

Um conjunto de dados geoespaciais multiescala e um Painel de Visualização Interativa para Epidemiologia Computacional e Aberta Pesquisa científica

Muhammad Usman , Departamento de Ciência da Informação e da Computação, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Arábia Saudita

Honglu Zhou , Seonghyeon Moon e Xun Zhang, Departamento de Ciência da Computação, Rutgers University, Piscataway, NJ, 08854, EUA

Petros Faloutsos , Departamento de Engenharia Elétrica e Ciência da Computação, Universidade de York, Toronto, ON, M3J 1P3 Canadá

Mubbassir Kapadia, Departamento de Ciência da Computação, Rutgers University, Piscataway, NJ, 08854, EUA

A doença do coronavírus (COVID-19) continuou a atacar como uma doença altamente infecciosa e de rápida disseminação em 2020 e 2021. À medida que a comunidade de pesquisa respondeu ativamente a essa pandemia, vimos o lançamento de muitos conjuntos de dados e painéis de visualização relacionados ao COVID-19. No entanto, os recursos existentes são insuficientes para suportar modelagem ou simulação multiescalar e multifacetada, o que é sugerido como importante pela literatura de epidemiologia computacional. Este trabalho apresenta um conjunto de dados geoespaciais multiescala com curadoria com um painel de visualização interativo no contexto do COVID-19. Este conjunto de dados aberto permitirá que os pesquisadores conduzam vários projetos ou análises relacionadas ao COVID-19 ou simplesmente estudos científicos relacionados ao geoespaço. A plataforma de visualização interativa permite que os usuários visualizem a propagação da doença em diferentes escalas (por exemplo, nível de país para bairros individuais) e permite que os usuários interajam com as políticas aplicadas nessas escalas (por exemplo, fechamento de fronteiras e bloqueios) para observar seus impactos na epidemiologia.

facilidade chamada coronavírus (COVID-19) foi relatada.

Em dezembro de 2019, uma nova doença infecciosa respiratória, como uma infecção grave de alto risco que ainda não havia sido identificada em humanos, observamos uma rápida eclosão dessa trágica epidemia.¹ Nos últimos dois anos e meio, a COVID-19 evoluiu para uma preocupação global e afetou todos os aspectos da sociedade, incluindo deslocamentos diários, admissões em faculdades, ações, economia, estilo de trabalho das pessoas e até mesmo eleições.

Com o objetivo comum de combater esta pandemia, pesquisadores científicos de várias áreas e organizações têm colaborado ativamente e feito grandes esforços. Um esforço notável que não pode ser negligenciado é a divulgação de vários dados relacionados ao COVID-19 repositórios e painéis de visualização, como o COVID-19 Multilanguage Tweets Dataset² e o 2019 Novel Coronavirus Visual Dashboard operado pelo Johns Hopkins University Center,³ respectivamente, que permitiram a oferta de novas oportunidades de pesquisa e adotaram abordagens mais bem pensadas percepções. Um grupo de trabalho aberto de dados COVID-19 foi formado como uma iniciativa global e multiorganizacional que visa permitir o compartilhamento rápido de informações de saúde pública confiáveis e abertas

0272-1716 2022 IEEE

Digital Object Identifier 10.1109/MCG.2022.3230444

Data de publicação 19 de dezembro de 2022; data da versão atual 31 de janeiro de 2023.

TABELA 1. Definição do nó de geolocalização em cada escala em nosso conjunto de dados.

Escala	Descrição	Contar
1	País	206
2	Divisão administrativa de primeira ordem de um país (ou seja, província, estado e equivalência)	1445
3	Divisão administrativa que é menor que o nó da escala 2, mas maior que um edifício funcional (por exemplo, condado, cidade, vila, área censitária, bairro e freguesia)	7420
4	Edifício funcional (ou seja, aeroporto e porto)	2890

TABELA 2. Descrição das propriedades do nó de geolocalização.

