municípios são unidades políticas, muitas vezes contendo mais de uma cidade funcional em seu território. O ideal para esse tipo de estudo seria contar com um banco de dados de cidades naturais (JIANG; MIAO, 2014) brasileiras.

Alguns autores (COTTINEAU et al., 2015; ARCAUTE et al., 2015; LOUF; BERTHELEMY, 2014) tem posto em questão a sensibilidade dos padrões de escalamento com relação à definição de cidade adotada. Por conta disso, uma análise de robustez dos valores do expoente de escalamento entre cada variável e a população das cidades foi realizada. A análise de robustez testou os expoentes de escalamento contra definições de cidades a partir de um valor limite de população e de densidade. Cada valor limite de população e densidade era usado como valor de corte no banco de dados. Com o banco novo, uma nova análise de escalamento era realizada.

4.2 Resultados

Uma análise dos histogramas das variáveis parece indicar uma distribuição contendo muitos valores extremos, fugindo da distribuição normal. As distribuições da área e dos quilômetros de ruas estão ainda mais afastadas da normal.

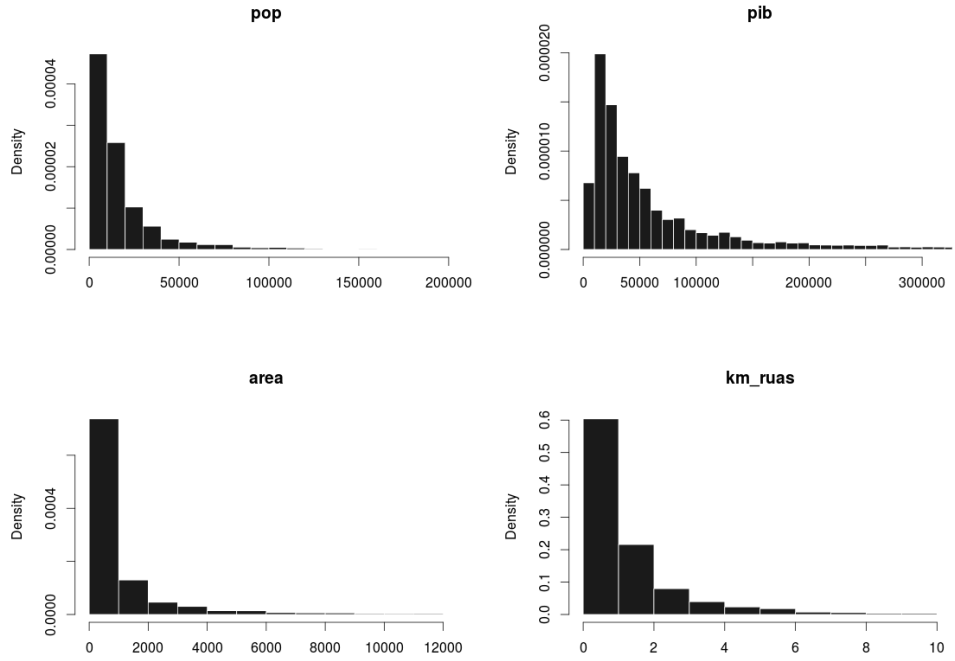


Figura 14 – Histogramas das variáveis estudadas. fonte:produção do próprio autor

A tabela 1 apresenta uma análise descritiva das variáveis estudadas.

	n	média	desvio padrão	mediana	mínimo	máximo
população	4930	38955.89	372238.91	10470.50	805.00	19683975.00
área	4930	1706.45	6215.04	453.40	3.60	159533.00
PIB	4930	344950.75	5213812.22	39152.90	3229.42	313164189.85
KM de ruas [x 100]	4930	2.02	8.90	0.78	0.00	333.94

Tabela 1 – Estatística Descritiva das variáveis analisadas

Porém, ao se plotar as variáveis estudadas contra a população o fit do modelo de regressão linear não parece acompanhar os dados. Para o PIB (Figura 15), os valores extremos, das maiores cidades, ficam acima da função de regressão.

É provável que isso aconteça por conta da definição do IBGE de município, que pode não traduzir bem as unidades espaciais funcionais que seriam as cidades naturais. É provável que muitas cidades carreguem valores de PIB que não derivem exclusivamente de dinâmicas sociais locais. Municípios com fortes características industriais ou com aportes externos de investimentos, como seria o caso dos municípios do litoral norte do estado do Rio de Janeiro que realizam exploração de petróleo off-shore podem influenciar a regressão.

A regressão entre a população e o PIB das cidades estudadas apresenta uma inclinação de 1.01, r^2 ajustado de 0.73 e um $valor_p$ muito próximo de 0. Ainda que pequena, já é possível observar uma relação de escalamento superlinear, de acordo com a previsão teórica. A linha preta representa uma relação de escalamento linear.

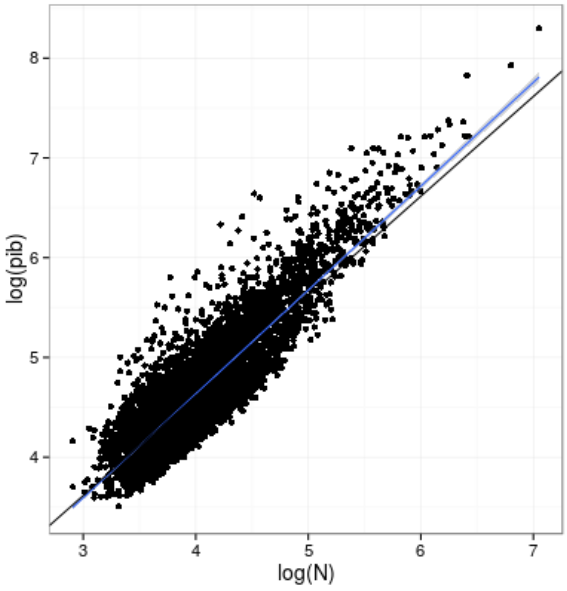


Figura 15 – Escalamento entre a População e o PIB das cidades estudadas.

Já a infraestrutura viária das cidades, contabilizado aqui como o total de quilômetros de ruas, parece seguir uma tendência de escalamento sublinear com a população das cidades. À mesma maneira do PIB, porém, o modelo de regressão linear não conseguiu acompanhar muito bem as maiores cidades do sistema urbano nacional.

O melhor ajuste dos dados fica abaixo das observações empíricas, influenciado pela ocorrência de muitos outliers entre as cidades pequenas. O problema nesse caso, porém, pode ser explicado por outro motivo. Por se tratar de um banco de dados colaborativos, onde usuários são responsáveis por cadastrar as ruas, é de se esperar que aquelas cidades

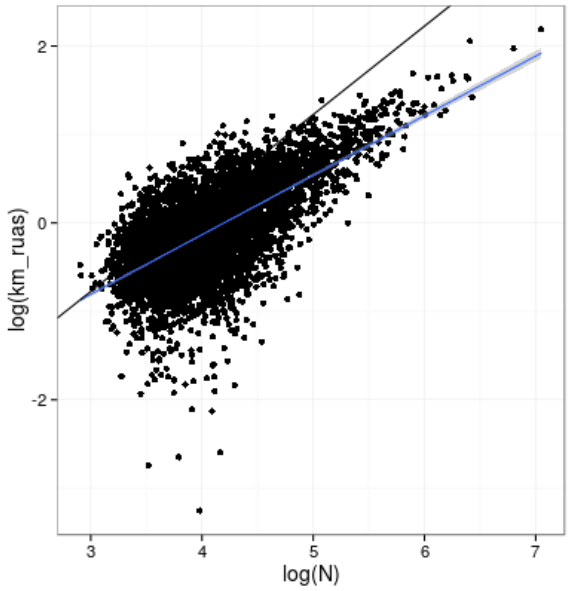


Figura 16 – Escalamento entre a População e infraestrutura viária das cidades estudadas.

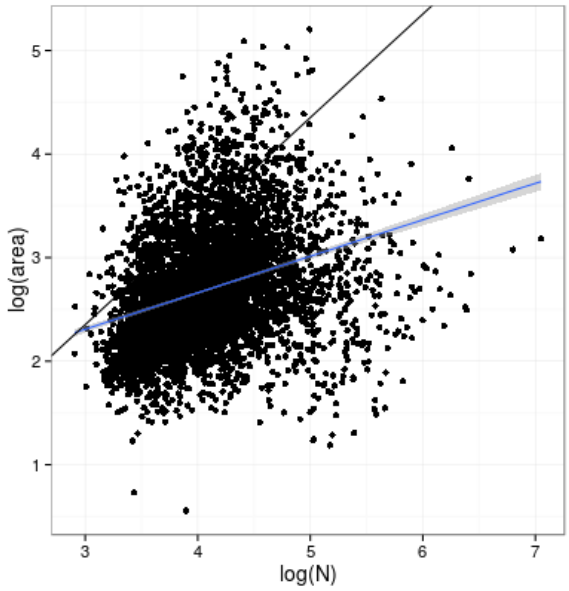


Figura 17 – Escalamento entre a População e á Área das cidades estudadas.

com menos usuários de internet tenham uma menor parcela de suas ruas cadastradas. O erro cresce em cidades de poucos colaboradores do Open Street Maps.

O modelo de regressão linear entre a população e o total de quilômetros de ruas nas cidades estudadas apresenta um valor de inclinação de 0.69 e um r^2=0.42 (Figura 16).

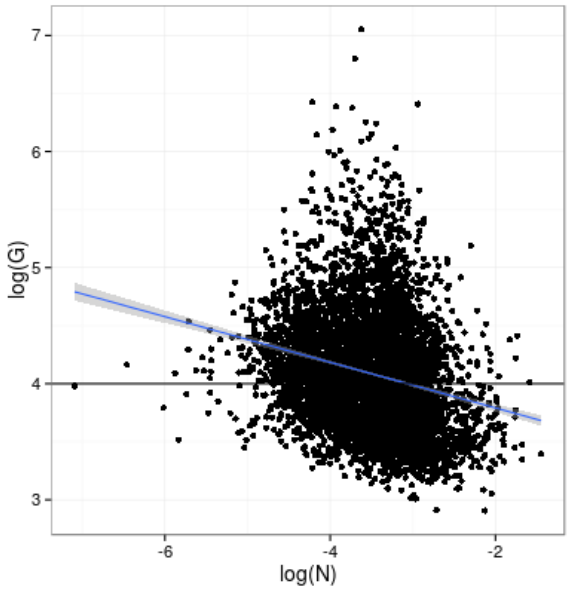


Figura 18 – Escalamento entre a População e o G das cidades estudadas
fonte:produção do próprio autor

Para a área, os dados apresentam grande variação, conforme apresentado na tabela 1. Essa variação claramente prejudica o modelo de regressão com a população. Nessa variável, já não se pode distinguir uma tendência tão clara como nas outras. Ainda que o escalamento da população com a área apresente um valor de inclinação sublinear, essa inclinação está abaixo daquilo que seria esperado teoricamente. Além disso, o $r^2 = 0.22$ indica uma correlação menor do que a das outras variáveis (Figura 17).

O escalamento de G com a população das cidades também não parece seguir completamente aquilo que seria esperado teoricamente. Existe um grupo de cidades de porte médio que afasta a regressão da inclinação nula. As variações no índice G podem indicar problemas nos dados de infraestrutu ra do OSM e anomalias no PIB dos municípios. O modelo de regressão linear encontra uma inclinação de -0.29 e um $r^2 = 0.15$(Figura 18). O escalamento das variáveis do banco de dados de Regiões Metropolitanas se comportou de maneira muito semelhante. A tabela 2 apresenta os resultados iniciais encontrados.

	intercepto	α [C.I.]	r^2	$valor_p$
população x PIB	5.27e-01	1.01 [0.99, 1.02]	0.73	2.2e-16
população x km de ruas	5.81e-01	0.69 [0.67, 0.71]	0.43	2.2e-16
população x área	-2.90e+00	0.54 [0.52, 0.57]	0.22	2.2e-16
população x G	-2.32	-0.29 [-0.33, -0.26]	0.15	2.2e-16

Tabela 2 – Escalamento das variáveis com a população das cidades estudadas

Claramente a área é a variável do IBGE que pior representa as dinâmicas populacionais de uma cidade. Diferentemente do PIB e do total de quilômetros de ruas, sua correlação com a população das cidades não parece ser significativa.

A maneira estatística de tratar os dados é adotar um critério de corte a partir da função de distribuição cumulativa. Seguindo essa metodologia, o corte deve ser feito de acordo com o inicio do regime de distribuição de lei de potência (CLAUSET; SHALIZI; NEWMAN, 2009). Alvez e colegas inplementaram (ALVES et al., 2014) essa metodologia para algumas variáveis sobre o banco de dados de cidades brasileiras. No presente trabalho, optou-se por utilizar outro método de corte no banco de dados, que é apresentado a seguir.

É muito provável que isso se de por conta da definição de município adotada pelo IBGE, que nenhuma relação tem com o centro urbano propriamente dito. A Figura 19 apresenta o mapa do município de Oriximiná-PA e de seus dois centros urbanos. É possível observar uma grande discrepância entre o tamanho do município e sua população urbana. De fato, Oriximiná possui uma área maior do que a de Portugal e uma população pouco maior do que 50 mil habitantes.

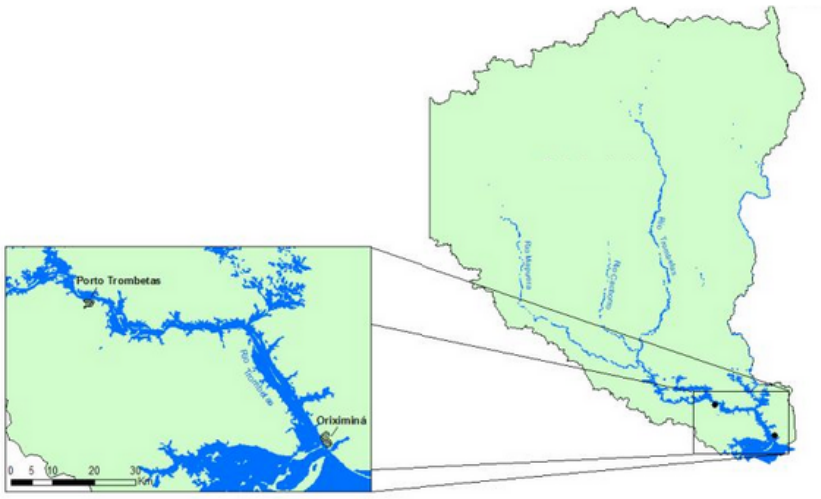


Figura 19 – Oriximiná-PA: relação entre área do município e seu centro urbano
fonte:adaptado de orixi.wordpress.com

Para os fins do estudo aqui apresentado a definição do IBGE de município parece não condizer com aquilo que se busca investigar: as cidades funcionais. De fato, parece que se esbarra em um problema de definição da unidade espacial estudada. Para fazer com que o banco de dados represente melhor as dinâmicas populacionais e as unidades espaciais adotadas correspondam melhor às cidades funcionais, optou-se por testar os expoentes de

escalamento contra diversas definições de cidade. As definições se baseavam em valores limites de população e densidade.

Com um script no R, realizou-se cortes recursivos na base de dados, adotando todos os valores de densidade entre 0 e 15 e de população entre 0 e 100000. Com as bases subdivididas, calculou-se um modelo de regressão simples entre a população e as variáveis básicas estudadas. Comparou-se os resultados encontrados para as duas bases de dados

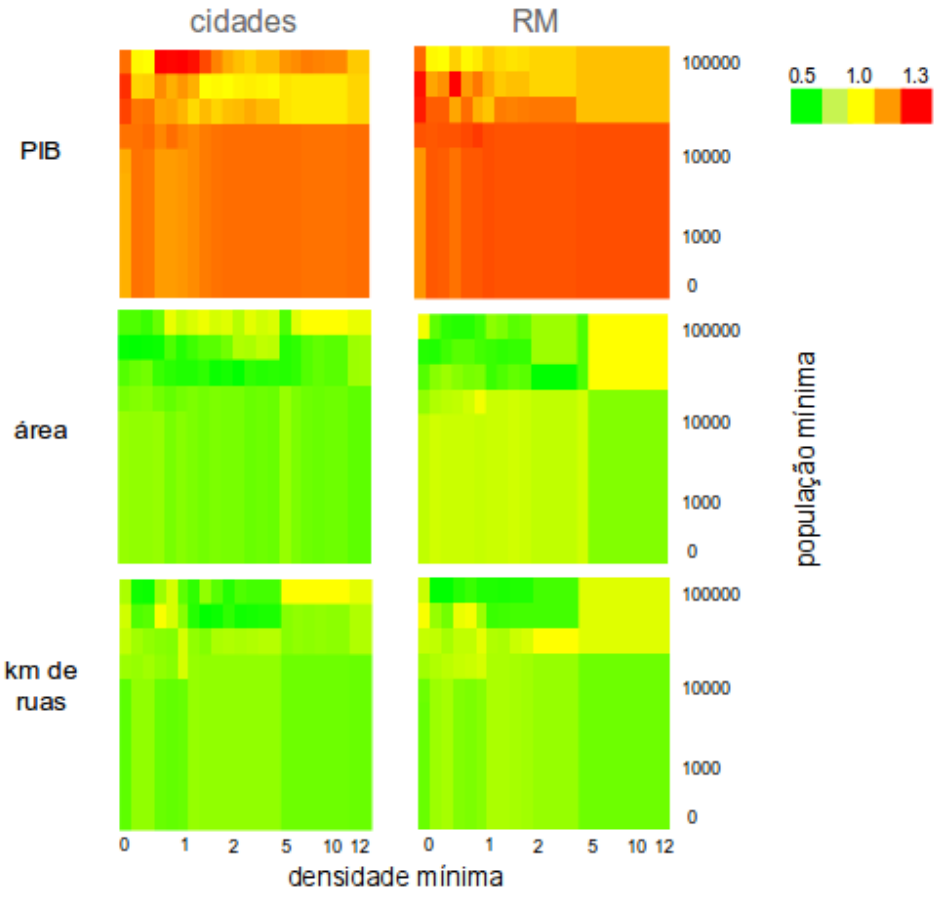


Figura 20 – Expoente de escalamento referente aos valores mínimos de corte para densidade e população da base de dados de Cidades (esquerda) e de Regiões Metropolitanas (direita)
fonte:produção do próprio autor

Os resultados indicam que os expoentes de escalamento são mais sensíveis à densidade do que à população. De fato, os expoentes só variam com a população a partir de um valor limite relativamente alto (cortes de 50000), quando poucas cidades restam à base. A variação com relação à densidade se observa de maneira mais contínua. Apesar disso, é

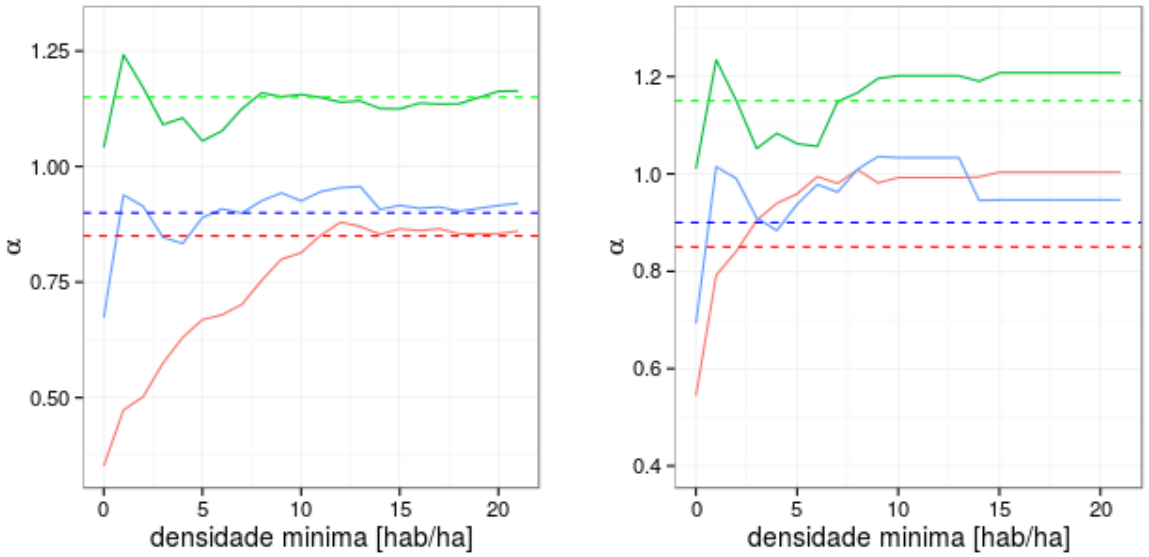
possível observar, também, um limite de densidade que parece causar uma descontinuidade nos valores dos expoentes de área e de quilômetros de ruas.

