

V!RUS13

para tempos difíceis
Boas notícias
Good news
for hard times

a n o 2 0 1 6 y e a r
s e m e s t r e 2 s e m e s t e r

revista do nomads.usp | nomads.usp journal
issn:2175.974x | CC BY - NC
DOI 10.4237/virus_journal

A IMAGEM DA CIDADE DE DADOS Lev Manovich Agustín Indaco

Como citar esse texto: MANOVICH, L.; INDACO, A. A imagem da cidade de dados. Traduzido do inglês por Marcelo Tramontano. **V!RUS**, São Carlos, n. 13, 2016. Disponível em: <<http://www.nomads.usp.br/virus/virus13/?sec=4&item=1&lang=pt>>. Acesso em: 00 m. 0000.

Lev Manovich é Doutor em Estudos Visuais e Culturais, Professor da City University of New York (CUNY), diretor da Software Studies Initiative. Estuda análise cultural, computação social, *big data* e sociedade, humanidades digitais, história da arte, teoria e história da mídia, estudos de software, visualização de dados.

Agustín Indaco é Mestre em Economia, pesquisador da Software Studies Initiative. Estuda a intersecção entre microeconomia aplicada, saúde e *big data*, além de análise de comportamento econômico através de dados obtidos em mídias sociais.

Palavras-chave: Big Data; Instagram; Coeficiente de Gini; Nova York; Redes sociais.

A imagem de uma cidade de dados

O conteúdo de redes sociais¹ compartilhado, hoje, nas cidades, como as imagens, *tags* e descrições do Instagram, é uma chave de leitura da vida urbana contemporânea. Ela informa as pessoas sobre onde se encontram atividades e lugares que lhes interessam, e permite-lhes compartilhar suas experiências urbanas e auto-representações. Ela cria uma "imagem de cidade" tanto para os residentes quanto para o mundo exterior. Pode-se argumentar que a identidade de qualquer cidade compõe-se, hoje, tanto pelo conteúdo mediatizado nela compartilhado através de redes sociais quanto por sua infraestrutura e suas atividades econômicas. É por isso que todo estudo sobre experiência urbana e auto-representação deve considerar o conteúdo compartilhado por redes sociais em uma determinada cidade.

A análise computacional de grandes quantidades de fotos e vídeos geradas e compartilhadas por usuários em áreas específicas pode também ajudar-nos a entender *como* as pessoas vivenciam a arquitetura e as estruturas urbanas e *o que* fazem nelas. Isso pode ser feito em qualquer escala, da urbana até o nível hiperlocal das ruas, edifícios ou seus interiores. É possível comparar a porcentagem de fotos do Instagram que mostram o ambiente construído em diferentes cidades, analisar quais enquadramentos são mais populares para cada monumento, e quais emoticons eles evocam segundo a hora do dia. Podemos comparar esses padrões segundo residentes e turistas, segundo diferentes gêneros, faixas etárias, etc.. Em resumo, pode ser muito útil estarmos aptos a analisar pegadas digitais do que grande número de pessoas *faz* no ambiente construído e como elas o vêem e o utilizam.



Fig. 1: Comparação de atividades no Instagram nas cidades de Bangcoc, Berlim, Moscou, Nova York, São Paulo e Tóquio (da esquerda para a direita, de alto a baixo). Cada visualização apresenta 20.000 imagens compartilhadas durante uma semana em cada cidade. Fonte: Lev Manovich, Jay Chow, 2013-2016. Copyright: Software Studies Initiative.

Em nosso *Software Studies Lab* (softwarestudies.com), localizado na Universidade da Califórnia, San Diego, e no *The Graduate Center*, da City University of New York, analisamos mais de 16 milhões de fotos do Instagram compartilhadas em 17 cidades globais a partir de 2012. As equipes de pesquisa incluíram cientistas de dados, desenvolvedores de *software*, designers de visualização de dados, teóricos dos meios digitais, historiadores da arte, economistas e *urban designers*. Começando por uma comparação geral entre 2,3 milhões de imagens compartilhadas em 13 cidades globais (*Phototrails*, 2013, <http://phototrails.net/>), nos concentramos, em seguida, em categorias mais específicas de imagens, filtradas por tipo de conteúdo: autorretratos (*Selfiecity*, 2014, <http://selfiecity.net>), uma rua específica (*On Broadway*, 2014, <http://on-broadway.net>), e uma combinação entre uma área de uma cidade e um período de tempo (o centro de Kiev [Ucrânia] durante a revolução Maidan, de 2014, em <http://www.the-everyday.net/>). As figuras que ilustram o presente ensaio apresentam alguns desses projetos.

Se o trabalho do laboratório demonstra quão úteis podem ser os dados de redes sociais para entender o hiperlocal, ele também revela as limitações desse tipo de dado. Em muitas áreas centrais urbanas, as redes sociais têm resolução espacial e temporal bem alta. Para o projeto *Inequaligram*, o laboratório coletou todas as 7.442.454 fotos geocodificadas do Instagram compartilhadas publicamente em

Manhattan durante os cinco meses entre março e agosto de 2014. Por exemplo, em uma única área de 30m X 100m, em Times Square, os usuários do Instagram compartilharam 43.541 imagens. Em muitas outras áreas de Manhattan, no entanto, as pessoas compartilharam apenas algumas dezenas de imagens nos mesmos cinco meses. Essa baixa densidade em várias partes das cidades limita a utilidade das redes sociais para se entender a vida urbana dessas áreas. Outra limitação é demográfica: por exemplo, em muitas cidades globais, pode ser que apenas pessoas jovens e com bom nível educacional postem conteúdo. Assim, enquanto em alguns casos as redes sociais são uma excelente fonte para o estudo de locações hiperlocais, em outros, a observação direta ou enquetes serão mais úteis. Análises computacionais em larga escala de redes sociais urbanas podem, portanto, apenas complementar - e não substituir - outros métodos de pesquisa em estudos urbanos, design e arquitetura.