Índice	Categoria	Nome	Data carimbada
1	localização geográfica	Latitude e longitude	Falso
2	População demográfica	Número da população	Falso
3	padrão de movimento	Séries temporais de tendências de mobilidade de seis funções sociais espaciais.	Verdadeiro
4	Intervenção não farmacêutica Intervenção governamental de "ficar em casa" para lidar com a COVID-19.		Verdadeiro

dados, bem como painéis de visualização para avançar na resposta a doenças infecciosas.⁴

Embora os recursos de dados existentes tenham se mostrado valiosos, descobrimos que eles são insuficientes para dar suporte à modelagem multiescalar e multifacetada de compreensão e simulação computacional de doenças. Na realidade, a propagação de doenças infecciosas, o movimento e comportamento humanos e as infra-estruturas sociais e civis estão intimamente interligados. Além disso, o mecanismo oculto que impulsiona sua coevolução e os fenômenos resultantes geralmente são dramaticamente distintos em diferentes escalas de observação. Compreender sua interação a partir de uma perspectiva multiescalar é fundamental para o desenho de políticas públicas e medidas de controle. Uma interpretação fácil para isso seria o fato de que as políticas geralmente são emitidas em múltiplas escalas, como política nacional ou política escolar, o que afeta diretamente as populações naquela escala correspondente e traz graus variados de efeitos. Por isso, acreditamos ser problemático ignorar as diferenças de escalas e modelar tudo junto. Para isso, oferecemos um painel de visualização interativo e uma plataforma multiescala

conjunto de dados geoespaciais, levando em consideração as propriedades geográficas e sociais dos locais e suas conexões, e as populações dentro deles.

Nosso conjunto de dados e painel são construídos sob o contexto do COVID-19. Partindo do entendimento geoespacial multiescala, consideramos cada geolocalização como um nó em nosso conjunto de dados. No total, quatro escalas são especificamente definidas, sendo a escala do país a mais grosseira e a dos edifícios individuais a mais refinada (ver Tabela 1). Os nós têm propriedades que caracterizam sua localização

geografia, dados demográficos da população, intervenções governamentais não farmacêuticas relacionadas ao COVID-19 e padrões de movimento da população com registro de data e hora (consulte a Tabela 2). Ao longo das escalas, as geolocalizações são conectadas por uma divisão administrativa social que denota a relação de contenção hierárquica. Dentro de cada escala, os nós de geolocalização são conectados por dois tipos de transporte, aéreo e terrestre, para possibilitar a consideração de suas diferentes capacidades de movimentação ou dispersão. Devido a questões como redundância de nomes e inconsistência de dados nas balanças, coletar e sincronizar tal conjunto de dados não é fácil. Para permitir a depuração e obter algum entendimento não superficial e demonstração visual do conjunto de dados multiescala, também construímos um painel de visualização interativo (consulte a Figura 1). Desenvolvemos um website para hospedar o conjunto de dados multiescala COVID-19, descrições e painel de visualização.

Em termos de novidades e contribuições, desenvolvemos um conjunto de dados geoespaciais multiescala que interliga localização geográfica, demografia populacional, intervenções governamentais não farmacêuticas relacionadas ao COVID-19 e padrões de movimento populacional com data e hora, todos em diferentes escalas (por exemplo, país, estado/município estado, condado/cidade e edifícios individuais). Aqui, em particular, gostaríamos de enfatizar os esforços manuais não negligenciáveis para limpar os dados e os desafios que encontramos, como 1) unificar e alinhar dados de várias fontes e 2) formular o

¹Um [On-line]. Disponível: <https://usmannn.github.io/COVID-19-Multiscale-Simulation/>