Divergindo dos resultados encontrados anteriormente em outros estudos (ARCAUTE et al., 2015; COTTINEAU et al., 2015), não se observa uma extrema sensibilidade dos expoentes com relação à definição adotada de cidade. De uma maneira geral, PIB per capita escala de maneira superlinear e a área o a infraestrutura viária escalam de maneira sublinear. Para algumas combinações específicas de valores limite de corte e densidade observa-se escalamentos lineares.

Os bancos de dados de cidades e de regiões metropolitanas apresentam alguns padrões ligeiramente distintos de sensibilidade. De uma maneira geral, os expoentes encontrados a partir do banco de dados de cidades se aproximam mais dos observados em outros sistemas urbanos do mundo. Principalmente os expoentes de área e quilômetros de ruas no banco de dados de Regiões Metropolitanas, que estabilizam em valores próximos do escalamento linear. Isso pode ser explicado pela heterogeneidade na definição de Região Metropolitana adotada pelos estados do país. Por isso, optou-se por trabalhar com o banco de dados de cidades. Nas análises que se seguem, esse será o banco utilizado.

O resultado encontrado para a inclinação α da regressão do escalamento nos dois bancos de dados referente a uma densidade populacional mínima é apresentada na Figura **??**. O valor de α esperado para cada uma das variáveis é indicado nas linhas horizontais tracejadas. As cores se referem à Área(vermelho), PIB(verde) e Km de ruas(azul). É possível observar que os padrões de escalamento, quando analisados a partir do banco de dados das Regiões Metropolitanas (direita), foge daquele observado em outros sistemas urbanos do mundo (BETTENCOURT et al., 2007). Os expoentes de escalamento do banco de dados de cidades parece convergir para os valores observados em outros sistemas urbanos do mundo. Além disso, as inclinações das regressões no banco de dados das regiões metropolitanas parecem variam de maneira mais volátil com relação ao corte na densidade do que no banco de dados de cidades.

É possível observar certa robustez dos padrões gerais de escalamento do banco de dados de cidades. O PIB varia numa faixa de inclinação sempre acima de 1 e as variáveis de infraestrutura (área e km de ruas) variam dentro de uma faixa de inclinação sempre menor do que 1. Mais do que isso, os valores do expoente de escalamento parecem convergir e se estabilizar em valores muito próximos dos esperados teoricamente. Esse resultado reitera a previsão teórica proposta por Bettencourt.



(a) cidades

(b) regiões metropolitanas

Figura 21 – inclinação α da função de regressão linear encontrada para cada uma das variáveis básicas, referente a uma densidade mínima de corte nos bancos de dados estudados
fonte:produção do próprio autor

Porém, observa-se que existe uma faixa muito estreita de corte densidade que otimiza todas as regressões simultaneamente. O valor ótimo de corte de densidade fica entre 7 e 8 habitantes por hectare. Além disso, é possível observar a grande volatilidade da área, que sai de valores abaixo de 0.5 de inclinação e chega a valores de próximos de 0.9, se estabilizando apenas a partir de um corte de cidade mais densas do que 13 habitantes por hectare. o PIB e a infraestrutura viária, por sua vez, se estabilizam mais cedo, respectivamente nos cortes de 8 e 5 habitantes por hectare.

A volatilidade da área fica evidente, também, na Figura 22, que indica o valor do r^2 das funções de regressão encontrados para cada valor de corte de densidade. Além de flutuarem menos, o PIB e a infraestrutura viária permanecem a maior parte do gráfico com valores elevados de r^2 (em torno de 0.8) ao passo que a área só atinge esse patamar de r^2 para valores de corte altos (em torno de 12 habitantes por hectare).

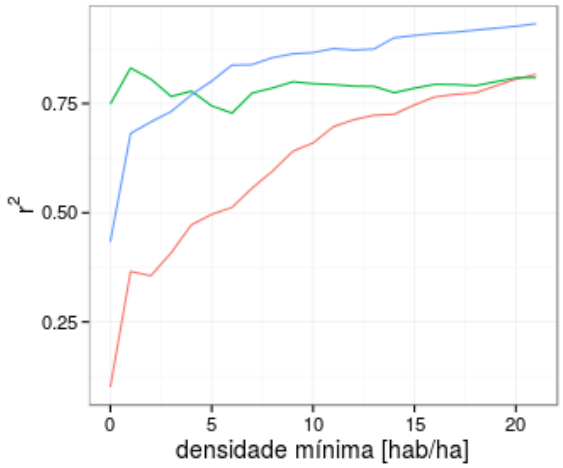


Figura 22 – r^2 da função de regressão linear referente a uma densidade mínima de corte da base de dados de cidades encontrada para cada uma das variáveis básicas
fonte:produção do próprio autor

O que a análise de robustez parece indicar é que a verdadeira inclinação da reta de escalamento da área com a população nas cidades brasileiras é de aproximadamente 0.88, se aproximando mais do expoente de escalamento da área política das cidades de Michigan (0.88) e do valor teórico encontrado por Bettencourt (0.86) do que das cidades de todo os EUA (0.78) ((BETTENCOURT, 2013)).

Baseado na análise de robustez, optou-se por trabalhar com os dados referentes às cidades brasileiras, descartando o banco de dados de Regiões Metropolitanas. Produziu-se um corte final nessa base de 10 habitantes por hectare e as análises de escalamento foram, então refeitas. A nova base de dados de cidades brasileiras, após a adoção do critério de corte, contém 95 cidades. A tabela **??** abaixo apresenta uma comparação dos parâmetros dos modelos de regressão encontrados para os dois bancos de dados após o corte de densidade (¿ 10 hab/ha). É possível observar que as os valores encontrados de α são maiores para todas as variáveis no banco de dados de regiões metropolitanas.

O PIB, após o corte, passou a escalar com uma inclinação α de 1.16, muito próximo daquilo esperado teoricamente (1.15), com um r^2 de aproximadamente 0.81.

Essa variável permite uma compração com a metodologia proposta por Clauset e colegas (CLAUSET; SHALIZI; NEWMAN, 2009) e implementada por Alvez e colegas (ALVES et al., 2014) para as cidades brasileiras. Usando o critéro de corte puramente estatístico do

Tabela 3 – Padrões de escalamento após o corte

	Cidades			Regiões Metropolitanas		
	intercepto	α [C.I.]	r^2	intercepto	α [C.I.]	r^2
Área	-2.85	0.81 [0.69, 0.93]	0.82	-3.41	0.99 [0.87, 1.11]	0.98
PIB	-0.03	1.16 [1.03, 1.27	0.81	-0.36	1.20 [1.03, 1.37]	1.00
Km de Ruas	-4.29	0.92 [0.85, 1.00]	0.97	-4.37	1.03 [0.78, 1.28]	0.94
G	-4.32	0.08 [-0.07, 0.23]	0.02	-4.73	0.23 [-0.13, 0.6]	0.29

início do regime de distribuição de lei de potência, Alvez e colegas encontram um expoente de escalamento do PIB com relação à população de 1.3, maior do que o encontrado utilizando a metodologia aqui proposta e maior do que os valores normalmente encontrados em outros sistemas urbanos do mundo.

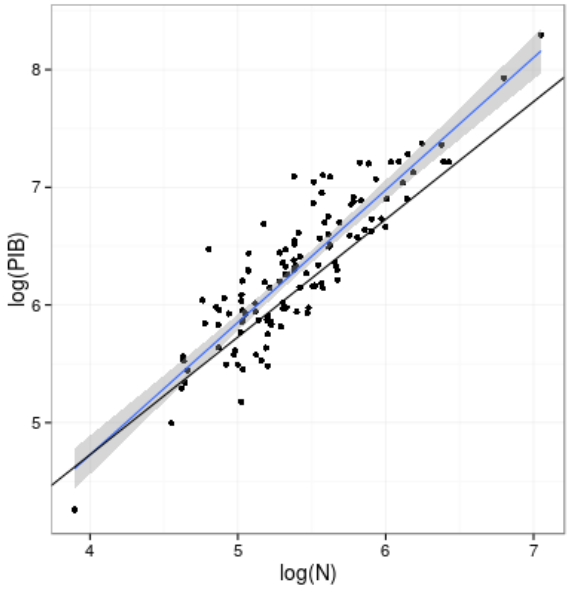


Figura 23 – Escalamento entre a População e o PIB das cidades estudadas após a adoção do critério de corte
fonte:produção do próprio autor

Após a adoção do critério de corte, a área passou a escalar com uma inclinação α de 0.81 e r^2 de 0.82. Assim como o PIB, o expoente de escalamento se aproxima razoavelmente bem daquilo encontrado na literatura.

A infraestrutura viária, após a adoção do critério de corte, passou a escalar com uma inclinação α de 0.92 e r^2 0.97. Assim como a área, o total de quilômetros de ruas das cidades brasileiras parece escalar com um expoente ligeiramente maior do que aquele usualmente encontrado na literatura.

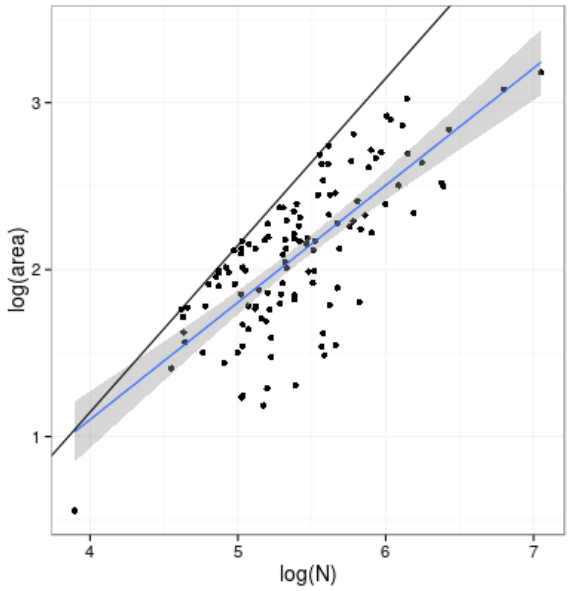


Figura 24 – Escalamento entre a População e a Área das cidades estudadas após a adoção do critério de corte
fonte:produção do próprio autor

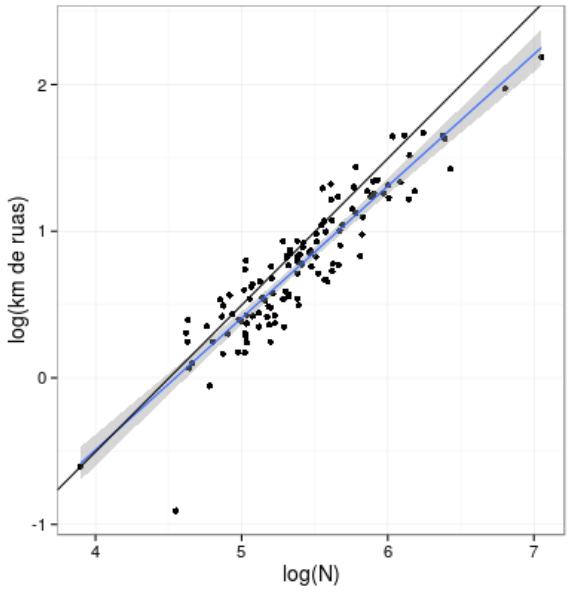


Figura 25 – Escalamento entre a População e infraestrutura viária das cidades estudadas após a adoção do critério de corte
fonte:produção do próprio autor

O G, após a adoção do critério de corte, passou a escalar com uma inclinação α de 0.08 e um r^2 de 0.02. Ele se afasta da esperada inclinação 0 apenas por conta das duas maiores cidades do País (Rio de Janeiro e São Paulo).

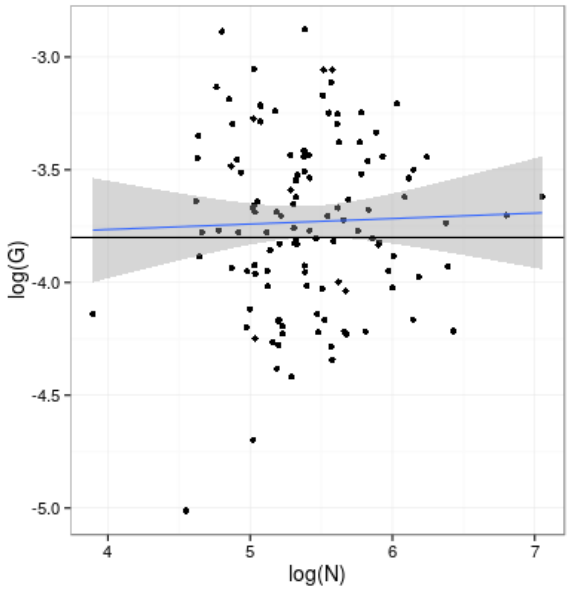


Figura 26 – Escalamento entre a População e do G das cidades estudadas após a adoção do critério de corte
fonte:produção do próprio autor

A tabela 4 apresenta os resultados consolidados.

	intercepto	α [C.I.]	r^2	$valor_p$
população x PIB	0.11905	1.16 [1.03, 1.27]	0.81	2.2e-16
população x km de ruas	-4.46614	0.92 [0.85, 1.00]	0.97	2.2e-16
população x área	-2.83477	0.81 [0.69, 0.93]	0.82	2.2e-16
população x G	4.58519	0.08 [-0.07, 0.23]	0.02	0.004302

Tabela 4 – Escalamento das variáveis com a população das cidades estudadas após adoção de critério de corte

Fica claro que, após a adoção do critério de corte, as cidades brasileiras parecem seguir de maneira muito próxima os padrões encontrados em outras cidades do mundo. A única variação perceptível parece ser o fato de que as cidades brasileiras escalam suas variáveis infraestruturais com um expoente de escalamento um pouco maior do que têm sido usualmente observado.

Buscando entender as implicações do fenômeno de escalamento das cidades de uma maneira mais ampla, levantou-se uma série de dados das cidades brasileiras em diversas fontes oficiais. A tabela 5 apresenta a lista de variáveis estudadas, bem como suas respectivas fontes.

Dado	Fonte
População	IBGE
Área	IBGE
PIB	IBGE
Km de Ruas	Open Street Maps
Pessoas Com Registro	IBGE
Pessoas Alfabetizadas	IBGE
Renda Média	IBGE
Moradores com Banheiros	IBGE
Moradores Conectados na Rede de Esgoto	IBGE
Moradores com Abasstecimento de Água de Rede	IBGE
Moradores com Lixo Coletado	IBGE
Moradores com Energia Elétrica	IBGE
Moradores com Banheiro Exclusico	IBGE
Moradores com Lixo Jogado em Terreno Baldio	IBGE
Moradores com Lixo Queimado	IBGE
Vítimas de Acidentes de Trânsito	DATASUS
Vítimass de Homicídios	DATASUS
Vítimas de Suicídio	DATASUS
Despesa Corrente	Ministério da Fazenda
Despesa de Capital	Ministério da Fazenda
Despesa Orçamentária	Ministério da Fazenda
Despesa por Função Total	Ministério da Fazenda
Densas por Transferência	Ministério da Fazenda
Receita Orçamentária	Ministério da Fazenda
Receita Corrente	Ministério da Fazenda
Receita Tributária	Ministério da Fazenda
Receita de Capital	Ministério da Fazenda
Receita Tributária de Impostos	Ministério da Fazenda
Receita Tributária IPTU	Ministério da Fazenda
Receita Tributária ISS	Ministério da Fazenda
Receita Tributária Taxas	Ministério da Fazenda

Tabela 5 – Variáveis estudadas e suas respectivas fontes

Com essa seleção de variáveis, buscou-se abranger uma gama de dados que representassem diversas instâncias da vida urbana. Todas as variáveis que se relacionam ao acesso dos moradores a serviços básicos (água, coleta de lixo, energia) buscarem mapear, em algum nível, a segregação com relação a esses bens. As variáveis de vítimas tentavam mapear se as interações crescentes (teoricamente) implicavam crimes igualmente crescentes. Já os dados de orçamento e receita se explicam pela análise de eficiência econômica das cidades. Os resultados encontrados para os escalamentos antes de aplicar o parâmetro de corte por densidade são apresentados na tabela 6. É possível notar que o padrão de escalamento das cidades brasileiras parecem possibilitar mais acesso às amenidades e qualidade de vida da população.

Tirando a raça Branca, todas as outras raças escalam superlinearmente, indicando que as cidades parecem ser compostas de populações mais heterogêneas à medida que crescem. As variáveis de Entorno e de Acesso a serviços Básicos também indicam melhor qualidade de vida nas cidades maiores. O total de moradores conectados à rede de esgoto, abastecidos de água a partir de rede e com o lixo coletado escalam de maneira superlinear com a população. O total de moradores com acesso à energia elétrica, a medidores de energia elétrica e a banheiros exclusivos pareceu escalar de maneira linear com a população. O total de moradores que tinham seu lixo queimado ou jogados em terreno baldio escalou de maneira sublinear com a população (ainda que essas duas últimas variáveis apresentaram baixo valor no r^2).