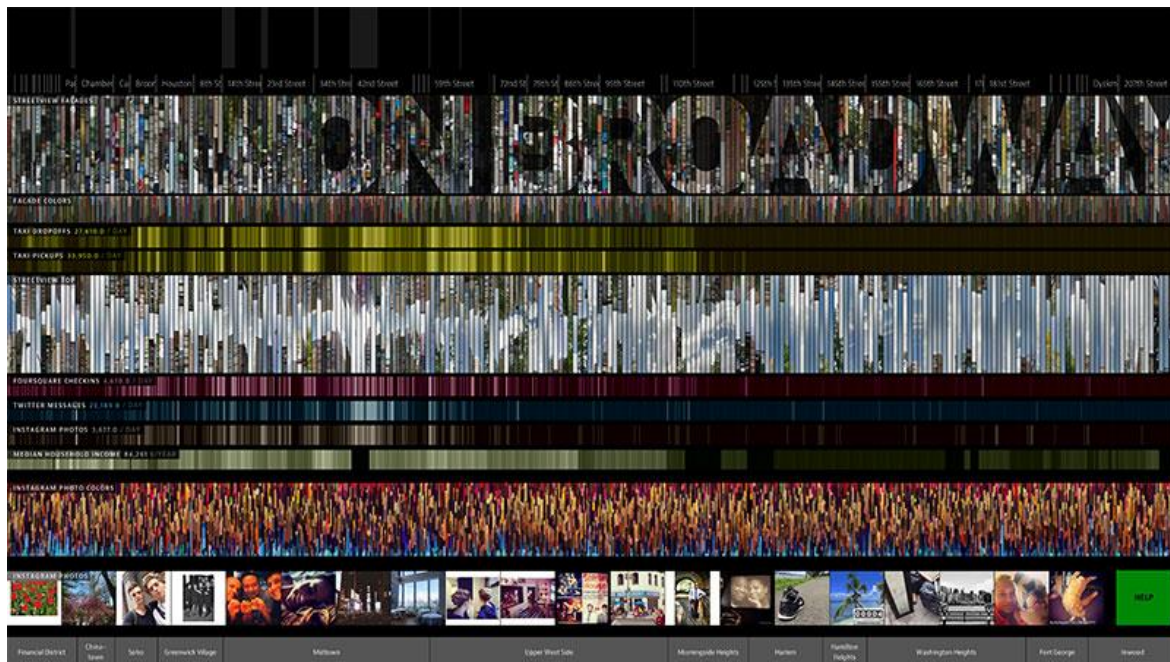


Fig. 2: Projeto *On Broadway* combinando uma instalação interativa e um website (<http://on-broadway.nyc>). Captura de tela da instalação interativa com uma vista em *zoom-out* completo mostrando toda a extensão (13 milhas) da rua Broadway em Manhattan. A instalação foi apresentada na New York Public Library, entre dezembro de 2014 e janeiro de 2016. Fonte: Daniel Goddemeyer, Moritz Stefaner, Dominikus Baur, Lev Manovich, 2014. Copyright: Software Studies Initiative.

On Broadway

Os dois estudos mais detalhados de redes sociais em escala hiperlocal produzidos até agora pelo *Software Studies Initiative* são os projetos *On Broadway* e *Inequaligram*. Encomendado pela Biblioteca Pública de Nova York, *On Broadway* (2015) baseou-se em um conceito original do *media designer* Daniel Goddemeyer. Moritz Stefaner foi o responsável pela direção artística e pelo design de visualização de dados, e Dominikus Baur, pelo desenvolvimento de *software*.

O foco do projeto foi uma única rua, muito longa: a parte da Broadway que cruza toda Manhattan (21 quilômetros). A equipe de projeto também quis incluir uma área um pouco mais larga do que a rua em si para captar atividades próximas. Para definir estas áreas, os pesquisadores dividiram a rua Broadway em segmentos de 30 metros de comprimento cada um e, a seguir, selecionaram áreas retangulares de 100 metros de largura em torno de cada segmento. O resultado foram 713 retângulos idênticos,



de 30m X 100m. O projeto visualiza e compara imagens de redes sociais e outros dados através dessas 713 áreas.

O principal objetivo do projeto era construir um novo mecanismo para se navegar em uma "data city", constituído de várias camadas de imagens e dados. Perguntamo-nos se haveria outro modo de visualizar estruturas urbanas além de mapas, gráficos e números. O resultado das múltiplas explorações é visualmente rico, uma interface centrada na imagem, na qual os números desempenham um papel secundário e mapas não são utilizados. Essa interface propõe uma nova metáfora visual para se pensar a cidade: uma pilha vertical de camadas de imagens e dados. Há 13 camadas assim no projeto, todas atreladas a locações ao longo da Broadway. Elas incluem imagens compartilhadas ao longo da rua no Instagram e no Twitter, imagens do Google Street View, *check-ins* no Foursquare, corridas de táxis e indicadores econômicos e sociais selecionados do U.S. Census [recenseamento geral dos Estados Unidos, n. t.]. Ao todo, utilizamos mais de 30 milhões de dados e imagens para representar atividades ao longo de uma única rua.