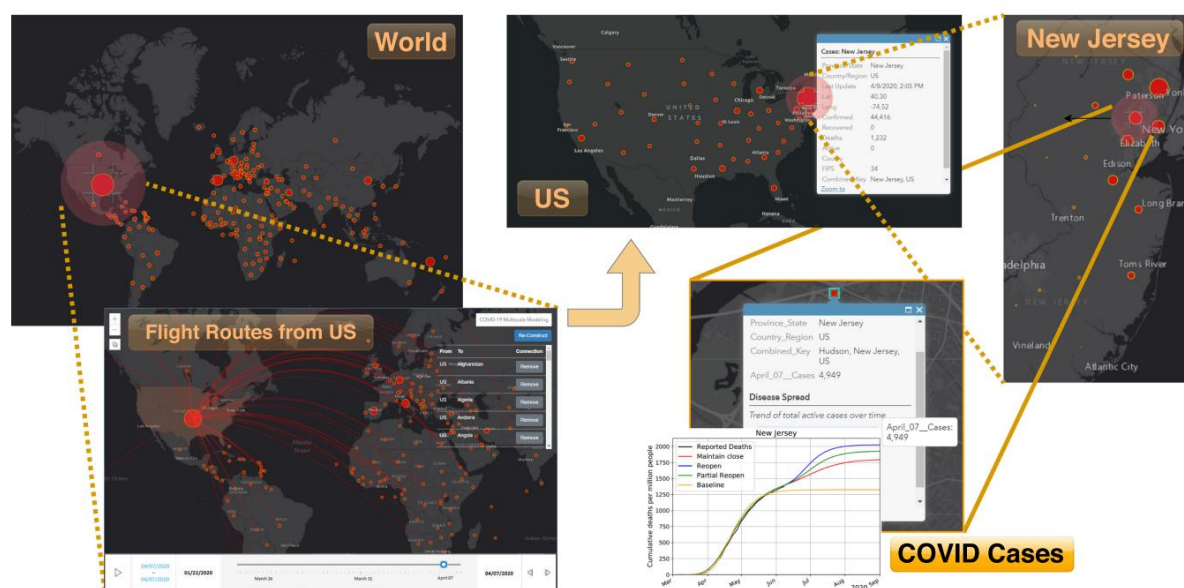


FIGURA 1. Protótipo preliminar de plataforma interativa multiescala com visualização da propagação da infecção em Nova Jersey. Os diferentes componentes na figura refletem a interatividade do usuário com o sistema. Os círculos vermelhos representam nós diferentes, enquanto as linhas representam as conexões entre os nós da rede. O tamanho dos discos está relacionado à exposição a doenças nesses nós.

novo multigrafo multiescala direcionado heterogêneo, com propriedades estáticas e dinâmicas, ou seja, informações limpas e consistentes (por exemplo, questões como ambiguidade e redundância são resolvidas com os melhores esforços manuais). Também desenvolvemos um painel de visualização interativo. Nossa plataforma permite que os usuários visualizem a propagação da doença em diferentes escalas (por exemplo, desde o nível do país até bairros individuais). Além disso, permite que os usuários interajam com as políticas aplicadas em diferentes escalas na rede (por exemplo, o fechamento de fronteiras e bloqueios nas comunidades) para observar seu impacto na epidemiologia.

Além disso, oferecemos análise baseada em correlação para examinar nosso conjunto de dados. Alguns dos insights mais importantes da análise de dados incluem o seguinte.

- Os padrões de correlação diferem distintamente entre as escalas (por exemplo, a escala 3 pode ser muito diferente das escalas 1 e 2 – veja as descrições das escalas na Tabela 1), indicando a necessidade de explorar estratégias de modelagem únicas para cada escala e a possibilidade de alcançar o desempenho ganhos encorajando a troca de informações entre escalas.
- Um modelo não linear pode capturar melhor os padrões ocultos em nosso conjunto de dados. Existem altas relações não lineares entre a população, os recursos relacionados ao voo e o horário de chegada do COVID-19.

- Quanto mais tarde for a data de implementação do regime de permanência em casa, maior será a proporção de infectados observada no nível estadual dos EUA.
- As séries temporais relacionadas ao COVID-19 e às tendências de mobilidade compartilham informações mútuas elevadas, indicando relacionamentos não independentes.

O restante deste artigo está organizado da seguinte forma. Revemos trabalhos relacionados e fornecemos uma visão geral da epidemiologia computacional e dos painéis de visualização na seção “Trabalhos relacionados”. As descrições do conjunto de dados estão na seção “Conjunto de dados geoespaciais multiescala”. Fornecemos um conjunto de análise de dados, como comparação de correlação de recursos multiescala, na seção “Análise de dados”. O serviço de dados interativo baseado na web (painel de visualização) é descrito na seção “Serviço de visualização de dados interativos”. Os detalhes da linguagem de implementação são apresentados na seção “Detalhes da implementação”. Destacamos a aplicabilidade de nosso conjunto de dados para futuros pesquisadores na seção “Aplicabilidade de dados”. Por fim, a seção “Conclusão” encerra este artigo.