Com relação às vítimas, suicídios e acidentes de trânsito parecem diminuir ao passo que Homicídios parecem aumentar, confirmando a hipótese das interações sociais como preditores dos escalamentos. Ao mesmo tempo, as despesas de uma cidade parecem escalar sublinearmente com relação à população das cidades, com a exceção de despesas por transferência. Já a receita parece escalar de maneira superlinear com a população, com impostos e taxas crescendo de maneira não proporcional à população.

Os padrões de escalamento encontrados após o corte de densidade (tabela 7) parecem seguir de perto aqueles encontrados antes do corte. Como variações, nota-se a população alfabetizada que sair de um regime de escalamento linear antes do corte para um regime de escalamento superlinear. A renda escala de maneira mais intensa após o corte. O total de vítimas em acidentes de trânsito passa a escalar de maneira superlinear, bem como o total de moradores com o lixo queimado. No resto das variáveis, observa-se apenas algumas acentuações, mas nenhuma mudança qualitativa no padrão de escalamento.

	Pré-Corte		
variável	**intercepto**	α **[C.I.]**	r^2
Área	1.68	0.54 [0.52, 0.57]	0.05
PIB	0.27	1.01 [0.99, 1.02	0.76
Km de Ruas	-2.88	0.69 [0.67, 0.71]	0.44
Pessoas Com Registro	-0.69	0.98 [0.94, 1.02]	0.64
Pessoas Alfabetizadas	-0.26	1.02 [0.98, 1.06]	0.66
Renda Média	2.30	0.12 [0.10, 0.13]	0.11
Moradores com Banheiros	-0.56	1.00 [0.99, 1.01]	0.96
Moradores Conectados na Rede de Esgoto	-3.98	1.51 [1.43, 1.58]	0.58
Moradores com Abastecimento de Água de Rede	-0.54	1.09 [1.05, 1.13]	0.70
Moradores com Lixo Coletado	-0.69	1.12 [1.07, 1.16]	0.70
Moradores com Energia Elétrica	-0.14	1.02 [0.98, 1.06]	0.67
Moradores com Banheiro Exclusico	-0.19	1.02 [0.97, 1.06]	0.67
Moradores com Lixo Jogado em Terreno Baldio	-0.97	0.72 [0.63, 0.81]	0.17
Moradores com Lixo Queimado	1.63	0.14 [0.07, 0.21]	0.01
Vítimas de Acidentes de Trânsito	-3.33	0.92 [0.89, 0.96]	0.68
Vítimass de Homicídios	-4.21	1.11 [1.07, 1.15]	0.73
Vítimas de Suicídio	-2.47	0.65 [0.62, 0.68]	0.58
Despesa Corrente	3.30	0.92 [0.91, 0.94]	0.93
Despesa de Capital	2.69	0.92 [0.89, 0.95]	0.76
Despesa Orçamentária	3.41	0.92 [0.90, 0.93]	0.92
Despesa por Função Total	3.41	0.92 [0.90, 0.93]	0.92
Densas por Transferência	0.82	1.10 [1.02, 1.19]	0.38
Receita Orçamentária	3.33	0.98 [0.96, 0.99]	0.92
Receita Corrente	3.35	0.97 [0.95, 0.99]	0.92
Receita Tributária	0.67	1.32 [1.28, 1.36]	0.80
Receita de Capital	2.95	0.75 [0.68, 0.80]	0.33
Receita Tributária de Impostos	0.59	1.32 [1.29, 1.36]	0.80
Receita Tributária IPTU	-1.21	1.53 [1.45, 1.61]	0.55
Receita Tributária ISS	-0.04	1.38 [1.34, 1.42]	0.78
Receita Tributária Taxas	-0.52	1.34 [1.27, 1.40]	0.61

Tabela 6 – Escalamento das variáveis com a população das cidades estudadas antes da adoção de critério de corte

		Pós-Corte	
variável	**intercepto**	α **[C.I.]**	r^2
Área	-2.23	0.81 [0.69, 0.93]	0.82
PIB	0.03	1.16 [1.03, 1.27]	0.81
Km de Ruas	-4.31	0.92 [0.85, 1.00]	0.97
Pessoas Com Registro	-0.60	1.00 [0.86, 1.15]	0.72
Pessoas Alfabetizadas	-0.80	1.07 [0.82, 1.31]	0.50
Renda Média	2.00	0.15 [0.07, 0.22]	0.18
Moradores com Banheiros	-0.61	1.01 [0.99, 1.02]	0.99
Moradores Conectados na Rede de Esgoto	-1.49	1.08 [0.95, 1.21]	0.78
Moradores com Abasstecimento de Água de Rede	-0.72	1.02 [0.87, 1.54]	0.74
Moradores com Lixo Coletado	-1.07	1.07 [0.98, 1.16]	0.89
Moradores com Energia Elétrica	-1.73	1.12 [0.99, 1.23]	0.82
Moradores com Banheiro Exclusico	-0.15	1.01 [0.89, 1.13]	0.78
Moradores com Lixo Jogado em Terreno Baldio	-2.91	1.14 [0.74, 1.55]	0.30
Moradores com Lixo Queimado	-0.27	0.63 [0.30, 0.97]	0.16
Vítimas de Acidentes de Trânsito	-4.93	1.18 [0.98, 1.39]	0.65
Vítimass de Homicídios	-4.32	1.13 [1.01, 1.25]	0.82
Vítimas de Suicídio	-4.22	0.94 [0.74, 1.14]	0.55
Despesa Corrente	3.07	0.98 [0.90, 1.07]	0.87
Despesa de Capital	2.58	0.98 [0.83, 1.12]	0.72
Despesa Orçamentária	3.22	0.98 [0.89, 1.07]	0.86
Despesa por Função Total	3.22	0.97 [0.88, 1.06]	0.85
Densas por Transferência	-0.32	1.32 [0.96, 1.68]	0.42
Receita Orçamentária	3.02	1.06 [0.96, 1.16]	0.86
Receita Corrente	3.12	1.05 [0.94, 1.14]	0.85
Receita Tributária	1.21	1.29 [1.11, 1.46]	0.73
Receita de Capital	0.99	1.11 [0.84, 1.37]	0.48
Receita Tributária de Impostos	1.15	1.29 [1.11, 1.47]	0.72
Receita Tributária IPTU	0.62	1.30 [1.09, 1.51]	0.67
Receita Tributária ISS	0.43	1.35 [1.16, 1.54]	0.72
Receita Tributária Taxas	0.41	1.23 [1.04, 1.43]	0.68

Tabela 7 – Escalamento das variáveis com a população das cidades estudadas após a adoção de critério de corte

5 Origem das Leis de Escala em Cidades com ABM

Neste capítulo, apresenta-se a metodologia e os resultados do modelo baseado em agentes proposto para explicar a origem das leis de escala nas cidades.

5.1 Metodologia

A fim de melhorar a documentação e facilitar a compreensão, optou-se por descrever o modelo proposto atravéz da metodologia ODD de descrição de modelos baseados em agentes (GRIMMA et al., 2006).

- **Propósito**

O modelo de campo médio proposto por Bettencourt (BETTENCOURT, 2013) apresenta uma prossível explicação para o surgimento das leis de escala nas cidades a partir de interações sociais no nível do indivíduo. A estrutura proposta por ele, porém, apresenta algumas limitações no que diz respeito à heterogeneidades dos indivíduos. A fim de testar hipóteses, mapear relações causais e o papel das diversas variáveis na origem das leis de escala nas cidades, traduziu-se o modelo proposto por Bettencourt numa estrutura de modelagem por agentes (ABM). Partindo da mesma fundamentação teórica, se manteve, sempre que possível, os mesmos nomes e símbolos das variáveis.

Inicialmente, adotou-se a linguagem de programação NetLogo mas, após alguns testes, optou-se por migrar para a linguagem Python por questões de eficiência computacional e programabilidade. Os modelos foram rodados no Google Compute Engine, um ambiente de máquinas virtuais, fornecido pela Google através de contratos de infraestrutura como um serviço.

- **Definição de Variáveis e Escalas**

Assim como no modelo descrito por Bettencourt, partiu-se de uma cidade ideal quadrada de área A e lado L, habitada por uma população N durante uma escala temporal de T passos de tempo. A cidade é comporta por quarteirões e ruas, que se distribuem no espaço segundo um grid regular. Nas simulações, trabalhou-se com 100 passos de tempo. A população varia de 10 a 10000 habitantes, sendo o escalamento acima desses níveis, apesar de importante, impossibilitado por questões computacionais. A tabela 8 sistematiza as variáveis sobre a cidade no modelo.

VARIÁVEL	DESCRIÇÃO	PARÂMETRO
Área	área da cidade	A
Lado	seção transversal da cidade	L
População	população da cidade	pop
Tamanho do Quarteirão	tamanho de um quarteirão	city block size
Tempo	número de passos de tempo simulados em cada rodada	τ

Tabela 8 – Definição de Parâmetros da cidade

No modelo, uma cidade é composta por habitantes caracterizados pelo seu acumulado de produção socioeconômica y realizada ao longo do tempo. Os agentes caminham aleatoriamente pelo espaço. Se, ao longo de seu percurso, outro agente entra dentro de sua área de interação a_0, os dois agentes produzem uma interação que se traduz numa produtividade socioeconômica média g, a ser adicionada ao acumulado y de cada um dos envolvidos no encontro. A tabela 9 apresenta, de maneira sistematizada, as variáveis utilizadas.

VARIÁVEL	DESCRIÇÃO	PARÂMETRO
rendimento acumulado	produção total do indívuo a partir dos encontros menos o custo de deslocamento	y
produtividade média por interação	quanto um indivíduo produz, em média, ao interagir com outro	g
área de interação	alcance da área de interação média dos indivíduos	a_0
custo de deslocamento	quanto é descontado do indivíduo a cada passo caminhado	ϵ

Tabela 9 – Definição de Parâmetros dos agentes

- **Visão geral do processo e programação**

Os valores de g e de Y são sorteados aleatoriamente no início da simulação, seguindo uma distribuição Gaussiana de probabilidades determinada pelo valor médio ($\bar{g_{ag}}$, $\bar{y_{ag}}$) e pelo desvio padrão (σg_{ag}, σy_{ag}). Diversos valores de média e de desvio padrão foram testados para essas duas variáveis.

A cada passo de tempo, cada agente faz uma opção aleatória de movimento para um dos possíveis espaços vizinhos, utilizando uma vizinhança de Von Neumann para fazer essa opção. Os espaços ocupados por quarteirões não são utilizados nessa escolha. Isso faz com que cada agente siga um padrão de deslocamento por Random Walk (PEARSON,

1905) em um grid regular. A cada passo dado, o custo de transporte ϵ é deduzido da energia individual Y do agente. Ao longo das explorações do modelo, a hipótese de que a densidade populacional refletia em custos de transporte maiores para o indivíduo foi testada. Nessas versões, o custo de transporte ϵ era corrigido por $\epsilon * (1 + densidade)$ onde *densidade* adotava um valor entre 0 e 1 calculado a partir do total de agentes nas ruas dividido pelo total de espaços disponíveis na cidade.

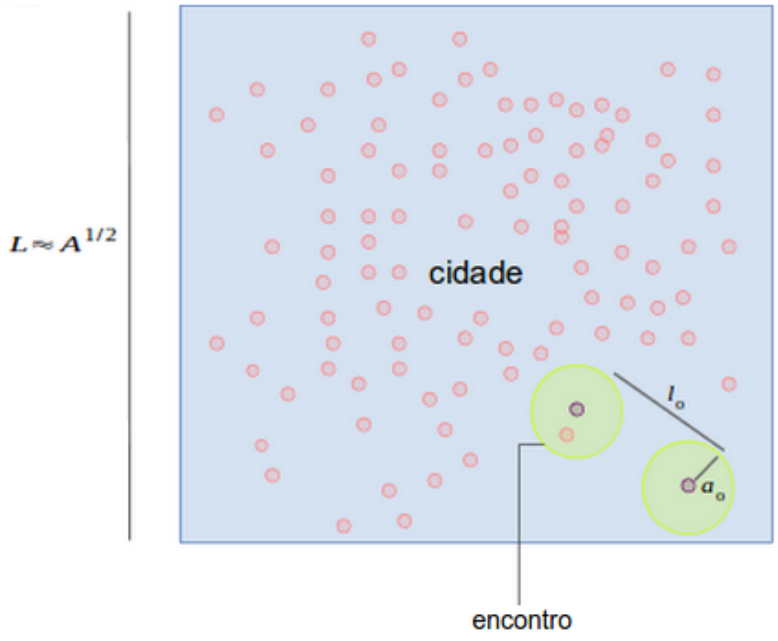


Figura 27 – Esquematização da visão geral do modelo proposto
fonte:produção do próprio autor

Após movimentar todos os agentes, computa-se as interações entre agentes que estão dentro da área de interação a_0 de outros agentes. Caso isso aconteça, cada participante da interação tem seu valor de produção por interação médio $\bar{g_{ag}}$ adicionado ao seu rendimento acumulado y.

Caso tenha seu valor de energia individual = 0 em algum passo de tempo, o agente morre e desaparece do modelo. O valor de produção média por interação $\bar{g_{ag}}$ se mantém constante com o tamanho da população da cidade. Para algumas configurações testadas, porém, o custo de deslocamento variava com a densidade populacional.

Diversas combinações de parâmetros foram testadas. Para se chegar aos expoentes de escalamento em cada parametrização, simulou-se cada população dentro de um vetor de populações $N = n_1, n_2, ..., n_n$ habitando cidades de cada área de um vetor de áreas

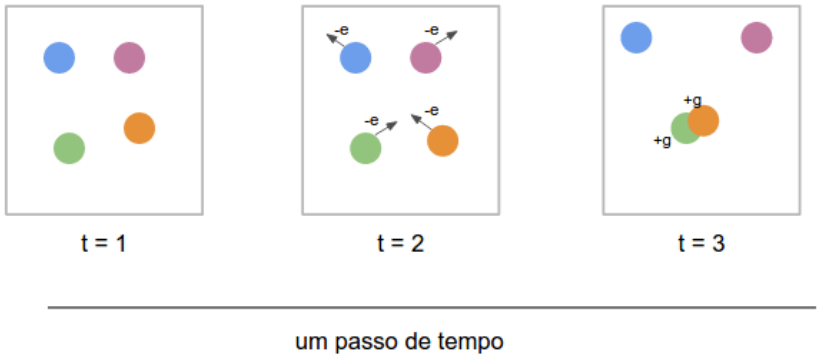


Figura 28 – Esquematização de um passo de tempo do modelo
fonte:produção do próprio autor

$A = a_1, a_2, ..., a_n$. Cada sistema urbano era composto por cidades em todas as combinações entre N e A. Assim, uma matriz de sistemas urbanos possíveis possuía dimensões N x A.

Nessa matriz, denominou-se $C_n a$ uma cidade de população n e área a. Para cada cidade da matriz, computou-se a taxa de morte como sendo $= \frac{N_i - N_f}{N_i}$, onde N_i é a população inicial em $t = 0$ e N_f é a população final no $t = 100$. Para cada valor de população n, selecionava-se apenas as cidades $C_n a$ que apresentavam taxas de morte inferior a 0.1. Com essas cidades, compunha-se um vetor de cidades prováveis, que era adotado como o vetor final de cidades da parametrização. Esse procedimento foi repetido para diversas parametrizações até que se encontrou uma parametrização satisfatória.

Em seguida, variou-se sistematicamente apenas um dos parâmetros do modelo e se avaliou seu impacto na produção socioeconômica Y final da cidade. Com isso, foi possível esclarecer o papel de cada parâmetro no desempenho socioeconômico das cidades.

- **Conceitos do projeto**

O modelo foi desenhado buscando encontrar o comportamento emergente das cidades no que diz respeito ao escalamento das variáveis urbanas com a população. Inclui-se, apenas, dinâmicas locais e, à medida que se analisa a produção comparada de cidades de tamanhos diferentes, observa-se a emergência das leis de escala.

O modelo, no presente estado de desenvolvimento, não apresenta características adaptativas. O Fitness dos agentes pode ser facilmente traduzido pelo Y individual. Ao se aproximar de 0, um agente não muda seu comportamento. Da mesma maneira, os agentes não condicionam seu comportamento a partir de previsões do futuro. O caminho percorrido

por ele é aleatório. A estocacidade está presente nesse caráter aleatório do caminho assim, na distribuição espacial dos agentes, e na distribuição dos parâmetros.

- **Inicialização**

A partir de um valor de área a e população n retirados do vetor $C_n a$, uma cidade é criada e quarteirões são inseridos seguindo um tamanho pré-determinado. Em seguida, os agentes que compõe a população n são distribuídos aleatoriamente pelo espaço nos trechos de rua da cidade.

- **Inputs**

Uma vez inicializado, o modelo não recebe novos inputs.

5.2 Resultados

O modelo proposto se constituí, inicialmente, de agentes, dispostos aleatoriamente no espaço. Seguindo os passos do modelo de campo médio proposto por Bettencourt (BETTENCOURT, 2013), implementou-se, então, um grid de quarteirões estilo Manhattan a fim de garantir que os agentes interagissem apenas nos espaços de infraestrutura. A Figura 29 apresenta uma visualização da condição inicial de distribuição espacial dos agentes no modelo já com a implementação do grid de quarteirões.

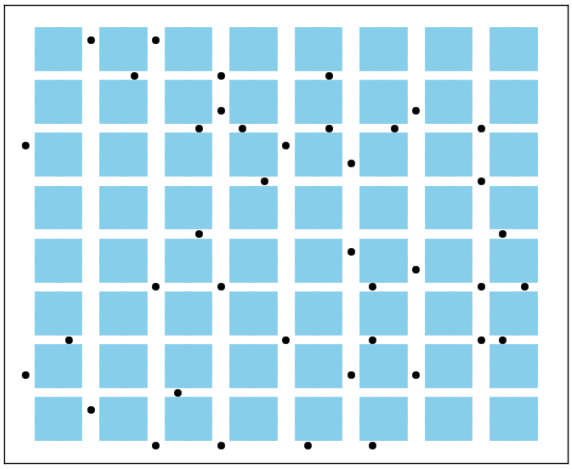


Figura 29 – Exemplo de condição inicial do modelo
fonte:produção do próprio autor

Num primeiro momento, testes de consistência do modelo foram realizados a fim de validá-lo. Um dos testes de consistência se constituía em observar a trajetória da população N, dos encontros e da produção socioeconômica Y de uma cidade muito pouco densa contra a evolução das mesmas variáveis em uma cidade consideravelmente densa.

A Figura 30 apresenta a evolução temporal da população N, da produção socioeconômica y e do total de encontros que acontecem dentro de uma cidade com baixa densidade populacional. Ao lado da condição inicial de distribuição populacional da cidade, apresenta-se as séries temporais das variáveis. Como era de se esperar, os indivíduos morrem a intervalos irregulares mas constantes até pouquíssimos habitantes sobreviverem. Os que sobrevivem são os que foram criados com maior y inicial e/ou que produziram mais encontros nos primeiros passos de tempo. Alguns poucos encontros acontecem enquanto uma parcela significativa da população ainda não morreu. Em seguida, a taxa de encontros cai para zero (0). A somatória da produção socioeconômica Y varia de maneira errática mas crescente durante os primeiros passos de tempo, até que começa a decair de maneira constante. Esse segundo regime de comportamento de Y corresponde aos poucos agentes que sobram na simulação, tendo o custo de transporte deduzido de seus somatórios e com uma densidade baixa demais para produzir qualquer encontro.