Ao mover-se ao longo da representação da rua, vê-se uma seleção de fotos do Instagram de cada área, imagens do Google Street View dos lados esquerdo e direito da rua e do céu, além das principais cores extraídas dessas imagens. Pode-se, ainda, ver o número médio de partidas e de chegadas de corridas de táxi, postagens no Twitter com imagens, e a renda familiar média nas áreas da cidade atravessadas pela Broadway. Para auxiliar a navegação, acrescentamos camadas adicionais com o nome dos bairros de Manhattan cortados pela avenida, das ruas transversais e de marcos urbanos.

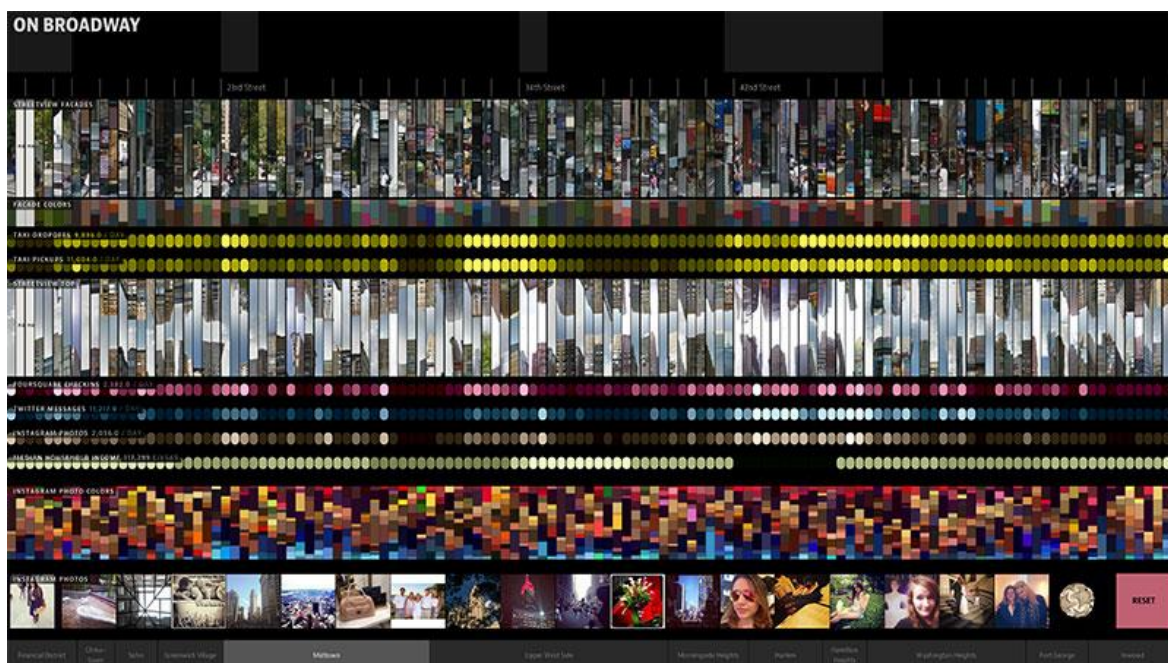


Fig. 3: *On Broadway*. Captura de tela da instalação interativa. Vista com zoom ao nível do bairro mostrando o centro de Manhattan. A instalação foi apresentada na New York Public Library, entre dezembro de 2014 e janeiro de 2016. Fonte: Daniel Goddemeyer, Moritz Stefaner, Dominikus Baur, Lev Manovich, 2014. Copyright: Software Studies Initiative.

V!RUS13

para tempos difíceis
Boas notícias
Good news
for hard times

ano 2016 year | semestre 2 semester
revista do nomads.usp | nomads.usp journal
issn:2175.974x | CC BY - NC
DOI 10.4237/virus_journal