TRABALHO RELATADO

Nesta seção, apresentamos um breve histórico sobre epidemiologia computacional e modelos multiescala de doenças infecciosas, seguido da revisão da literatura sobre o

Sistemas de Visualização COVID-19 (consulte a seção “Sistemas de Visualização COVID-19”) e sobre a compreensão dos efeitos das intervenções não farmacêuticas para limitar a propagação do COVID-19 (consulte a seção “Entendendo o Efeito das Intervenções Não Farmacêuticas”).

A epidemiologia computacional é um campo multidisciplinar que visa entender melhor as questões centrais da epidemiologia, como a disseminação de doenças ou a eficácia de uma intervenção de saúde pública. As técnicas anteriores incluem modelos matemáticos de equações diferenciais baseadas em taxas, como o popular modelo suscetível infectado-recuperado (SIR)⁵ e suas variantes,⁶ modelagem baseada em agentes⁷ e outros.⁸ Na área de modelos multiescala de doenças infecciosas, um grande número de artigos publicados enfocou a modelagem da dinâmica de doenças infecciosas e não aborda as escalas multiespaciais desses sistemas de doenças.⁹ Wang et al.¹⁰ consideraram a previsão espaço-temporal multirresolução. No entanto, os autores consideram apenas a escala do condado e do estado com apenas dois estados nos Estados Unidos.

Sistemas de visualização do COVID-19 Vários esforços foram feitos para visualizar o início da pandemia do COVID-19. Uma interface interativa é apresentada para rastrear casos de COVID-19 em todo o mundo em tempo real.³ Yang et al.¹¹ propuseram um sistema de análise visual em toda a cidade para simular a mobilidade humana e observar o status da infecção por COVID-19 em resposta. Em outro sistema, Samant et al.¹² apresentaram uma rede multicamadas como um painel interativo para visualizar vários aspectos da COVID-19 nos Estados Unidos. Literatura adicional sobre plataformas de visualização e o que a solução de visualização interativa proposta oferece de diferente é apresentada em detalhes na seção “Serviço de Visualização Interativa de Dados”.

Compreendendo o efeito das intervenções não farmacêuticas À medida que medidas especiais de controle são feitas em nível global, nacional, estadual e municipal, os pesquisadores estão tentando entender seu efeito na disseminação do COVID-19. Por exemplo, Chinazzi et al.¹³ usaram um modelo de transmissão de doenças metapopulacionais globais para projetar o impacto das limitações de viagens na disseminação nacional e internacional do COVID-19.

Seus resultados de modelagem indicam que as restrições sustentadas de viagens de 90% de e para a China continental afetam apenas modestamente a trajetória da epidemia, a menos que combinadas com uma redução de 50% ou mais da transmissão na comunidade. Adiga et al.¹⁴ avaliaram o impacto das suspensões de companhias aéreas internacionais na disseminação global precoce do COVID-19. Wilder-Smith e Freedman¹⁵ analisou o papel fundamental de intervenções, como

isolamento, quarentena, distanciamento social e contenção comunitária, para medidas de saúde pública de estilo antigo no surto de COVID-19. Percepções diretas e bem pensadas devem ser extraídas de maneira mais oportuna para os formuladores de políticas. Além disso, não está claro como cada medida de controle única, tomada em diferentes escalas, variando de individual a global, influencia as doenças infecciosas.

GEOESPACIAL MULTIESCALA

CONJUNTO DE DADOS

Nesta seção, descrevemos o conjunto de dados geoespaciais multiescala proposto. Apresentamos primeiro a visão geral do conjunto de dados na seção “Sistemas de visualização COVID-19”. O processo de coleta de dados e a descrição detalhada dos dados são descritos na seção “Entendendo o efeito das intervenções não farmacêuticas”.

Visão geral dos dados

Partindo do entendimento geoespacial multiescala, consideramos cada geolocalização em diferentes escalas como um nó no conjunto de dados proposto. No geral, quatro escalas são especificamente definidas, desde a escala de país mais grosseira até a escala de construção funcional mais refinada. Isso ocorre porque as implicações geográficas e sociais únicas dos locais, dentro ou entre escalas, podem levar a resultados ou impactos drasticamente diferentes de uma doença infecciosa. A Tabela 1 fornece a descrição e as estatísticas das geolocalizações/nós em cada escala. No total, o conjunto de dados contém informações sobre 11.961 geolocalizações/nós.