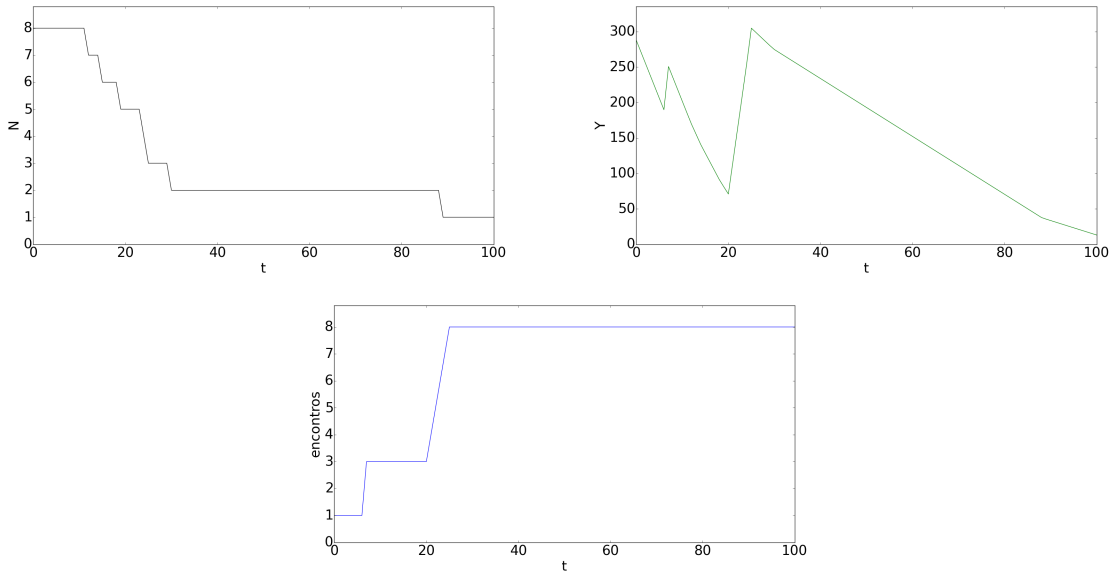


Figura 30 – Evolução temporal de uma cidade pouco densa
fonte:produção do próprio autor

Quando se trata de uma cidade mais densa, porém, a dinâmica muda substancialmente (Figura 31). A população se sustenta, em larga escala, ao longo dos passos de tempo. Após uma fase inicial com algumas poucas mortes decorrentes de flutuações

aleatórias, a população da cidade se estabiliza em um valor fixo. O somatório da produção socioeconômica Y e o total de encontros cresce de maneira constante. Com isso, o valor acumulado final de produção socioeconômica das cidades depende da quantidade de passos temporais simulados. Mais passos de tempo levam a maior produção socioeconômica de maneira linear e constante.

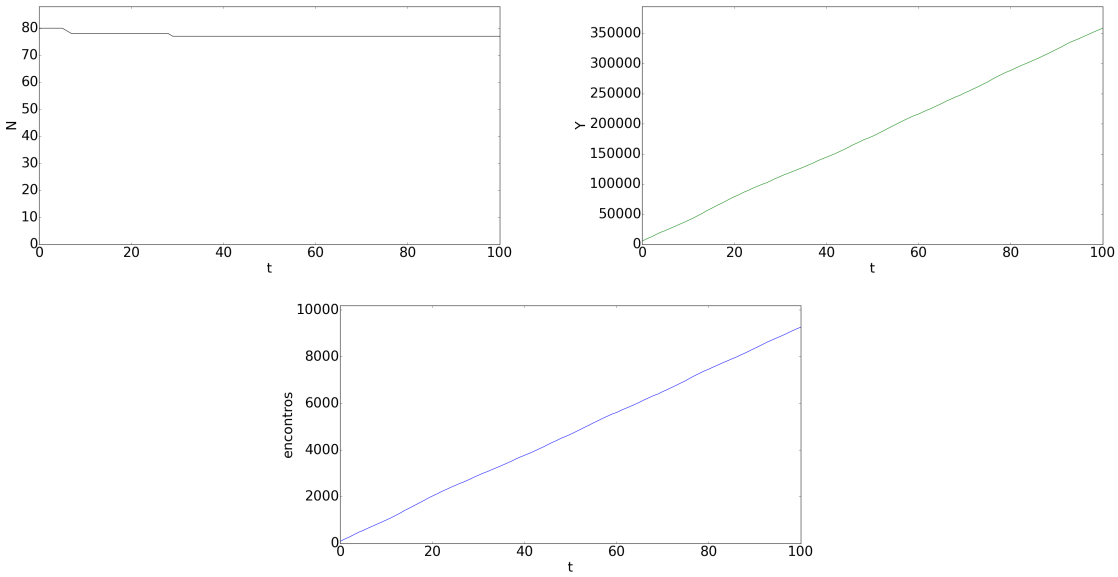


Figura 31 – Evolução temporal de uma cidade muito densa
fonte:produção do próprio autor

O modelo se comportou conforme o esperado nas situações de densidade populacional distintas. Na cidade que possui uma grande quantidade de indivíduos por metro quadrado, um indivíduo realiza encontros suficientes para se sustentar. Seu acumulado de produção socioeconômica y cresce no tempo. Já na situação de uma cidade com baixa densidade populacional, o indivíduo realiza menos encontros do que seria necessário para se manter vivo ao longo do tempo. Ainda que esse indivíduo produza alguns encontros esporádicos (descontinuidades no gráfico), seu acumulado de produção socioeconômica individual y tende à zero.

Porém, Quando se testa o modelo para diversas combinações de população N e área A algumas inconsistências podem ser observadas. A Figura 32 apresenta a taxa de morte encontrada em cada cidade $C_n a$ na matriz de valores de N e A testados. Nota-se que, de acordo com o esperado, cidades com áreas muito grandes e populações muito pequenas não sustentam suas populações no tempo (altas taxas de morte). Não existe, porém, nenhum feedback no modelo de controle da densidade. Assim, cidades irrealmente densas apresentam baixas taxas de morte. De fato, não existe limite para a densidade

populacional nessa versão do modelo. Essa característica é contrafactual, dado que cidades reais se adensam com a população, mas não de maneira indefinida. Em última instância, existe uma limitação física para o crescimento da densidade, expressa nos termos na capacidade de suporte dos materiais de construção e no espaço físico propriamente dito.

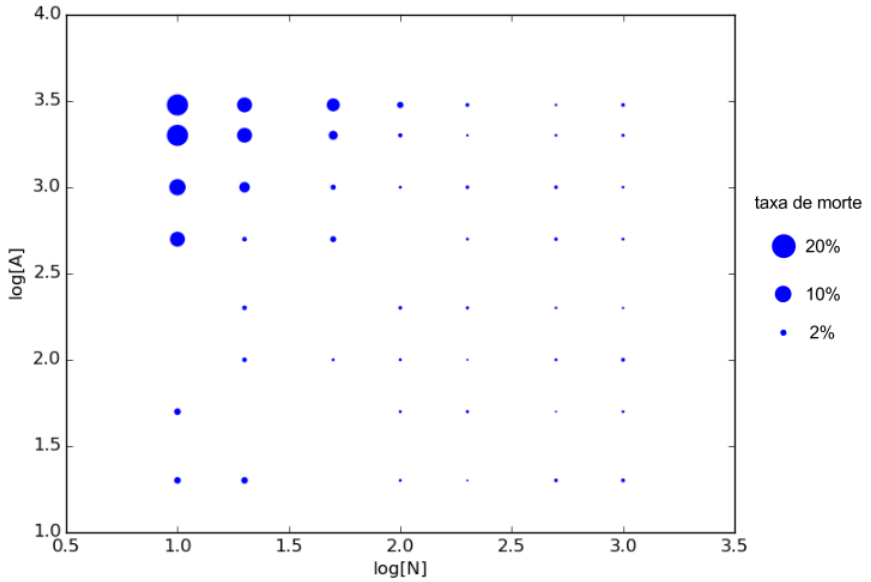


Figura 32 – Taxa de morte referente àos valores iniciais de População e Área na versão onde o custo de transporte é invariante com a densidade
fonte:produção do próprio autor

Para solucionar esse problema, propõe-se uma versão do modelo na qual o custo de transporte se comporte como uma função da densidade populacional. O pressuposto é que cidades mais densas têm custos de transporte associados maiores. Seja em termos do trânsito ou de investimentos em sistemas de transporte de massa mais caros (metrô, trem). Nessa nova versão do modelo, o custo de transporte ϵ é definido pela função

$$\epsilon = passo * (1 + densidade)$$

A Figura 33 apresenta a taxa de morte encontrada em cada cidade $C_n a$ na matriz de valores de N e A testados. É possível notar que as cidades muito densas passam a apresentar elevadas taxas de morte. Apenas uma faixa intermediária de combinações N e A se sustenta ao longo do tempo. Essa é uma configuração muito mais factual e, por conta disso, o modelo utilizado nas parametrizações e nos testes de hipótese tem o custo de transporte variando como uma função da densidade.

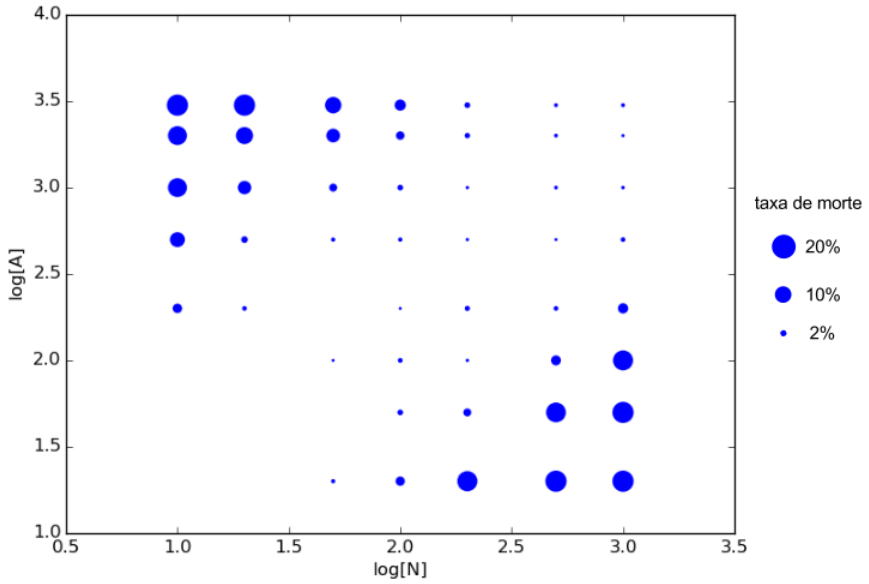


Figura 33 – Taxa de morte referente àos valores iniciais de População e Área na versão onde o custo de transporte é uma função da densidade fonte:produção do próprio autor

Com o funcionamento básico do modelo corrigido, dedicou-se à sua parametrização. Foram testadas pouco mais de 1000 combinações entre os parâmetros e os resultados do expoente de escalamento de Y, é apresentado na figura 34 . Cada ponto no gráfico representa uma rodada específica do modelo, com parametrizações específicas. É possível notar que, ainda que alguns grupos de parâmetros apresentaram valores sublineares para o escalamento da produção socieconômica Y com a população N das cidades, a grande maioria se encontra na faixa de valores superlineares, conforme o esperado. Mais do que isso, valores extremamente altos, muito acima do esperado teoricamente também se mostraram raros. O sistema parece convergir para os valores esperados (linhas tracejadas).

A partir dos valores encontrados, selecionou-se a parametrização que mais se aproximava dos valores esperados pela teoria (BETTENCOURT, 2013). A tabela **??** apresenta os valores das variáveis da parametrização escolhida.

Essas variáveis produzem os padrões de escalamento que são observados nas figuras 35 e 36 .

A parametrização selecionada gerou um expoente de escalamento $\beta = 1.168$ e $\alpha = 0.8785$ muito próximos àquilo encontrado pela estrutura proposta por Bettencourt (BETTENCOURT, 2013).

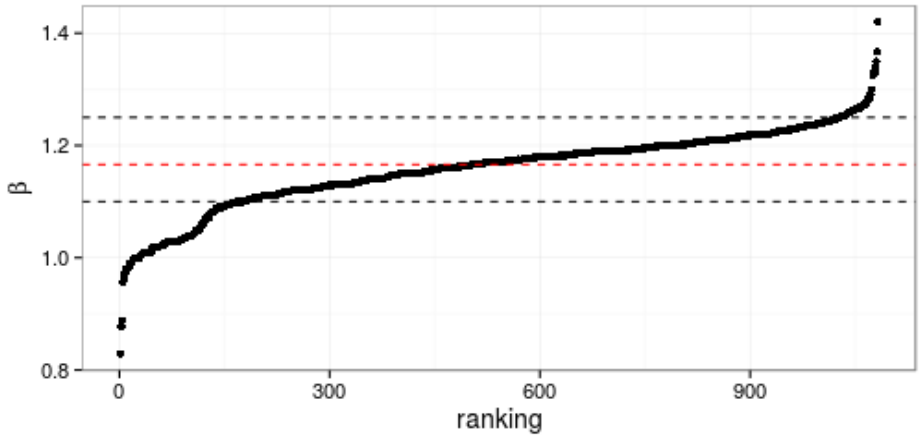


Figura 34 – expoentes entre Y e N para diversas parametrizações
fonte:produção do próprio autor

Tabela 10 – Parametrização capaz de reproduzir a hipótese de escalamento

Variável	Valor
a_0	2
ϵ	1.5
$\bar{g_{ag}}$	10
σg_{ag}	5
$\bar{y_{ag}}$	50
σy_{ag}	10

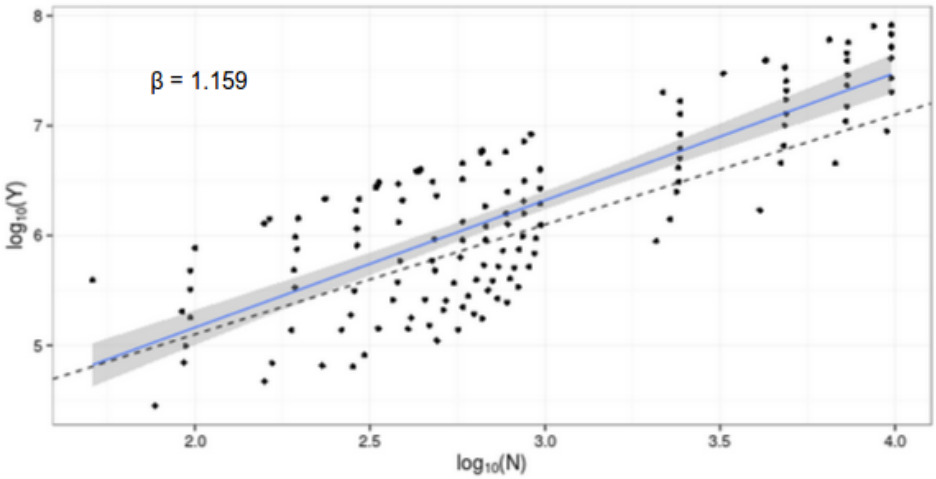


Figura 35 – expoentes entre Y e N na melhor parametrização
fonte:produção do próprio autor

A estrutura de ABM permite uma exploração apurada dos parâmetros do modelo. O passo seguinte foi entender o papel de cada variável na produção socioeconômica das cidades. Para isso, variou-se cada um dos parâmetros sistematicamente, mantendo todos

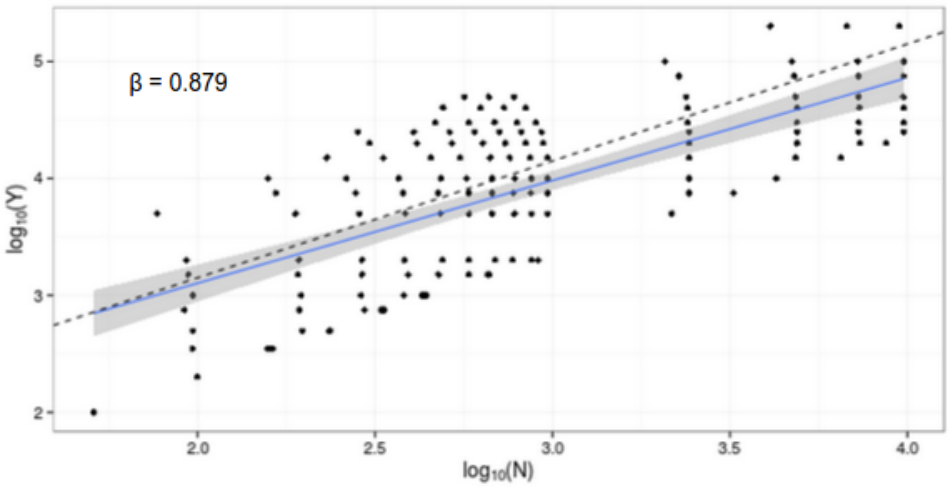


Figura 36 – expoentes entre *Area* e N na melhor parametrização
fonte:produção do próprio autor

os outros constantes e se mediu o total da produção socioeconômica final. As simulações foram rodadas 10 vezes para cada valor de parâmetro e os resultados foram mediados para evitar possíveis flutuações aleatórias. Os resultados são apresentados a seguir.

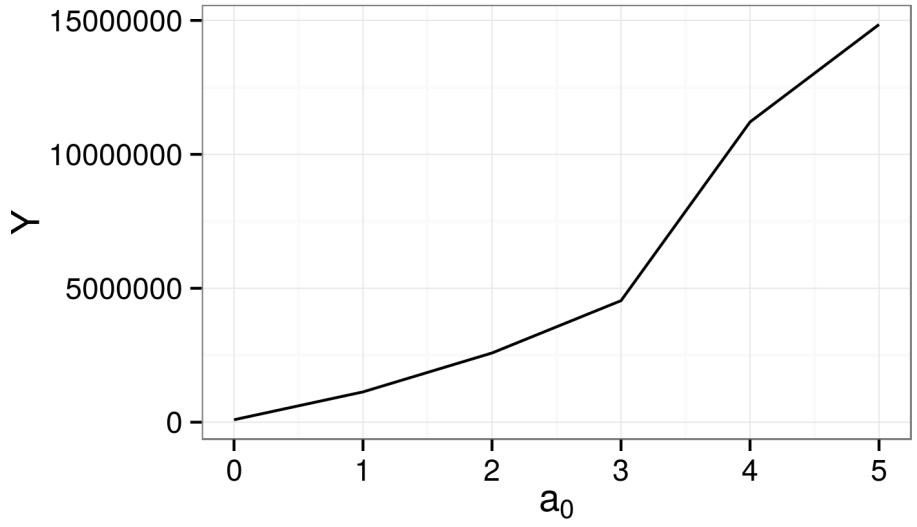


Figura 37 – Y final da cidade com relação ao valor de a_0
fonte:produção do próprio autor

A relação entre a produção socioeconômica Y e a área de interação a_0 segue, conforme esperado, uma relação exponencial (Figura 37). A área de interação cresce exponencialmente, uma vez que o parâmetro que a define no modelo é ao raio do círculo de interação. Dessa forma, é de se esperar que a produção socioeconômica cresce da

mesma maneira. Essa variável pode estar relacionada ao sistema de transporte urbano e às tecnologias de informação. O padrão de influência de a_0 em Y indica a importância desses dois fatores na produtividade de uma cidade. Ainda que intuitivo, é possível levantar a importância de se inverstir em facilitar a mobilidade e a comunicação para fomentar as trocas e a produção socioeconômica de um centro urbano.