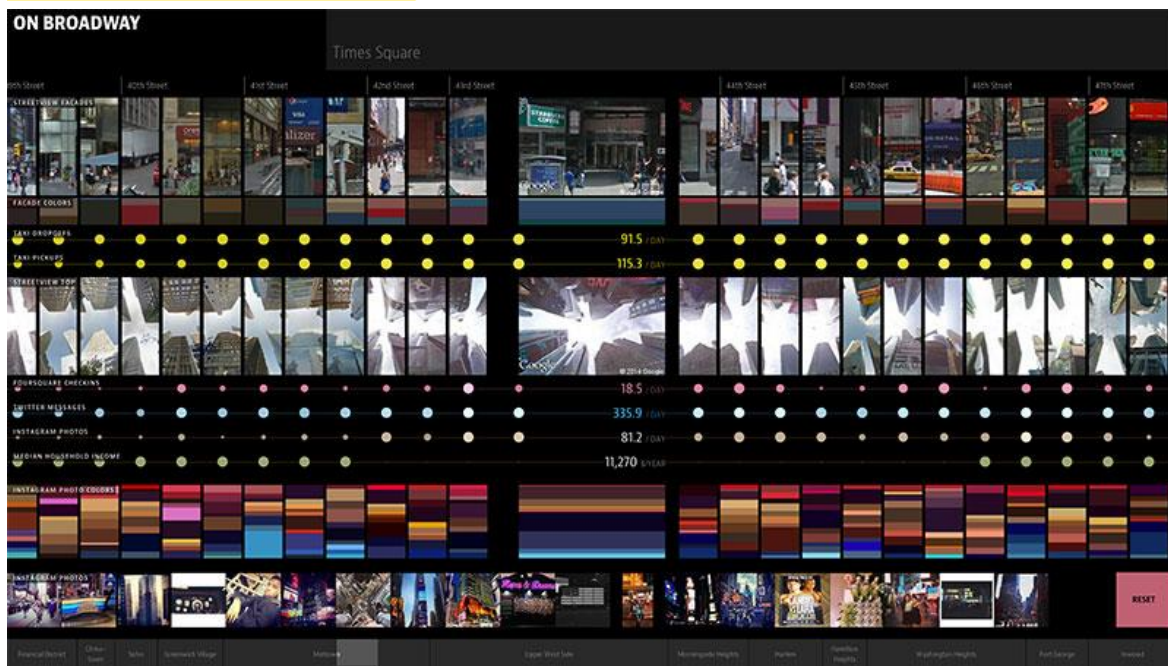


Fig. 4: On Broadway. Captura de tela da instalação interativa. Vista com zoom ao nível do quarteirão na região de Times Square, em Manhattan. Fonte: Daniel Goddemeyer, Moritz Stefaner, Dominikus Baur, Lev Manovich, 2014. Copyright: Software Studies Initiative.



Fig. 5: On Broadway. Gráfico comparando as camadas de dados utilizadas para representar a rua Broadway no projeto. A rua está projetada sobre o eixo horizontal (Norte-Sul torna-se esquerda-direita). A altura do gráfico em cada localização corresponde ao volume de determinada camada de dados nesta localização. Fonte: Daniel Goddemeyer, Moritz Stefaner, Dominikus Baur, Lev Manovich, 2014. Copyright: Software Studies Initiative.

Esse "corredor" multi-camadas da Broadway pode ser explorado em diferentes escalas. Em uma visão em *zoom out*, vêem-se os 21 quilômetros da rua. Para isso, dispusemos estreitas fatias verticais de cada foto do Google Street View. Quando se começa a aproximar com o *zoom*, as fatias vão se tornando mais largas. Finalmente, na visão mais próxima do *zoom*, a imagem da área selecionada é mostrada inteiramente. Todas as peças visuais de todas as camadas e todos os números referentes a atividades agregadas são instantaneamente atualizados quando o usuário move seu campo de visão para a direita ou para a esquerda, ou altera o nível de *zoom*.

Ao explorar as camadas de dados ao longo da Broadway, a equipe do projeto percebeu que os volumes das camadas correlacionavam-se fortemente. Informalmente, isso pode ser visualizado na Figura 5, que apresenta os volumes de todas as variáveis: as alturas das variáveis aumentam e diminuem juntas. Como é possível interpretar essa "cidade correlacionada"? Os dados sugerem que a desigualdade social e a fratura digital são agora acompanhadas por uma fratura ainda mais extrema nas redes sociais. Em áreas ricas, as pessoas ganham mais dinheiro, usam táxis e postam mais imagens no Instagram e no Twitter. Em áreas pobres, as pessoas ganham menos, raramente usam táxis e postam muito menos imagens em redes sociais *online*.

Inequaligram

Os membros do laboratório decidiram explorar mais profundamente estas conexões. Isso resultou em um novo projeto chamado *Inequaligram* (2016), criado pelos autores deste ensaio. Nele, são analisadas características das postagens no Instagram e indicadores socioeconômicos extraídos do U.S. Census para Manhattan.

O U.S. Census reporta características socioeconômicas agregadas de populações usando um tipo de divisão chamada "setores". Existem 287 setores censitários em Manhattan. A população média de cada um deles situa-se entre 3 mil e 4 mil pessoas, e sua área média é de 0,36 km². O projeto utiliza estes setores para comparar padrões de compartilhamento no Instagram e indicadores tais como renda e taxa de desemprego.

A equipe do projeto escolheu o Instagram para essa análise porque ele apresenta a mais forte identidade geográfica e espacial dentre todos os principais serviços *online* de redes sociais. Enquanto tweets e postagens no Facebook podem também ter geocoordenadas e informam sobre eventos locais no entorno do usuário no momento da postagem, com frequência as imagens no Instagram capturam diretamente tais eventos e mostram os usuários em lugares específicos. E uma vez que as postagens no Instagram contêm ou uma imagem, ou um vídeo, metadados sobre data e tempo, descrições e *hashtags*, elas permitem estudar representações coletivas de vida urbana em cada uma dessas dimensões separadamente. Por exemplo, podemos comparar o número de imagens compartilhadas em áreas distintas, a presença de diferentes assuntos nessas imagens, as *hashtags* mais populares e as menos comuns, a maneira como as pessoas se vestem, etc.. Estas e muitas outras características podem ser automaticamente extraídas das postagens no Instagram utilizando-se técnicas de *data science*, disponíveis em programas computacionais de código aberto.

Os conteúdos de redes sociais compartilhados em uma determinada área podem combinar contribuições de diversos tipos de usuários: pessoas que residem na área, pessoas que vivem em outras partes da cidade ou em periferias mas que, em dias

úteis, permanecem um tempo significativo nessa área a trabalho, turistas nacionais ou internacionais em visita à cidade, empresas localizadas na área, entre outros. Juntos, os conteúdos compartilhados por todos esses usuários criam uma "voz" coletiva desta área específica da cidade. De fato, uma cidade pode ser comparada a um coral formado por todas essas vozes, ainda que, é claro, elas não estejam necessariamente interpretando uma mesma peça musical. A aplicação do conceito de desigualdade a um conjunto selecionado dessas vozes urbanas pode nos trazer novos modos de entender a cidade e prover uma métrica adicional para se comparar inúmeras cidades ao redor do mundo.