O custo de transporte ϵ também segue um padrão esperado. Quanto maior ϵ, menor é a produção socioeconômica total Y da cidade(Figura 38). A relação aqui, porém, parece seguir um padrão de decaimento linear. Esse padrão observado é mais uma indicação da importância de produzir sistemas de transporte e comunicação eficientes e inclusivos. As simulações indicam que a cidade, como um todo, ganha sob essas condições.

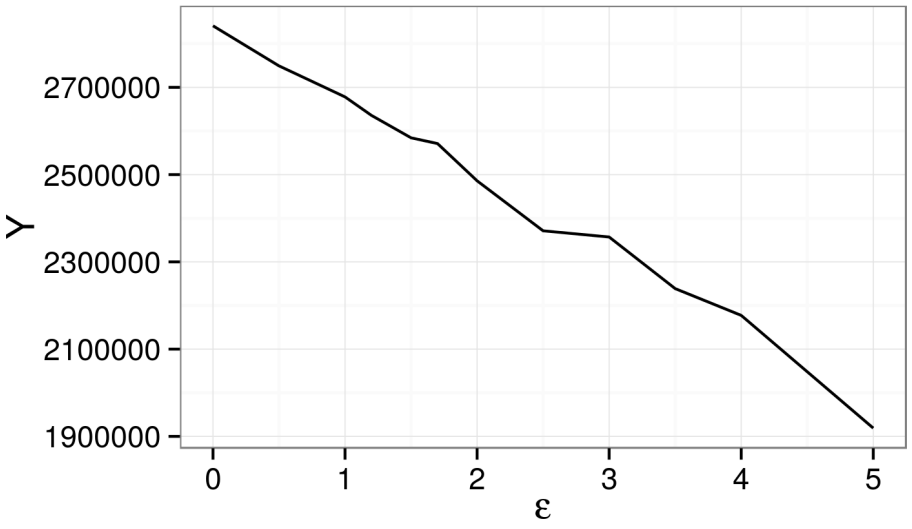


Figura 38 – Y final da cidade com relação ao valor de ϵ
fonte:produção do próprio autor

A produtividade média por encontro guarda uma relação linear e direta com Y(Figura 39). Quanto maior $\bar{g_{ag}}$, maior Y. Isso pode nos indicar a necessidade de produzir interações fortes e produtivas. É possível atingir esse tipo de interação através do fomento à trocas construtivas e do aumento confiança entre indivíduos de uma cidade. Uma cidade muito perigosa, por exemplo, seria uma cidade de baixo $\bar{g_{ag}}$. De acordo com as simulações, isso levaria à uma cidade menos produtiva.

A área de interação a_0, o custo de transporte ϵ e a produtividade média por encontro $\bar{g_{ag}}$ se comportam de maneira esperada. Algumas variáveis, porém, não apresentam padrões tão previsíveis de impacto em Y. O desvio padrão da produtividade média por encontro

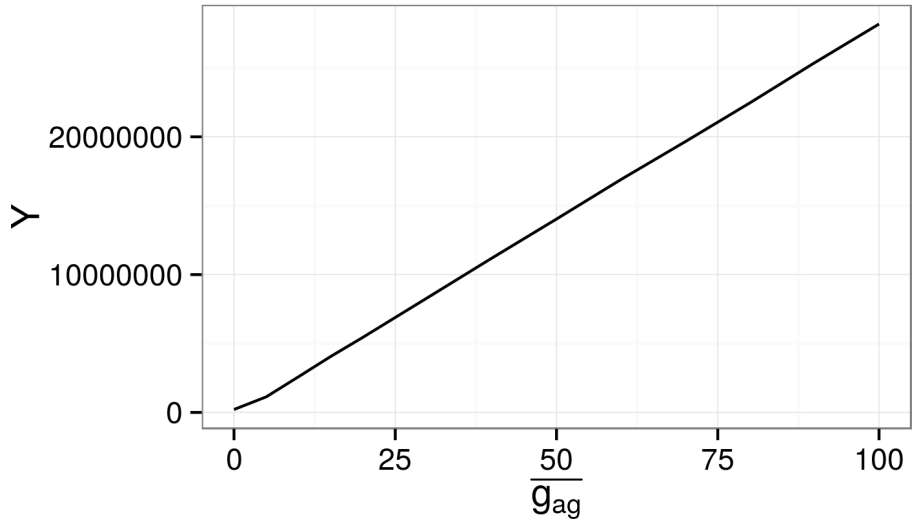


Figura 39 – Y final da cidade com relação ao valor de $\bar{g_{ag}}$
fonte:produção do próprio autor

σg_{ag}, por exemplo, impacta positivamente a produção socioeconômica total da cidade Y(Figura 40). A relação parece ser exponencial, indicando que quanto mais desigual for a taxa de produtividade da população no modelo, maior é a produção final total Y da cidade. A distribuição de g_{ag} simulada é simétrica. Ou seja, para cada grande produtor, se tem um agente com baixa produtividade. Esse padrão de crescimento indica a importância dos grandes empreendedores na produtividade total de uma cidade. Os grandes produtores parecem compensar os agentes com baixa taxa de produtividade, aumentando a produtividade total da cidade.

A produção socioeconômica total final apresenta um crescimento assintótico com relação ao acumulado inicial de produtividade socioeconômica $\bar{y_{ag}}$(Figura 41). O crescimento de Y com relação à $\bar{y_{ag}}$ no início do gráfico é intuitivo: dado que a cidade produz uma quantidade suficiente de encontros no tempo, quanto maior o estoque inicial, maior será o estoque final. Entretanto, a partir de um valor (aproximadamente 30 para a parametrização estudada), o limite temporal adotado se torna curto para a convergência das trajetórias individuais. Uma cidade insuficientemente densa, com um $\bar{y_{ag}} = 80$, por exemplo, não desaparece em 100 passos de tempo. Isso acontece porque pouquíssimas interações são necessárias para garantir um somatório de y igual ao total custo de transporte ao longo de todos os passos de tempo.

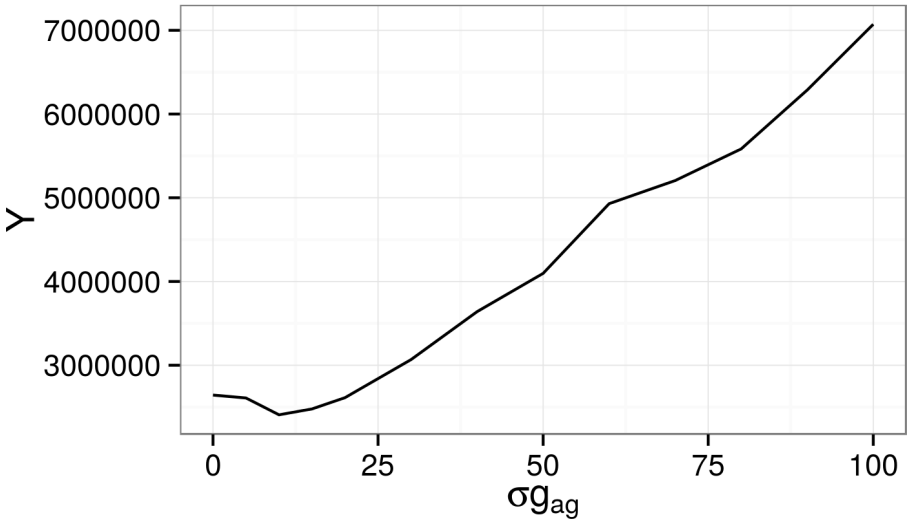


Figura 40 – Y final da cidade com relação ao valor de σg_{ag}
fonte:produção do próprio autor

O padrão de crescimento assiontótico indica, também, a importância das derivadas de produtividade em detrimento do estoque. A partir de algum monento, grandes estoques inicais já nao levam a grandes estoques finais dado que a produtividade dos agentes não se modifica.

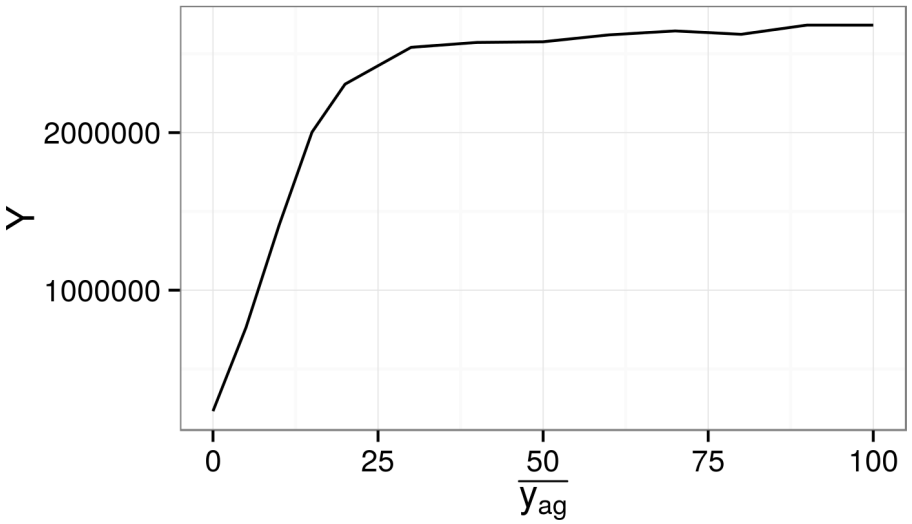


Figura 41 – Y final da cidade com relação ao valor de $\bar{y_{ag}}$
fonte:produção do próprio autor

O papel do desvio padrão do estoque inicial de produtividade socioeconômica σy_{ag} em Y também se mostra esclarecedor na compreensão dos fundamentos das leis de escala

em cidades. Os resultados do modelo indicam que, quando maior for σy_{ag}, menor será Y(Figura 42). Ou seja, quanto mais desigual for a distribuição do estoque de produtividade numa cidade, menos produtiva ela é no fim da simulação. Isso pode ser causado pelo fato de, em cidade muito desiguais, alguns habitantes não conseguirem se manter ao longo do tempo. Isso faz com que o total de habitantes se reduza e, por conseguinte, as taxas de encontros são reduzidas para a população toda. Isso resulta em somatórios finais de produção socioeconômicas inferiores para a cidade como um todo. É possível inferir que cidades brasileiras sejam penalizadas por esse problema. Com muita desigualdade, a população de baixa renda dessas cidades se vê obrigada a morar longe, o que reduz o tempo gasto em lazer e amenidades, fomentando os encontros e a microeconomia local.

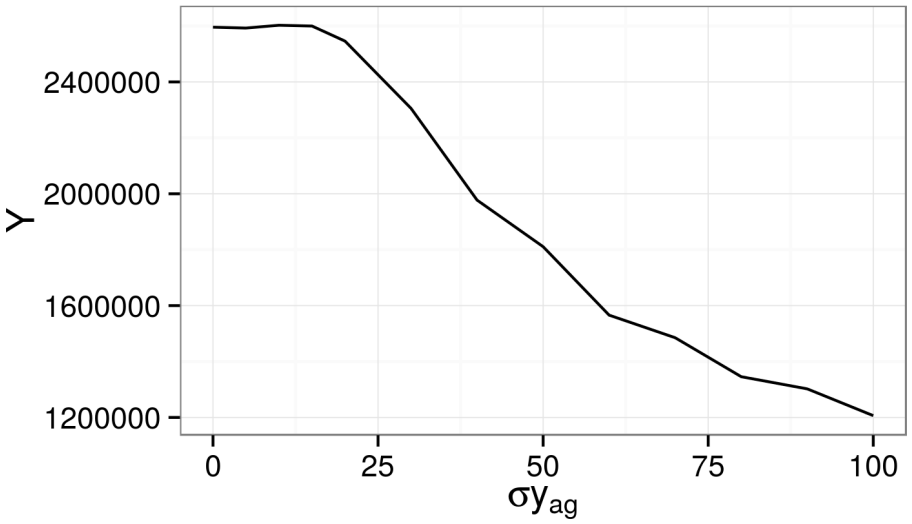


Figura 42 – Y final da cidade com relação ao valor de σy_{ag}
fonte:produção do próprio autor

Em seguida, analisou-se a papel dos parâmetros no escalamento da produção socioeconômica de um sistema urbano. Para cada variação sistemática dos parâmetros, simulou-se cada população dentro de um vetor de populações $N = n_1, n_2, ..., n_n$ habitando cidades de cada área de um vetor de áreas $A = a_1, a_2, ..., a_n$. O mesmo procedimento de excluir as cidades que não sobrevivem à parametrização foi feito para cada par de população N e área A e calculou-se o expoente de escalamento das cidades resultantes.

Divergindo da exploração dos parâmetros, esses testes do comportamento do sistema urbano não puderam ser feitos repetidamente, visto o custo de processamento e o tempo

necessário. Assim, apenas uma rodada foi realizada para cada valor testado. Os resultados são apresentados a seguir.

A relação entre o expoente de escalamento a área de interação a_0 parece seguir uma relação de crescimento exponencial até $a_0 = 4$, onde o crescimento para (Figura 43). Entre $a_0 = 3$ e $a_0 = 4$ é possível observar um crescimento acentuado. Isso pode ser explicado pelo comprimento dos quarteirões das cidades, que é de 3. Assim, quando se atinge um valor de $a_0 = 4$, os agentes passam a interagir com agentes de outras ruas.

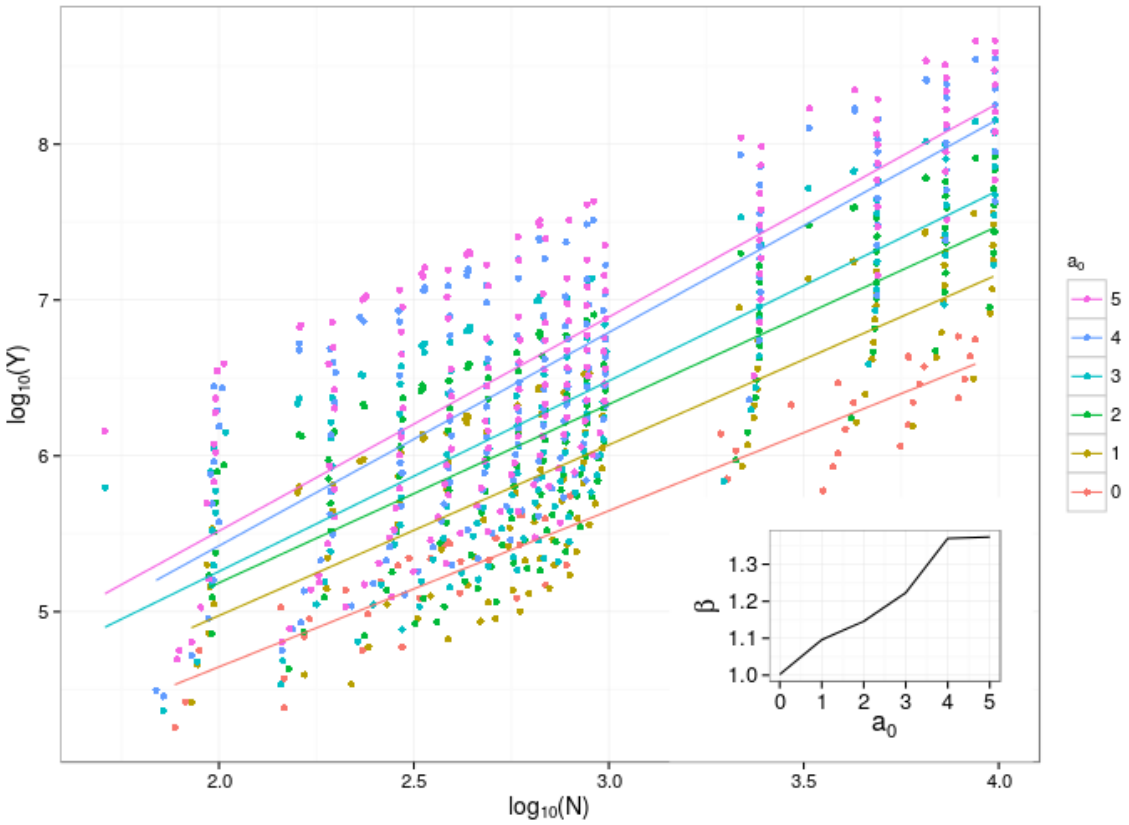


Figura 43 – β final do sistema urbano com relação ao valor de a_0
fonte:produção do próprio autor

Esse padrão de crescimento parece indicar que as vantagens de sistemas de transporte e comunicação mais eficientes são mais bem exploradas em cidades maiores. A medida que a área de interação a_0 cresce para o sistema urbano inteiro, a produtividade socioeconômica total Y cresce mais rapidamente em cidades maiores. É possível compreender isso de maneira intuitiva. Com mais pessoas interagindo, áreas de interações maiores serão exponencialmente mais utilizadas em cidades maiores. Trata-se de um efeito de densidade.

O custo de transporte ϵ parece se relacionar com o expoente de escalamento através de uma função de decaimento exponencial. Quanto maior o custo de transporte, menor é o expoente de escalamento da produção socioeconômica com a população do sistema urbano. A provável explicação para isso reside no fato de que, para cidades maiores e mais densas, muito mais pessoas sofrem a influência de um alto custo de transporte. Como resultado, cidades maiores são mais penalizadas. Dessa forma, o sistema urbano caminha rapidamente para um padrão de escalamento linear.

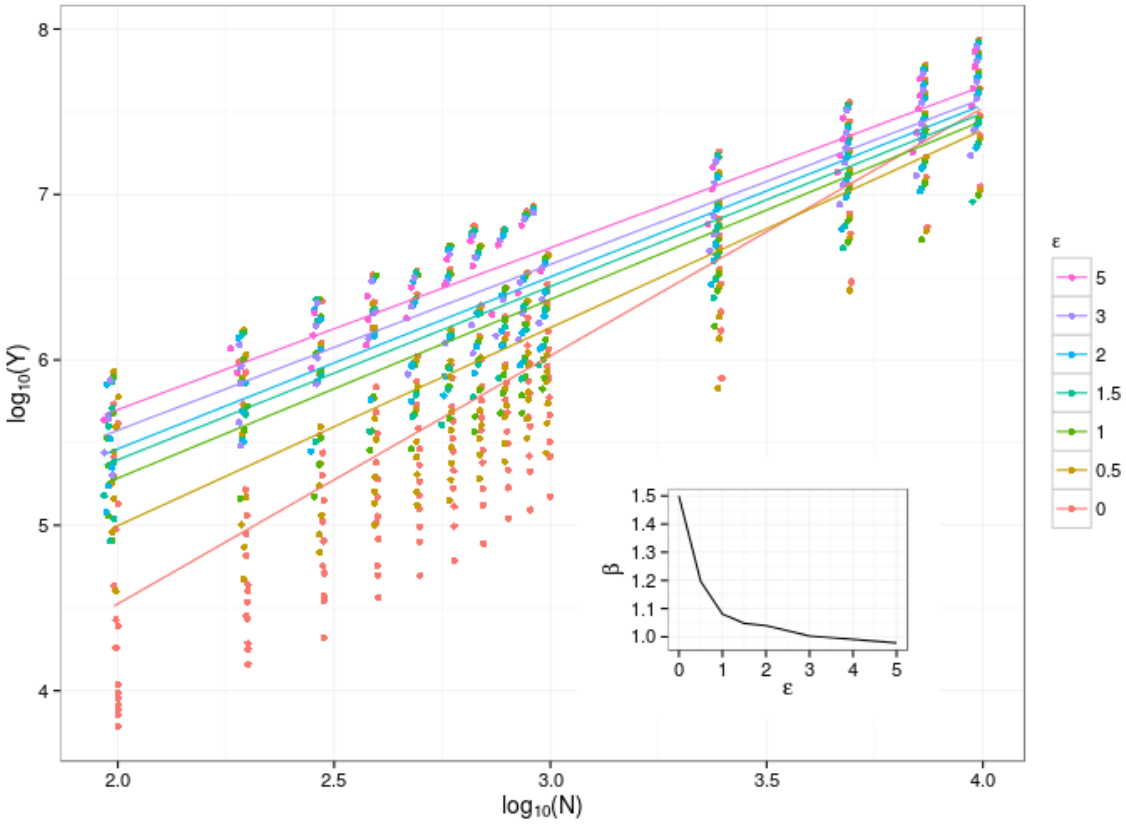


Figura 44 – β final do sistema urbano com relação ao valor de ϵ
fonte:produção do próprio autor

Esse padrão de influência do custo de transporte no expoente de escalamento de Y com relação à N pode, talvez, explicar os padrões de escalamento lineares encontrados em países de alto custo de vida, como o Reino Unido (ARCAUTE et al., 2015). Essa havia sido apontada como uma falha na hipótese de escalamento. A explicação que vinha sendo aceita era a de que a rede urbana do Reino Unido é densa demais, fazendo com que o país inteiro correspondesse a uma única cidade(ARCAUTE et al., 2015). A explicação do custo de transporte gerando um escalamento linear entre Y e N parece mais aderente à realidade.

A produtividade média por encontro $\bar{g_{ag}}$ parece guardar uma relação de crescimento assintótico com o expoente de escalamento(Figura 45). Quanto maior o valor de $\bar{g_{ag}}$, maior o valor final de β. Esse efeito parece ser mais acentuado para mudanças entre valores baixos de $\bar{g_{ag}}$. Altos valores de $\bar{g_{ag}}$ parecem convergir para algum β máximo. É dificil de confirmar o crescimento assintótico pela falta de certeza que uma rodada única produz. as variações observadas podem derivar de flutuações aletaórias das simulações. A maneira de tratar isso seria rodando diversas vezes as simulações e mediando os resultados.

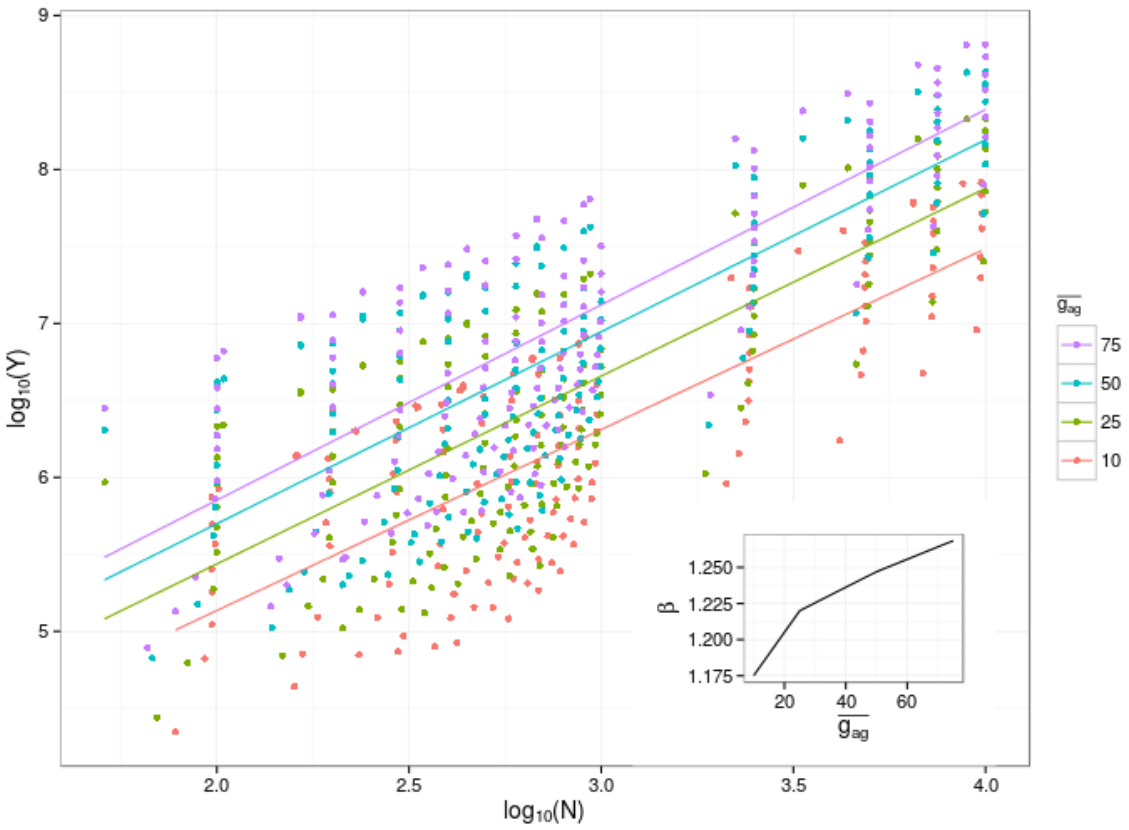


Figura 45 – β final do sistema urbano com relação ao valor de $\bar{g_{ag}}$
fonte:produção do próprio autor

Esse padrão de comportamento do expoente de escalamento da produtividade socioeconômica Y com relação à taxa de produtividade $\bar{g_{ag}}$ parece ser mais um fruto das dinâmicas de aglomerações. Taxas de produtividade maior influenciam mais cidades maiores, porque essas produzem mais encontros do que cidades menores. Assim, para altos valores de $\bar{g_{ag}}$, as cidades maiores se afastam mais das cidades menores.

O desvio padrão da produtividade média por encontro σg_{ag} parece produzir o efeito oposto. Quanto maior seu valor, menor é o expoente de escalamento da produção socioeconômica. Isso indica que, quanto mais desigual for a taxa de produtividade dos agentes no modelo, menor é o expoente de escalamento entre a população e a produção socioeconômica do sistema urbano. De maneira análoga à $\bar{g_{ag}}$, esse efeito parece ser mais acentuado para valores baixos de σg_{ag}.

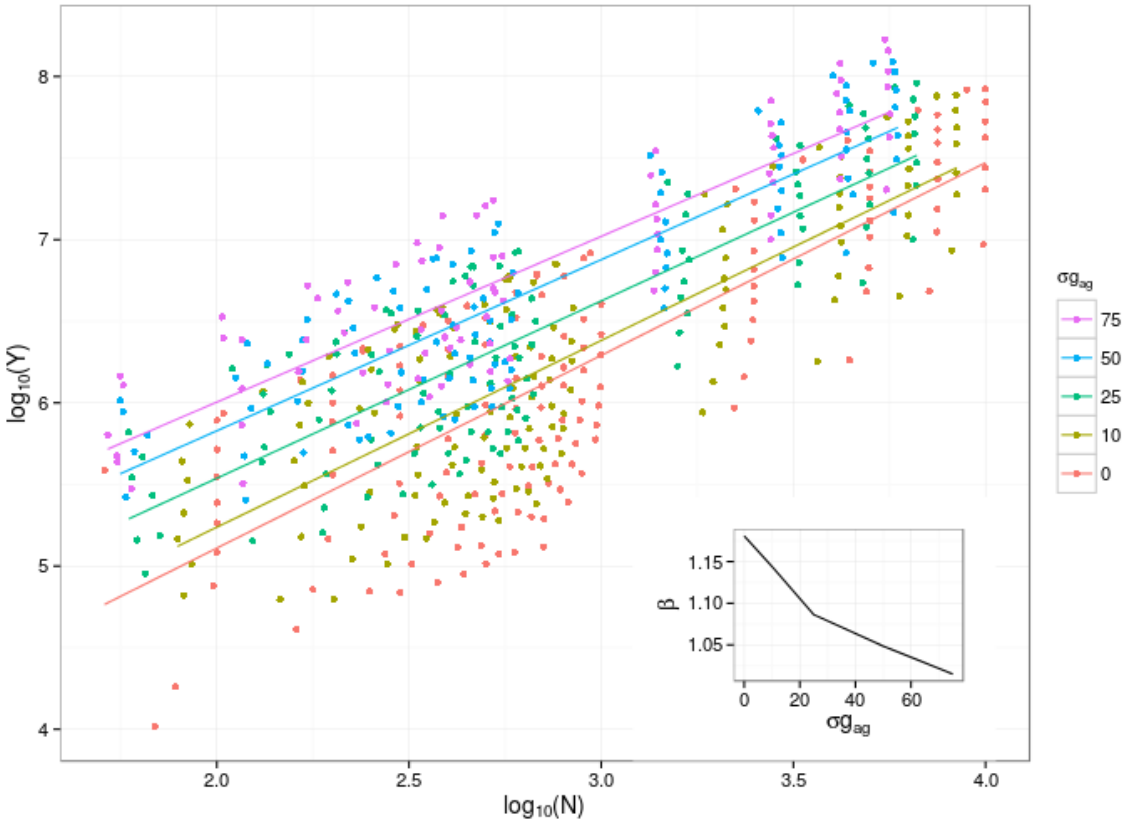


Figura 46 – β final do sistema urbano com relação ao valor de σg_{ag}
fonte:produção do próprio autor

Aqui, os efeitos de densidade parecem operar no sentido oposto ao observado para $\bar{g_{ag}}$. O decaimento do expoente de escalamento β indica que o paper dos grandes empreendedores, agentes com alto g_{ag}, observado para uma cidade sozinha(Figura 40), é mais forte nas cidades menores do que nas cidades maiores. Assim, à meidade que uma sociedade se torna mais desigual, as cidades menores parecem se aproximar das cidades maiores em termos de produção socioeconômica total.

O expoente de escalamento apresenta um crescimento assintótico com relação ao estoque médio inicail dos indivíduos $\bar{y_{ag}}$(Figura 47). O valor de β varia muito para valores pequenos de $\bar{y_{ag}}$ e, depois de algum ponto, parece estabilizar. Mais uma vez, se faz importante frizar a falta de certeza desses padrões específicos visto a falta de observações recorrentes. Essa variável parece ser muito determinante para o padrão final de escalamento do sistema urbano, produzindo uma variação total em β de 0.25.

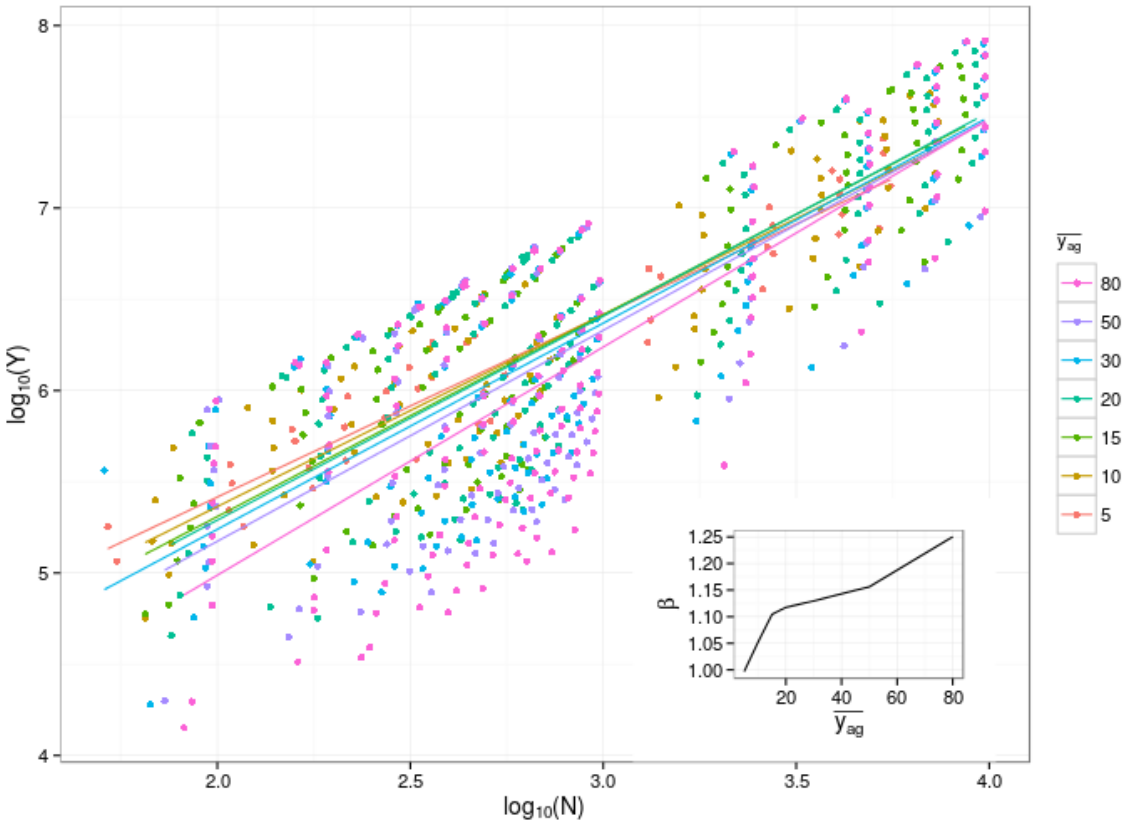


Figura 47 – β final do sistema urbano com relação ao valor de $\bar{y_{ag}}$
fonte:produção do próprio autor

Mais uma vez, é possível notar o papel do efeito de densidade. O gráfico indica que o Y final das cidades maiores varia muito pouco com relação à $\bar{y_{ag}}$. Isso deve se dar por conta dessas cidades produzirem tantas interações, que o estoque inicial se torna quase irrelevante para a sobrevivencia das cidades. Já as cidades menores parecem ser consideravelmente influenciadas pela variável. Para baixos valores de $\bar{y_{ag}}$, apenas sobrevivem as cidades com parametrizações muito produtivas, fazendo com que apenas cidades muito produtivas apareçam nesse sistema urbano.

O desvio padrão da renda inicial média dos indivíduos σy_{ag} parece se relacionar de maneira negativa com o expoente de escalamento (Figura 48). Para valores de σy_{ag} pequenos (entre 0 e 30) não se observa nenhuma relação sistemática. Na faixa dos valores maiores, porém, nota-se que quanto maior o valor da variável, menor é o expoente de escalamento da produção socioeconômica. Isso indica que, quanto mais desigual for a população no modelo, menor é o expoente de escalamento entre a população e a produção socioeconômica do sistema urbano.

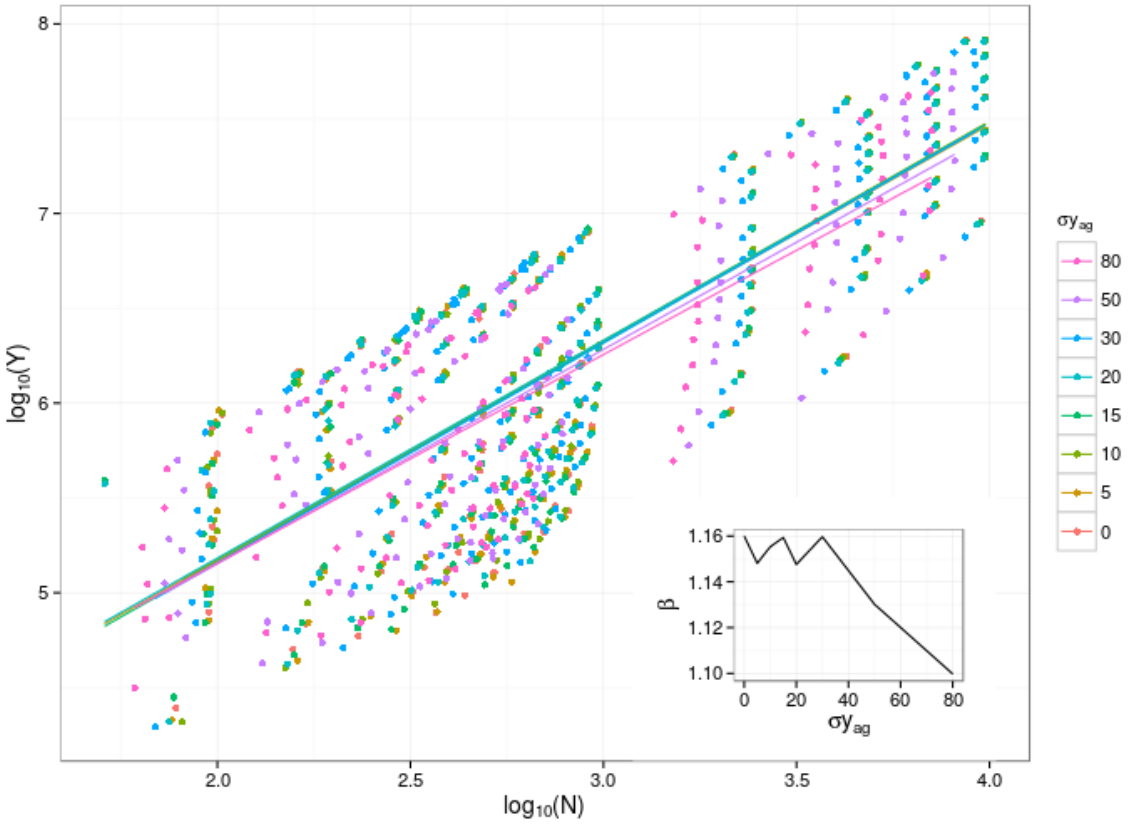


Figura 48 – β final do sistema urbano com relação ao valor de σy_{ag}
fonte: produção do próprio autor

De maneira oposta à $\bar{y_{ag}}$, as cidades maiores parecem ser as responsáveis pelo padrão observado. Quando maior σy_{ag}, menor é o Y final da cidade. Isso parece atuar, principalmente, sobre as cidades maiores. É provável que essas cidades maiores sintam mais a perda da parcela da população que é criada com baixo y_{ag} inicial. Por serem maiores, essa parecela também é maior. Logo, a impacto global da morte desses agentes recai mais pesadamente sobre as cidades maiores.