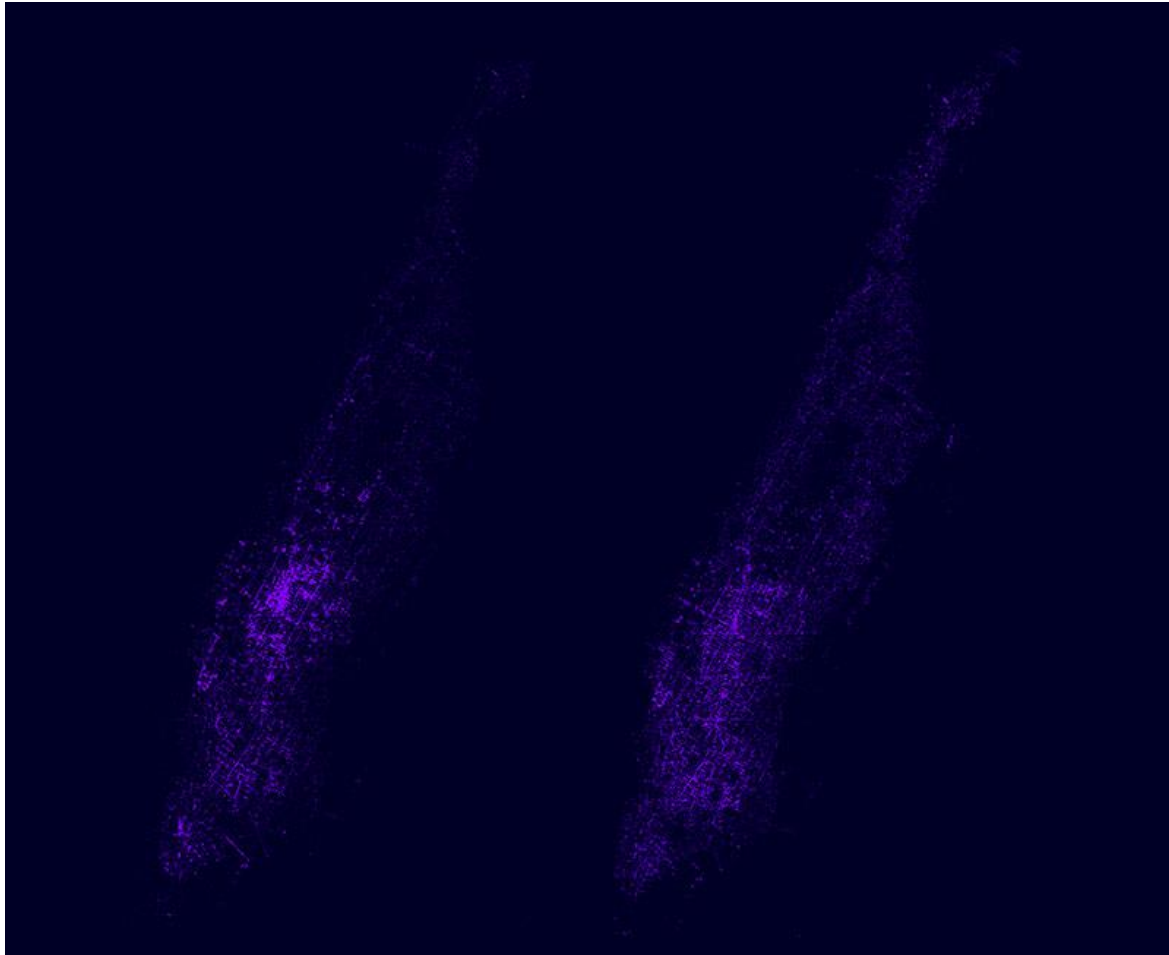


Fig. 6: *Inequaligram*. Locações de imagens do Instagram compartilhadas por visitantes de Nova York (esquerda) e locais (direita). Cada mapa utiliza um conjunto de 100.000 imagens, extraídas do conjunto maior de 7.442.454 imagens georreferenciadas compartilhadas publicamente em Manhattan entre março e julho de 2014. Fonte: Agustin Indaco, Lev Manovich, 2016. Copyright: Software Studies Initiative.

Em contraste com outros serviços de redes sociais, as imagens e as locações veiculadas pelo Instagram criam uma "imagem de cidade" tanto para os locais como para os visitantes. Desse modo, temos que entender o que essas representações coletivas contêm e como suas características se relacionam tanto com a estrutura arquitetônica da cidade (por exemplo, presença de pontos turísticos) como com a estrutura socioeconômica (por exemplo, locações em áreas ricas ou pobres).

Planejadores urbanos e arquitetos sabem mapear as estruturas físicas das cidades, mas quais seriam, para eles, os meios mais informativos para se mapear e analisar

redes sociais? Em uma cidade como Nova York, as pessoas compartilham uma quantidade enorme de imagens via Instagram, em certas áreas, e muito pouco, em outras. As imagens compartilhadas em algumas áreas podem, ainda, conter mais *hashtags* e descrições sobre a arquitetura local do que aquelas compartilhadas em outras áreas. Ao mapearmos essas características nas postagens, usando sua geolocalização, percebemos que sua distribuição espacial é bastante irregular.