6 Conclusões

Neste capítulo, apresenta-se as conclusões do estudo.

Nesta dissetação, estudou-se as leis de alometria urbana sob duas perspectivas distintas e complementares: a da análise empírica e da simulação por agentes.

De uma maneira geral, os dados das cidades brasileiras parecem reiterar a hipótese capitaneada por Bettencourt (BETTENCOURT, 2013), com variáveis socioeconômicas escalando superlinearmente e variáveis infraestruturais escalando sublinearmente com relação à população das cidades.

Quanto à simulação, a metodologia de modelagem por agentes (ABM) se mostrou relevante para a exploração dos parâmetros do modelo que explica a origem das leis de escala. Se valendo do ABM, foi possível esclarecer laços causais entre as variáveis e a produção socioeconômica das cidades.

6.1 Conclusões sobre as análises de dados das cidades brasileiras

As regressões entre variáveis urbanas no Brasil, quando feitas a partir dos dados brutos, indicaram os padrões de escalamento superlinear para variáveis socioeconômicas e sublinear para variáveis infraestruturais. Porém, os expoentes de escalamento fogem dos valores esperados teoricamente e o fit do modelo não se comporta bem para cidades grandes. É provável que isso aconteça por conta da definição do IBGE de município, que pode não traduzir bem as unidades espaciais funcionais que seriam as cidades naturais. É provável que muitas cidades carreguem valores de PIB, ou área que não derivem exclusivamente de dinâmicas sociais locais.

Divergindo de estudos em outros países, a sensibilidade do expoente de escalamento do sistema urbano brasileiro parece ser baixa, adotando valores bem definidos num espectro abrangente de valores limites de corte de densidade e população. A densidade populacional pareceu ser um parâmetro de corte mais consistente do que a população.

Encontrou-se, ainda, que os bancos de dados de cidades e de regiões metropolitanas apresentam alguns padrões ligeiramente distintos de sensibilidade. De uma maneira geral, os expoentes encontrados a partir do banco de dados de cidades se aproximam mais dos observados em outros sistemas urbanos do mundo. Principalmente os expoentes de área e quilômetros de ruas no banco de dados de Regiões Metropolitanas, que estabilizam em

valores próximos do escalamento linear. Isso pode ser explicado pela heterogeneidade na definição de Região Metropolitana adotada pelos estados do país.

No banco de dados das cidades brasileiras, foi possível observar certa robustez dos padrões gerais de escalamento. O PIB varia numa faixa de inclinação sempre acima de 1 e as variáveis de infraestrutura (área e km de ruas) variam dentro de uma faixa de inclinação sempre menor do que 1. Mais do que isso, os valores do expoente de escalamento parecem convergir e se estabilizar em valores muito próximos dos esperados teoricamente.

Após um corte no banco de dados no menor valor de densidade para o qual os valores convergem (12 habitantes por hectare), encontrou-se valores de escalamento do PIB de 1.14, área 0.88, quilômetros de rua 0.96. Isso indica que o sistema urbano brasileiro parece ser um pouco menos eficiente do que a média mundial.

O expoente de escalamento foi calculado para outras variáveis.

Tirando a raça Branca, todas as outras raças escalam superlinearmente, indicando que as cidades parecem ser compostas de populações mais heterogêneas à medida que crescem. As variáveis de Entorno e de Acesso a serviços Básicos também indicam melhor qualidade de vida nas cidades maiores. O total de moradores conectados à rede de esgoto, abastecidos de água a partir de rede e com o lixo coletado escalam de maneira superlinear com a população. O total de moradores com acesso à energia elétrica, a medidores de energia elétrica e a banheiros exclusivos pareceu escalar de maneira linear com a população. E o total de moradores que tinham seu lixo queimado ou jogados em terreno baldio escalou de maneira sublinear com a população (ainda que essas duas últimas variáveis apresentaram baixo valor no r-quadrado).

Com relação às vítimas, Suicídios e Acidentes de Trânsito parecem diminuir ao passo que Homicídios parecem aumentar, confirmando a hipótese das interações sociais como preditores dos escalamentos.

Ao mesmo tempo, as despesas de uma cidade parecem escalar sublinearmente com relação à população das cidades, com a exceção de despesas por transferência. Já a receita parece escalar de maneira superlinear com a população, com impostos e taxas crescendo de maneira não proporcional à população.

Como continuação desses estudos, seria interessante estudar os padrões de escalamento de variáveis que indiquem desigualdades sociais e acessibilidade à amenidades com a população das cidades urbanas. Além disso, faz-se importante testar as distribuições

de leis de potência estatisticamente usando algum método estatístico robusto para tal (GOMEZLEVIANO; YOUN; BETTENCOURT, 2012).

6.2 Conclusões sobre o modelo de origem das leis de escala

O modelo baseado em agentes conseguiu incorporar as dinâmicas sociais básicas propostas por Bettencourt (BETTENCOURT, 2013). Com o custo de transporte variando como uma função da densidade populacional, o modelo foi capaz de produzir padrões de escalamento muito próximos daqueles observados na realidade.

Os expoentes de escalamento encontrados variam de acordo com a parametrização adotada, mas se mantém dentro dos padrões esperados. Para algumas combinações de parâmetros, é possível encontrar os valores teóricos de escalamento indicados pela estrutura de campo médio proposto por Bettencourt (GOMEZLEVIANO; YOUN; BETTENCOURT, 2012). A parametrização que mais se aproxima dos valores propostos por Bettencourt (BETTENCOURT, 2013) gera um expoente de escalamento das produção socioeconômica $Y = 1.168$ e da área $A = 0.8785$.

A relação entre a produção socioeconômica Y e a área de interação a_{zero} segue, conforme esperado, uma relação exponencial. O custo de transporte ϵ também segue um padrão esperado. Quanto maior o custo de transporte, menor é a produção socioeconômica total Y da cidade. A relação aqui, porém, parece seguir um padrão de decaimento linear.

A produtividade média por encontro guarda uma relação linear e direta com Y. Quanto maior $g_{ag}mean$, maior Y. O desvio padrão da produtividade média por encontro impacta positivamente a produção socioeconômica total da cidade Y. A relação parece ser exponencial. Isso indica que, quanto mais desigual for a taxa de produtividade da população no modelo, maior é a produção final total Y da cidade.

A produção socioeconômica total final apresenta um crescimento assintótico com relação à renda inicial média $y_{ag}mean$. A partir de um valor (aproximadamente 30 para a parametrização estudada), o limite temporal adotado se torna curto para a convergência das trajetórias individuais. Uma cidade insuficientemente densa, com um $y_{ag}mean$ de 80, por exemplo, não desaparece em 100 passos de tempo. Isso acontece porque pouquíssimas interações são necessárias para garantir um somatório de y igual ao total custo de transporte ao longo de todos os passos de tempo. Os resultados do modelo indicam que, quando maior for o desvio padrão da renda inicial média $y_{ag}stdv$, menor será Y. Em cidade muito

desiguais, alguns habitantes não conseguem se manter ao longo do tempo, por mais densa que seja a cidade. Isso faz com que o total de habitantes se reduza e, por conseguinte, as taxas de encontros são reduzidas para a população toda. Isso resulta em somatórios finais de produção socioeconômicas inferiores para a cidade como um todo.

O impacto das variáveis sobre o expoente de escalamento do sistema urbano também foi estudado.

A relação entre o expoente de escalamento a área de interação a_{zero} parece seguir uma relação de crescimneto exponencial até o limite do tamanho de comprimento de quarteirão adotado no modelo, depois do qual a produção socioeconômica Y para de crescer. O custo de transporte ϵ parece se relacionar com o expoente de escalamento através de uma função de decaimento exponencial. Quanto maior o custo de transporte, menor é o expoente de escalamento da produção socioeconômica com a população do sistema urbano.

A produtividade média por encontro $g_{ag}mean$ guarda uma relação linear de crescimento assintótico com o expoente de escalamento. O desvio padrão da padrão da produtividade média por encontro $g_{ag}stdv$ parece produzir o efeito oposto. Quanto maior seu valor, menor é o expoente de escalamento da produção socioeconômica.

O expoente de escalamento apresenta um crescimento assintótico com relação à renda inicial média $y_{ag}mean$. O desvio padrão da renda inicial média dos indivíduos $y_{ag}stdv$ parece se relacionar de maneira negativa com o expoente de escalamento

Como futuros desenvolvimentos dessa linha de pesquisa, seria interessante testar os padrões de escalamento em topologias de rede complexa para o grid de ruas, e não apenas em grids de estilo Manhattan. Outro caminho de pesquisa a ser seguido seria o de compreender o papel de distintas funções de distribuição de renda e de taxa de produção na população.

Referências[1]

ALVES, L. et al. Distance to the Scaling Law: A Useful Approach for Unveiling Relationships between Crime and Urban Metrics. *PLoS ONE*, v. 8(12), p. –, aug 2013. Citado na página 32.

ALVES, L. et al. Empirical analysis on the connection between power-law distributions and allometries for urban indicators. *Physica A*, v. 409, p. 175–182, set 2014. Citado 2 vezes nas páginas 47 e 51.

ANTONIO, F. et al. Growth patterns and scaling laws governing AIDS epidemic in Brazilian cities. *Physics and Society*, v. 9(10), p. –, oct 2014. Citado na página 32.

ARCAUTE, E. et al. Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface*, v. 12(102), p. 543–567, jan 2015. Citado 5 vezes nas páginas 15, 30, 42, 49 e 75.

BAK, P. *How Nature Works: the science of self-organized criticality*. New York, NY, USA: Springer, 1996. Citado 3 vezes nas páginas 19, 20 e 24.

BAR-YAM, Y. Concept: Power Law. *New England Complex Systems Institute Concept Map*, acessado em junho de 2014. Disponível em: <http://www.necsi.edu/guide/concepts-/powerlaw.html>. Citado 2 vezes nas páginas 19 e 21.

BARABASI, A.-L.; ALBERT, R. Emergence of Scaling in Random Networks. *Science*, v. 286 (5439, p. 509–512, oct 1999. Citado na página 24.

BATTY, M. *The New Science of Cities*. Massachusetts, USA: MIT press, 2013. Citado 6 vezes nas páginas 6, 7, 14, 28, 30 e 32.

BELEVOLO, L. *A Historia da Cidade*. Sao Paulo, BRA: Perspectiva, 1997. Citado na página 27.

BETTENCOURT, L. The Origins of Scaling in Cities. *Science*, v. 340, p. 1438–1441, jul 2013. Citado 22 vezes nas páginas 6, 7, 8, 14, 15, 26, 29, 30, 31, 32, 33, 34, 35, 36, 39, 41, 51, 59, 63, 67, 80 e 82.

BETTENCOURT, L. *Cidades como sistemas complexos. In: Furtado, B. et all (Org.). Modelagem de Sistemas Complexos Para Politicas Publicas*. Brasilia, BR: Ipea, 2015. Citado na página 28.

BETTENCOURT, L. et al. Growth, innovation, scaling, and the pace of life in cities. *PNAS*, v. 104(17), p. 7301–7306, apr 2007. Citado 3 vezes nas páginas 14, 31 e 49.

BETTENCOURT, L.; WEST, G. A unified theory of urban living. *Nature*, v. 467, p. 912–913, oct 2010. Citado na página 29.

BRAND, S. *Whole Earth Discipline: Why Dense Cities, Nuclear Power, Transgenic Crops, RestoredWildlands, and Geoengineering Are Necessary*. New York, USA: Atlantic Books, 2010. Citado na página 30.

[1] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

BURRY, K. *Statistical Methods in Applied Science*. New York, USA: John Wiley, 1975. Citado na página 18.

CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. Powerlaw distributions in empirical data. *SIAM*, v. 51 (4), p. 661–703, nov 2009. Citado 5 vezes nas páginas 19, 24, 25, 47 e 51.

COTTINEAU, C. et al. Paradoxical Interpretations of Urban Scaling Laws. *Arxiv*, v. 1507.07878, p. –, jul 2015. Citado 4 vezes nas páginas 15, 30, 42 e 49.

FRAGKIAS, M. et al. Does size matter? Scaling of CO2 emissions and U.S. urban areas. *PLoS ONE*, v. 8, p. 6, jun 2013. Citado na página 30.

GALILEI, G. *Two New Sciences*. traduzido por Henry Crew e Alfonso de Salvio (Dover Publications, New York, NY, 1954): Originalmente publicado por Elzevir, 1638. Citado na página 17.

GIRARDET, H. *Giant Footprints*. Stockholm, Suécia: Our Planet, 1996. Citado na página 14.

GLAESER, E. *Triumph of the City*. New York, USA: Penguin Books, 2011. Citado na página 29.

GOMEZLEVIANO, A.; YOUN, H.; BETTENCOURT, L. The statistics of urban scaling and their connection to Zipf's law. *Plos ONE*, v. 7 (7), p. –, jul 2012. Citado 3 vezes nas páginas 24, 25 e 82.

GRIMMA, V. et al. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, v. 198 (1-2), p. 115–126, jun 2006. Citado na página 59.

IBGE. *Arranjos Populacionais e Concentraçoes Urbanas do Brasil*. Brasilia, BR: IBGE, 2015. ISBN 9788524043390. Citado na página 41.

JACOBS, J. *The Life and Death of the Great American City*. New York, USA: [s.n.], 1961. Citado 2 vezes nas páginas 15 e 28.

JIANG, B.; MIAO, Y. The Evolution of Natural Cities from the Perspective of Location-Based Social Media. *The Professional Geographer*, v. 67(2), p. 295–306, nov 2014. Citado na página 42.

KRUGMAN, P.; FUJITA, M.; VENABLES, A. *The Spatial Economy*. New York, USA: Perspectiva, 1999. Citado na página 28.

LAM, H. et al. Corrections to the Central Limit Theorem for Heavy-tailed Probability Densities. *Journal of Theoretical Probability*, v. 24, p. 895–927, jul 2011. Citado na página 24.

LOUF, R.; BARTHELEMY, M. How congestion shapes cities: from mobility patterns to scaling. *Scientific Reports*, v. 4, p. 5561, jul 2014. Citado na página 30.

LOUF, R.; BERTHELEMY, M. Scaling: lost in the smog. *Environment and Planning B*, v. 41, p. 767–769, jul 2014. Citado 3 vezes nas páginas 30, 35 e 42.

MANAGEMENT, O. of; BUDGET. *2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas*. Washington, USA: [s.n.], 2010. Citado na página 37.

MANDELBROT, B. *The Fractal Geometry of Nature*. New York, USA: W. H. Freeman and Company, 1982. Citado na página 21.

MARSHALL, A. *Principles of Economy*. New York, USA: Macmillian, 1890. Citado na página 28.

MELO, H. P. et al. Statistical Signs of Social Influence on Suicides. *Scientific Reports*, v. 1402.2510, p. –, feb 2014. Citado na página 32.

MITCHELL, M. *Complexity: A Guided Tour*. Oxford, GB: Oxford University press, 2011. Citado na página 19.

NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, v. 46, p. 323–351, nov 2005. Citado 5 vezes nas páginas 17, 18, 21, 23 e 24.

NICOLIS, G.; NICOLIS, C. *Foundations of Complex Systems*. Singapura: World Scientific Publishing Company, 2007. Citado 4 vezes nas páginas 19, 21, 22 e 25.

OLIVEIRA, E. A.; ANDRADE, J. S.; MAKSE, H. A. Large cities are less green. *Scientific Reports*, v. 4, p. 4235, feb 2014. Citado na página 30.

ORTMAN, S. G. et al. The Pre-History of Urban Scaling. *PLOSone*, v. 9 (2), p. –, feb 2014. Citado na página 32.

PEARSON, K. The Problem of Random Walk. *Nature*, v. 72, p. 294–342, jun 1905. Citado na página 61.

PETERSON, M. A. Galileo's discovery of scaling laws. *American Journal of Physics*, v. 70, n. 6, p. 575–580, 2002. Citado na página 17.

PISARENKO, V.; RODKIN, M. *Heavy-Tailed Distributions in Disaster Analysis*. New York, NY, USA: Springer Science, 2010. Citado na página 18.

RICE, J. *Mathematical Statistics and Data Analysis*. New York, USA: Duxbury Press, 2001. Citado 2 vezes nas páginas 17 e 18.

SCHLAPFER, M. et al. The scaling of human interactions with city size. *Journal of the Royal Society Interface*, v. 11 (98), p. –, jul 2014. Citado 3 vezes nas páginas 8, 34 e 35.

SCHROEDER, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. New York, USA: Dover Books on Physics, 2009. Citado na página 24.

SOLE, R. *Phase Transitios*. New Jersey, USA: Princeton University Press, 2012. Citado na página 24.

THOMAS, K. *O Homem e o mundo natural*. São Paulo, SP, BRA: Compania das Letrass, 1996. Citado na página 27.

UN-HABITAT. *State of the World's Cities 2012/2013*. New York, USA: UN-Habitat Press, 2013. Citado 2 vezes nas páginas 14 e 28.

UN-HABITAT. *World Urbanization Prospects: The 2014 Revision*. New York, USA: UN-Habitat Press, 2014. Citado 3 vezes nas páginas 8, 14 e 27.

WEAVER, W. Science end Complexity. *American Scientist*, v. 36, p. 536–547, oct 1948. Citado na página 28.

WEST, G.; BROWN, J. *Scaling in Biology*. Oxford, UK: Oxford University Press, 2000. Citado 4 vezes nas páginas 8, 21, 25 e 26.

WEST, G. B.; BROWN, J. H. The origin of allometric scaling laws in biology from genomes to ecosystems. *Jounral of Experimental Biology*, v. 208, p. 1575–1592, aug 2005. Citado na página 26.

WIDOM, B. Scaling laws. acessado em junho de 2014. Disponível em: <www.scalinglaws-.com>. Citado na página 22.