Para quantificar exatamente quão irregulares são essas distribuições, a equipe do *Inequaligram* desenvolveu um conceito novo, o de "desigualdade em redes sociais". Tal conceito permite comparar quantitativamente padrões espaciais em atividades relevantes em redes sociais, entre partes de uma cidade, entre várias cidades, ou entre quaisquer outras áreas. Esse conceito foi definido a partir de uma analogia com o conceito de desigualdade econômica. A desigualdade econômica indica como uma característica econômica, como renda, riqueza ou consumo, é distribuída em uma cidade, um país, ou entre países. Da mesma forma, a desigualdade em redes sociais indica como uma característica de conteúdos compartilhados em redes sociais é distribuída em áreas geográficas. Exemplos dessas características são o número de fotos compartilhadas por todos os usuários de uma rede social, como o Instagram, em uma determinada área da cidade, o número de *hashtags* e de *hashtags* pouco comuns.

Para compensar as diferenças das dimensões geográficas entre os setores, os dados Instagram foram normalizados segundo o tamanho do setor. O projeto também usou as datas das imagens compartilhadas para estimar se um usuário específico vive em Manhattan, ou está apenas visitando. Dados do Escritório Norte-americano das Indústrias de Viagens e Turismo (*U.S. Office of Travel and Tourism Industries*) indicam que o visitante médio permanece 10,5 dias em Nova York. Decidimos usar um período um pouco maior, de 12 dias, e considerar um usuário como um "visitante" se ele postou todas as suas fotos em um único período de 12 dias, do total dos cinco meses de nossa coleta de dados. Por outro lado, se um utilizador compartilhou pelo menos duas fotos dentro de qualquer intervalo maior do que 12 dias, considerou-se esta pessoa um "local". Embora este método muito simples não seja preciso, a análise dos dados mostra que, de forma eficaz, ele realmente diferencia capturas entre estes dois grupos. O nosso conjunto de dados contém 5.918.408 milhões de imagens de 366.539 contas Instagram de residentes locais, e 1.524.046 imagens de 505.345 contas pertencentes aos visitantes.

Locals in selected Manhattan areas below 59th street

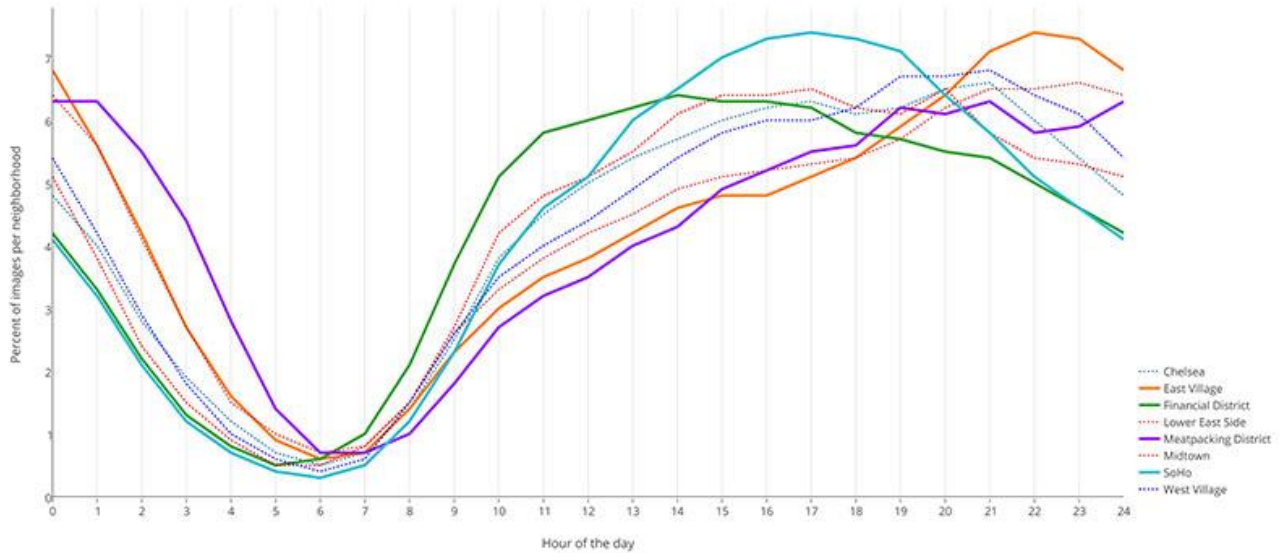


Fig. 7: *Inequaligram*. Proporções horárias de imagens compartilhadas por locais em bairros selecionados de Manhattan abaixo da rua 59. O gráfico usa informações temporais de 5.918.408 imagens compartilhadas por 366.539 residentes. Fonte: Agustin Indaco, Lev Manovich, 2016. Copyright: Software Studies Initiative.

Visitors in selected Manhattan areas below 59th street

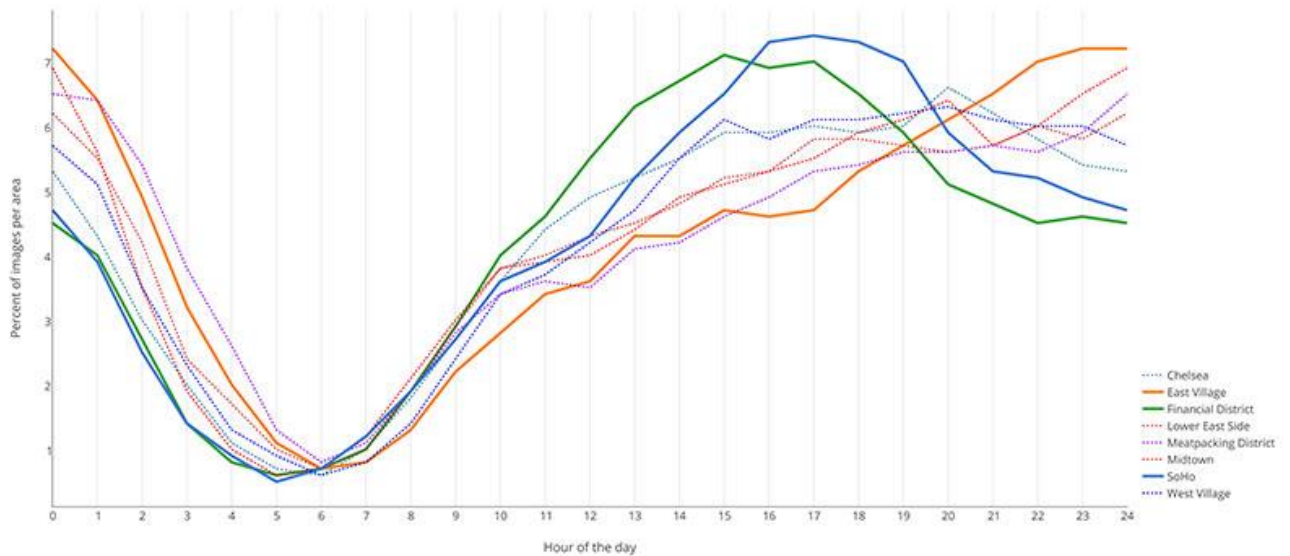


Fig. 8: *Inequaligram*. Proporções horárias de imagens compartilhadas por visitantes em bairros selecionados de Manhattan abaixo da rua 59. O gráfico usa informações temporais de 1.524.046 imagens compartilhadas por 505.345 visitantes. Fonte: Agustin Indaco, Lev Manovich, 2016. Copyright: Software Studies Initiative.

Locals in selected Manhattan areas above 59th street

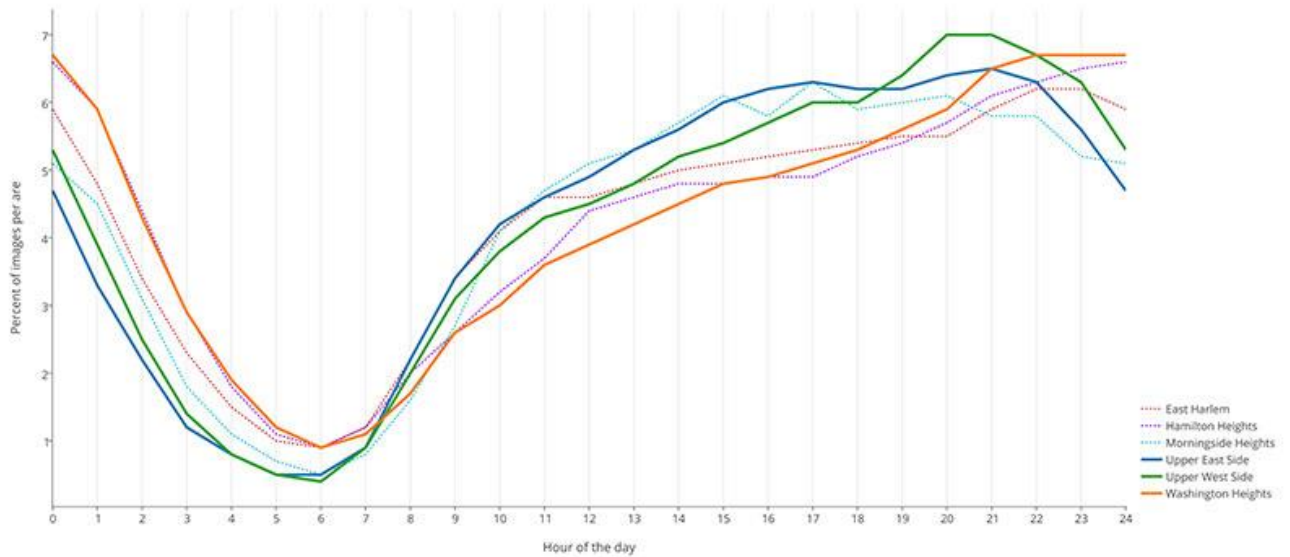


Fig. 9: Inequaligram. Proporções horárias de imagens compartilhadas por locais em bairros selecionados de Manhattan acima da rua 59. O gráfico usa informações temporais de 5.918.408 imagens compartilhadas por 366.539 residentes. Fonte: Agustin Indaco and Lev Manovich, 2016. Copyright: Software Studies Initiative.

Para comparar a desigualdade em redes sociais entre esses dois grupos, em Manhattan, o *Inequaligram* decidiu usar o mais popular instrumento de medição de desigualdade econômica: o coeficiente de Gini. Este instrumento é usado na maioria das discussões sobre desigualdade de renda e de riqueza, tanto em Economia quanto na grande imprensa. No caso do Instagram, se as pessoas compartilhassem exatamente o mesmo número de imagens cada uma, em cada setor censitário da cidade, isso significaria igualdade completa e o coeficiente de Gini seria igual a 0. Se, ao contrário, as pessoas compartilhassem todas as imagens em apenas um setor e nenhuma imagem em nenhum outro setor, isso significaria completa desigualdade e o coeficiente de Gini seria igual a 1.