Apêndice A – Código da análise dos dados em R

```
1
2  #carregas as bibliotecas necessarias
3    library(ggthemes)
4    library(plyr)
5    library(ggplot2)
6    library(reshape2)
7    library(gplots)
8    library(colorRamps)
9
10 #Funcao para computar ordem de magnitude
11   log10_ceiling <- function(x) {
12     10^(ceiling(log10(x)))
13   }
14
15 #agregar os dados das cidades dentro das regioes metropolitanas
16   #conectar com a pasta onde estao os dados e subir para o r como um data-frame
17   setwd("/home/joaomeirelles/Documents/USP/TESE/data/RM")
18   getwd()
19     #dados ibge
20     base <- read.csv(file="cidades_ibge.csv", head=TRUE, sep=",")
21     #nome das regioes metropolitanas
22     rm_nomes <- read.csv(file="rm_nomes.csv", head=TRUE, sep=",")
23     #nome das cidades que pertencem as regioes metropolitanas
24     rm_exclude <- read.csv(file="rm_exclude.csv", head=TRUE, sep=",")
25
26   #limpa os dados
27   base <- read.csv(file="basetst.csv", head=TRUE, sep=",")
28   base <- base[! duplicated(base$Codigo),]
29   base2 <- subset(base, km_ruas>0)
30   base <- subset(base, pop>0)
31   base <- subset(base, pib>0)
32
33   #agrega as cidadess em regioes metropolitanas
34   rm_nomes[,c("pop", "area", "pib", "km_ruas")] <- NA
35   basepars <- base[-c(2:3)]
36   temp = list.files(pattern="*.csv")
37   for (i in 1:length(temp)) {
38     rm <- assign(temp[i], read.csv(temp[i]))
39     basepars <- base[-c(2:3)]
40     rm_include <- ddply(basepars, .(basepars$Codigo %in% rm$Codigo), summarise, pop=sum(
         pop), area=sum(area), pib=sum(pib), km_ruas=sum(km_ruas))
41     rm_nomes[rm_nomes$Codigo==i,]$pop <- rm_include[2,2]
42     rm_nomes[rm_nomes$Codigo==i,]$pib <- rm_include[2,4]
43     rm_nomes[rm_nomes$Codigo==i,]$area <- rm_include[2,3]
44     rm_nomes[rm_nomes$Codigo==i,]$km_ruas <- rm_include[2,5]
45   }
46   base <- rbind(base, rm_nomes)
47
48   #salva novo csv com dados das regioes metropolitanas
49   baserm <- base[ ! base$Codigo %in% rm_exclude$Codigo, ]
50   write.csv(baserm, file = "baserm.csv")
51
52
53 #analise de escalamento
54   #importa dados
55   baserm <- read.csv(file="baserm.csv", head=TRUE, sep=",")
```

```
56  str(base)
57  summary(base)
58
59  #calcula o G das cidades
60  base$G <- (base$pib/base$pop)*(base$km_ruas/base$pop)
61
62  basepars <- base[-c(1:4)]
63  par (mfrow=c(1,1))
64  colnames <- dimnames(base)[[2]]
65  parsnames <- colnames[-c(1:4)]
66
67  #BOXPLOT
68  basescaled <- scale(basepars)
69  ggplot(data = melt(basescaled), aes(x=Var2, y=value)) + geom_boxplot() + xlab("") +
        ylab("std.dev") + theme_economist()
70  ggplot(data = melt(basescaled), aes(x=Var2, y=value)) + geom_boxplot() + xlab("") +
        ylab("std.dev") + theme_few()
71  ggplot(data = melt(basescaled), aes(x=Var2, y=value)) + geom_boxplot() + xlab("") +
        ylab("std.dev") + theme_igray()
72  ggplot(data = melt(basescaled), aes(x=Var2, y=value)) + geom_boxplot() + xlab("") +
        ylab("std.dev") + theme_classic()
73  ggplot(data = melt(basescaled), aes(x=Var2, y=value)) + geom_boxplot() + xlab("") +
        ylab("std.dev") + theme_bw(base_size = 14, base_family = "Helvetica")
74
75  #HISTOGRAMS
76  for (i in c(1:ncol(basepars))) {
77    hist(basepars[,i],
78         xlim=c(0, (max(basepars[,i])/((0.000001*log10_ceiling(max(basepars[,i])))))),
79         breaks=seq(0, (log10_ceiling(max(basepars[,i]))), 10000),
80         main=parsnames[i],
81         probability=TRUE,
82         xlab="",
83         col="gray10", border="white")
84  }
85
86  #SCALING
87  scaling <- data.frame(scale = character((ncol(basepars))),
88                        intercept= numeric((ncol(basepars))),
89                        alpha= numeric((ncol(basepars))),
90                        r.squared= numeric((ncol(basepars))),
91                        p.value= numeric((ncol(basepars))),
92                        stringsAsFactors=FALSE)
93
94  scale <- list()
95  options(scipen = 10)
96  for (i in c(1:ncol(basepars))) {
97    scale[[paste("pop",colnames(basepars[i]),sep="x")]] <- lm(log10(basepars[,i]) ~ log10
        (basepars$pop))
98
99    #save the variables on the data.frame
100   scaling[i, "scale"] <- paste("pop",colnames(basepars[i]),sep="x")
101   scaling[i, "r.squared"] <- summary(scale[[paste("pop",colnames(basepars[i]),sep="x")
        ]])$r.squared
102   scaling[i, "p.value"] <- summary(scale[[paste("pop",colnames(basepars[i]),sep="x")]])
        $coefficients[2,4]
103   scaling[i, "alpha"] <- summary(scale[[paste("pop",colnames(basepars[i]),sep="x")]])$
        coefficients[2,1]
104   scaling[i, "intercept"] <- summary(scale[[paste("pop",colnames(basepars[i]),sep="x")
        ]])$coefficients[1,1]
105 }
```

```
106
107  #PLOT SCALING
108  for (i in c(1:ncol(basepars))) {
109    print(ggplot(basepars, aes(log10(pop), log10(basepars[,i])), na.rm=TRUE)
110          + geom_point()
111          + ggtitle(parsnames[i])
112          + xlab(paste("Log","(",(parsnames[1]),")",sep=""))
113          + ylab(paste("Log","(",(parsnames[i]),")",sep=""))
114          + geom_smooth(method=lm,fullrange=T)
115          + geom_abline(intercept=scaling[i,"intercept"])
116    )
117  }
118
119  #corte por densidade
120  base$densidade <- (base$pop / base$area)
121  base$densidademi <- (base$densidade / 0.38610)
122  base$areaha <- (base$area * 100)
123  base$densidadeha <- (base$pop / base$areaha)
124  base$areaurbana <- (base$area - base$areaplantada)
125  base_corte <- subset(based, densidadeha>12)
126
127  #analise de robustez
128  baseparsp <- base[-c(1:3)]
129  scaling_rob <- data.frame(dens_min = numeric((ncol(baseparsp))),
130                            pop_min = numeric((ncol(baseparsp))),
131                            cases = numeric((ncol(baseparsp))),
132                            pop_area_intercept= numeric((ncol(baseparsp))),
133                            pop_area_alpha= numeric((ncol(baseparsp))),
134                            pop_area_r.squared= numeric((ncol(baseparsp))),
135                            pop_area_p.value= numeric((ncol(baseparsp))),
136                            pop_pib_intercept= numeric((ncol(baseparsp))),
137                            pop_pib_alpha= numeric((ncol(baseparsp))),
138                            pop_pib_r.squared= numeric((ncol(baseparsp))),
139                            pop_pib_p.value= numeric((ncol(baseparsp))),
140                            pop_ruas_intercept= numeric((ncol(baseparsp))),
141                            pop_ruas_alpha= numeric((ncol(baseparsp))),
142                            pop_ruas_r.squared= numeric((ncol(baseparsp))),
143                            pop_ruas_p.value= numeric((ncol(baseparsp))),
144                            stringsAsFactors=FALSE)
145  scalearea <- list()
146  scalepib <- list()
147  scaleruas <- list()
148
149  pop_lim <- c(0, 100, 1000, 5000, 7000, 10000, 25000, 50000, 75000, 100000)
150  dens_lim <- c(0,0.1,0.2,0.5,0.7,1,1.2,1.5,1.7,2,2.5,3,3.5,4,5,6,7,8,9,10,11,12)
151
152
153  rob_matrx_y <- matrix(NA, nrow = length(pop_lim), ncol = length(dens_lim))
154  rob_matrx_area <- matrix(NA, nrow = length(pop_lim), ncol = length(dens_lim))
155  rob_matrx_ruas <- matrix(NA, nrow = length(pop_lim), ncol = length(dens_lim))
156
157  rownames(rob_matrx_y) <- pop_lim
158  colnames(rob_matrx_y) <- dens_lim
159
160  rownames(rob_matrx_area) <- pop_lim
161  colnames(rob_matrx_area) <- dens_lim
162
163  rownames(rob_matrx_ruas) <- pop_lim
164  colnames(rob_matrx_ruas) <- dens_lim
165
```

```
166  c <- 0
167  r <- 0
168  f <- 0
169  for (j in c(0,10, 50, 100, 500, 1000, 5000, 10000, 50000, 100000, 500000)){
170    basepars <- subset(baseparsp, pop>j)
171    r <- r+1
172
173
174    for (i in c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,18,20,21)){
175      c <- c+1
176      f <- f+1
177
178      basepars <- subset(basepars, densidadeha>i)
179      scalearea <- lm(log10(basepars$area) ~ log10(basepars$pop))
180      scalepib <- lm(log10(basepars$pib) ~ log10(basepars$pop))
181      scaleruas <- lm(log10(basepars$km_ruas) ~ log10(basepars$pop))
182
183      #save the variables on the data.frame
184      scaling_rob[c, "pop_min"] <- j
185      scaling_rob[c, "dens_min"] <- i
186      scaling_rob[c, "cases"] <- nrow(basepars)
187      scaling_rob[c, "pop_area_intercept"] <- summary(scalearea)$coefficients[1,1]
188      scaling_rob[c, "pop_area_alpha"] <- summary(scalearea)$coefficients[2,1]
189      scaling_rob[c, "pop_area_r.squared"] <- summary(scalearea)$r.squared
190      scaling_rob[c, "pop_area_p.value"] <- summary(scalearea)$coefficients[2,4]
191      scaling_rob[c, "pop_pib_intercept"] <- summary(scalepib)$coefficients[1,1]
192      scaling_rob[c, "pop_pib_alpha"] <- summary(scalepib)$coefficients[2,1]
193      scaling_rob[c, "pop_pib_r.squared"] <- summary(scalepib)$r.squared
194      scaling_rob[c, "pop_pib_p.value"] <- summary(scalepib)$coefficients[2,4]
195      scaling_rob[c, "pop_ruas_intercept"] <- summary(scaleruas)$coefficients[1,1]
196      scaling_rob[c, "pop_ruas_alpha"] <- summary(scaleruas)$coefficients[2,1]
197      scaling_rob[c, "pop_ruas_r.squared"] <- summary(scaleruas)$r.squared
198      scaling_rob[c, "pop_ruas_p.value"] <- summary(scaleruas)$coefficients[2,4]
199
200      rob_matrx_y[r,f] <- summary(scalepib)$coefficients[2,1]
201      rob_matrx_area[r,f] <- summary(scalearea)$coefficients[2,1]
202      rob_matrx_ruas[r,f] <- summary(scaleruas)$coefficients[2,1]
203    }
204    f <- 0
205  }
206
207  scaling_rob_alpha <- scaling_rob[,c(1,5,9,13)]
208  scaling_rob_intercept <- scaling_rob[,c(1,3,7,11)]
209  scaling_rob_r <- scaling_rob[,c(1,6,10,14)]
210  scaling_rob_pvalue <- scaling_rob[,c(1,7,11,15)]
211
212  melt_scaling_rob_alpha <- melt(scaling_rob_alpha, id.vars="dens_min")
213  melt_scaling_rob_intercept <- melt(scaling_rob_intercept, id.vars="dens_min")
214  melt_scaling_rob_r <- melt(scaling_rob_r, id.vars="dens_min")
215  melt_scaling_rob_pvalue <- melt(scaling_rob_pvalue, id.vars="dens_min")
216
217
218  #heatmap
219  my_palette <- colorRampPalette(c("yellow", "red"))(n = 50)
220  heatmap(rob_matrx_y, Rowv=NA, Colv=NA, col = my_palette)
221
222  my_palette <- colorRampPalette(c("green", "yellow"))(n = 50)
223  heatmap(rob_matrx_area, Rowv=NA, Colv=NA, col = my_palette)
224  heatmap(rob_matrx_ruas, Rowv=NA, Colv=NA, col = my_palette)
225  dev.off()
```

```
226
227
228  #plot scaling final
229  ggplot(melt_scaling_rob_alpha, aes(x = dens_min, y = value, colour = variable)) + geom_
         line()
230  ggplot(melt_scaling_rob_intercept, aes(x = dens_min, y = value, colour = variable)) +
         geom_line()
231  ggplot(melt_scaling_rob_r, aes(x = dens_min, y = value, colour = variable)) + geom_line
         ()
232  ggplot(melt_scaling_rob_pvalue, aes(x = dens_min, y = value, colour = variable)) + geom
         _line()
```

Apêndice B – Código do modelo em Python

```
1
2  import math
3  import numpy as np
4  import random as rn
5  import matplotlib.pyplot as plt
6  from matplotlib.pylab import *
7  import pandas as pd
8  import statsmodels.api as sm
9  import csv
10 from pandas import DataFrame
11
12 runs = [0,1,2,3]
13 populacoes = [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2500, 5000, 7500,10000]
14
15 #areas em ordem decrescente
16 areas = [30000, 25000,20000,15000, 10000,7500, 5000, 2000, 1500, 1000, 750, 500, 350,
       200, 100, 50, 20]
17
18
19 Tempo = 100
20 encontros = 0
21
22 b=2
23 passo = 1
24 gag_mean = 20
25 gag_stdv = 10
26 yag_mean = 30
27 yag_stdv = 10
28 a_zero = 2
29
30 porc_morte_corte = 20
31
32 #define as funcoes que serao utilizadas
33 #-------------------------------------------------------------------------
34 def cria_pop(populacao):
35   p = []
36   city_streets_pop = set(city_streets).copy()
37   for i in range(populacao):
38     if len(city_streets_pop) > 0:
39       x, y = city_streets_pop.pop()
40       ag={'id':i, 'latag': x  ,'lonag': y, \
41         'gag': rn.gauss(gag_mean, gag_stdv), 'yag': rn.gauss(yag_mean, yag_stdv) ,'tag':
              10,'azero': a_zero,'lag': passo} #distribuicao
42       p.append(ag)
43   return p
44
45 #-------------------------------------------------------------------------
46 def plota_city(p,cor):
47   for x, y in city:
48     plt.plot(x,y,"s", ms=0)
49
50   for x, y in city_blocks:
51     plt.plot(x,y,"s", ms=block_print)
52
53   for ag in p:
54     plt.plot(ag['latag'],ag['lonag'],cor)
```

```
55      plt.ylim([-1,Lcity+1])
56      plt.xlim([-1,Lcity+1])
57  #-------------------------------------------------------------------------
58  def ag_walk(p):
59    p_tmp = []
60    for ag in p:
61      pss = [-passo, passo]
62      directions = ['x', 'y']
63      possible2 = 0
64
65      possible = set([])
66      possible.add(((ag["latag"]+1) % Lcity, ag["lonag"]))
67      possible.add(((ag["latag"]-1) % Lcity, ag["lonag"]))
68      possible.add((ag["latag"], (ag["lonag"]+1) % Lcity))
69      possible.add((ag["latag"], (ag["lonag"]-1) % Lcity))
70
71      possible2 = set.intersection(set(city_streets), possible)
72
73      choice = rn.choice(list(possible2))
74
75      ag['latag'] = choice[0]
76      ag['lonag'] = choice[1]
77
78      ag['yag'] = ag['yag'] - (passo*(b+densidade))
79
80      if ag['yag']>0 :
81        p_tmp.append(ag)
82    return p_tmp
83
84
85  #-------------------------------------------------------------------------
86  def ag_interact(p):
87    for i in xrange(len(pop)):
88      for j in xrange(i+1, len(pop)):
89          if (abs((pop[i]['latag']-pop[j]['latag'])) <= pop[i]['azero'] ):
90
91            if (abs((pop[i]['lonag']-pop[j]['lonag'])) <= pop[i]['azero']):
92
93              pop[i]['yag'] = pop[i]['yag'] + pop[i]['gag']
94              pop[j]['yag'] = pop[j]['yag'] + pop[j]['gag']
95              global encontros
96              encontros = encontros + 1
97
98  #-------------------roda o modelo-------------------------#
99  lines = []
100
101 data = []
102 data_pop_inicio = []
103 data_pop_criada = []
104 data_pop_final = []
105 data_pop_ponderada = []
106 data_encontros = []
107 teorico_encontros = []
108 teorico_area = []
109 pop_tempo = []
110
111 alpha_inicial = []
112 alpha_final = []
113 alpha_media = []
114 alpha_ponderada = []
```

```
115
116 beta_inicial = []
117 beta_final = []
118 beta_media = []
119 beta_ponderada = []
120
121 n_cidades = []
122
123 time_step = []
124 data_area = []
125 pessoas = []
126 Y = []
127
128 time = []
129 p = []
130 i = []
131
132 for run in runs:
133   print run
134
135   for area in areas:
136     city= set([])
137     city_blocks = set([])
138     city_streets=set([])
139     city_streets_pop=set([])
140     Area = area
141     print ("AREA: " + str(Area))
142
143     city_block_size = 4
144     Lcity = int(math.ceil(np.sqrt(Area)))
145
146     for x in range(Lcity):
147       if x % city_block_size!=0:
148         for y in range(Lcity):
149           if y % city_block_size !=0:
150             city_blocks.add((x,y))
151
152     for x in range(Lcity):
153       for y in range(Lcity):
154         city.add((x,y))
155
156
157     city_streets = city - city_blocks
158
159     city_streets = list(city_streets)
160     np.random.shuffle(city_streets)
161     city = list(city)
162     city_blocks = list(city_blocks)
163
164     for population in populacoes:
165       data_pop_inicio.append(population)
166       print ("POP: ", population)
167       pop  = cria_pop(population)
168       print len(pop)
169       data_pop_criada.append(len(pop))
170       densidade = city_block_size * population*1.0 / Area
171
172       for t in range(Tempo+1):
173         if len(pop) > 0:
174           densidade = city_block_size * len(pop)*1.0 / Area
```

```
175             pop = ag_walk(pop)
176             ag_interact(pop)
177             time_step.append(t)
178             pop_tempo.append(len(pop))
179         data_pop_ponderada.append(sum(pop_tempo)/len(pop_tempo))
180         data_pop_final.append(len(pop))
181         data_area.append(Area)
182         Y.append (sum(i['yag'] for i in pop))
183         pop_tempo = []
184         pop = []
185
186
187     def fx(x, y):
188         return ((y-x)/float(y))*100
189
190
191     data = zip(data_pop_inicio, data_pop_ponderada, data_pop_final, data_area, Y)
192     print ("Y" + str(Y))
193     print data
194     df = pd.DataFrame(data)
195     df.columns = ['pop_inicio', 'pop_ponderada','pop_final', 'area', 'Y']
196     df['porc_morte'] =  np.vectorize(fx)(df['pop_final'], df['pop_inicio'])
197     print df
```