O *Inequaligram* verificou que o coeficiente de Gini para o número de imagens compartilhadas em Manhattan, considerando todos os setores censitários, é de 0,494 para os locais e de 0,669 para os visitantes. Para o número total de *hashtags*, o coeficiente é ainda mais alto: 0,514 para os locais e 0,678 para os visitantes. Contextualizando, a desigualdade relacionada às imagens dos visitantes de Manhattan (Gini = 0,669) é maior do que a desigualdade de renda no país mais desigual do mundo (Seychelles, onde Gini = 0,658). Já as imagens compartilhadas em redes sociais por locais têm um coeficiente de Gini similar ao de países que situam-se entre a 25a. e a 30a. posições na lista de países segundo a desigualdade de renda. São países como Costa Rica (0,486), México (0,481) e Equador (0,466).

O que leva à alta desigualdade de compartilhamento no Instagram entre as regiões de Manhattan? Os visitantes compartilham a maioria de suas imagens no centro de Manhattan (área de grandes lojas e hotéis), em torno de monumentos famosos, como a Times Square e o Flatiron Building, e nas áreas de restaurantes e bares noturnos, como East Village e Lower East Side. No caso dos locais, nossa análise sugere que as diferenças de atividade nas redes sociais entre as regiões da cidade são, em grande medida, impulsionadas por padrões de deslocamento. Durante a semana, nos horários de trabalho, os moradores das regiões menos ricas, tais como aquelas acima da rua 100, trabalham nas regiões mais ricas, abaixo da rua 100 e, particularmente, no

centro da cidade. É aí que eles compartilham imagens no Instagram durante o dia e, por isso, suas ações são contabilizadas junto a essas áreas.

Observando os padrões de desigualdade em compartilhamentos no Instagram de moradores e visitantes em conjunto, o *Inequaligram* percebeu que as áreas de Manhattan abaixo da rua 100, onde se localiza a maioria dos negócios, são também as mais populares entre os visitantes. Temos, assim, um efeito de dupla amplificação: as contribuições dos residentes abastados dessas áreas em redes sociais são amplificadas pelas contribuições de pessoas que se deslocam até lá a trabalho, vindo de outras partes de Manhattan, e também por contribuições de visitantes de fora da cidade. Comparando as estatísticas de redes sociais com indicadores do Censo para os setores de Manhattan, verificamos que a desigualdade do número de imagens do Instagram, entre setores, é maior do que as desigualdades nos níveis de renda e desemprego. Os coeficientes de Gini são 0,32 (renda mediana), 0,22 (renda mediana), 0,35 (taxa de desemprego) e de 0,49 (número de imagens no Instagram compartilhadas por moradores locais). Este é um resultado muito interessante e original.

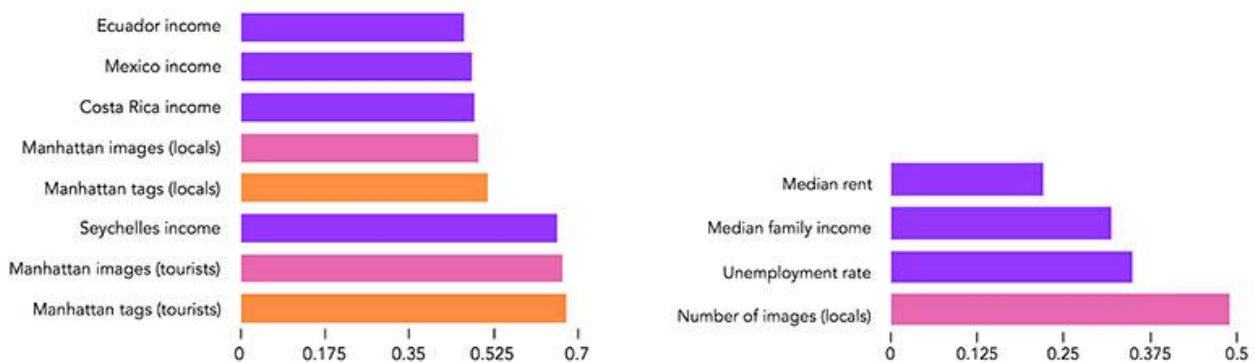


Fig. 10: *Inequaligram*. Esquerda: Coeficiente Gini de desigualdade para imagens compartilhadas no Instagram, em Manhattan, comparado com coeficientes de desigualdade de renda em países selecionados. Direita: Coeficiente de Gini de desigualdade para imagens compartilhadas no Instagram por locais em 287 setores censitários de Manhattan e indicadores econômicos selecionados do Censo (renda, desemprego) para os mesmos setores. "Setores" são divisões espaciais utilizadas pelo Censo nacional norte-americano em relatórios de resultados das pesquisas. Os coeficientes de Gini para indicadores econômicos são calculados usando dados do Censo de 2014. Source: Agustin Indaco, Lev Manovich, 2016. Copyright: Software Studies Initiative.

Estudando a vida urbana na era dos dados

Muitas são as possibilidades analíticas oferecidas pelos *big data* extraídos de redes sociais aos estudiosos urbanos e arquitetos que podem ser exploradas além do que foi discutido aqui. Ao fazer o *download*, análise e visualização de fotos compartilhadas por usuários, juntamente com as suas *tags*, descrições, informações temporais e coordenadas geográficas, os pesquisadores do Software Studies Lab construíram uma "imagem de cidade" coletiva e conseguiram perceber como ela se altera ao longo do tempo. O conceito de desigualdade em redes sociais nos permite mensurar alterações dessa imagem de uma região da cidade a outra, além de comparar essas imagens em escalas espaciais arbitrárias.



Notas

¹ Nota da tradução: o termo em inglês "social media", referindo-se às redes sociais *online* via Internet, será traduzido aqui apenas como "redes sociais".