Para resumir as informações fornecidas para os nós: todos os nós têm propriedades de descrição de dados associadas que caracterizam sua localização geográfica. Propriedades, como dados demográficos da população, são fornecidas para uma parte dos nós. Para apoiar o estudo de caso, as intervenções governamentais não farmacêuticas relacionadas ao COVID-19 e os padrões de movimento durante esse período de tempo para os nós também são fornecidos como propriedades de nós variáveis no tempo. Consulte a Tabela 2 para obter uma enumeração dos tipos e nomes das propriedades do nó fornecidas no conjunto de dados.

Intuitivamente, os nós de geolocalização são conectados pela divisão administrativa entre as escalas, denotando a relação de contenção hierárquica. Por exemplo, o nó “US” de escala 1 contém (é conectado por) 56 nós de nível de estado de escala 2 (ou equivalente). Dentro de cada escala, os nós de geolocalização são conectados por diferentes tipos de transporte, especificamente, aéreo, terrestre e marítimo. Consulte a Tabela 3 para conexões de nó que conectam nós em cada escala dada.

TABELA 3. Descrição da conexão do nó de geolocalização

dentro de uma determinada escala.

Índice	Categoria	Nome	Atribuído
5	Ar	Voo internacional	Verdadeiro
6	Ar	Voo doméstico	Verdadeiro
7	Terra	borda de compartilhamento	Falso

Como resultado, nosso conjunto de dados formula um multigrafo multiescalab com 1) tipos heterogêneos de nós e arestas bidirecionais e 2) atributos estáticos e dinâmicos de nós e arestas.

Os principais estágios de pré-processamento de dados são os seguintes:

- 1) coleta das fontes de informação disponíveis; 2) identificando todas as geolocalizações de todas as escalas de todas as fontes;
- 3) formar os nós do grafo removendo duplicatas de geolocalização e resolvendo problemas, como ambigüidade e redundância de nomenclatura; 4) adicionando conexões de nós (bordas de voo e bordas de compartilhamento);
- 5) preencher atributos de nó e aresta.

Mantivemos propositalmente o foco de nossa coleta de dados para 2020, conhecido por ser o auge da pandemia de COVID-19, com um conjunto diversificado de intervenções demográficas e de políticas de viagens em diferentes países. Isso não torna nossa solução multimodelo (como em nossa sugestão discutida na seção "Conclusão" orientada por nossa análise de dados) ou a plataforma de visualização inutilizável por um período ou ano diferente.

O conjunto de dados coletados, incluindo a visualização, pode ser encontrado em nosso painel baseado na Web, e o painel será descrito na seção "Serviço de visualização de dados interativos".

Coleta e descrição de dados Nesta seção, descrevemos as fontes de dados, nosso método de processamento de dados, bem como qualquer suposição feita.^c

^b Na teoria dos grafos, um multigrafo é um grafo que pode ter múltiplas arestas, ou seja, arestas que possuem os mesmos nós finais. ^c Os dados de infecção por COVID-19 são obtidos do Corona virus Visual Dashboard operado pela Johns Hopkins University.³

Identidade de geolocalização e conexão por escala Os tipos de geolocalização são demonstrados na Tabela 1.

Os países são identificados por seu código ISO2 exclusivo e os aeroportos são identificados pelo código IATA exclusivo.

As escalas no nome da geolocalização são separadas por "/" para evitar ambigüidade, pois o nó filho de dois países pode ter o mesmo nome (por exemplo, "AU/New South Wales/Bathurst/Bathurst Airport" e "CA/New Brunswick/Bathurst/Aeroporto de Bathurst"). Uma vez que uma geolocalização pode ter vários aliases de nomes, a redundância de nó foi resolvida ao máximo (por exemplo, "CN/Macau" e "CN/Macau" são redundantes. Em nosso conjunto de dados, apenas "CN/Macau" pode ser encontrado).

Para os Estados Unidos, temos as informações mais refinadas. No total, temos 56 locais de escala 2 para os EUA: 50 estados dos EUA mais seis casos especiais: Distrito de Columbia (um distrito federal), Samoa Americana, Guam, Ilhas Marianas do Norte, Ilhas Virgens e Porto Rico. Vale a pena mencionar que a cidade de Nova York contém na verdade cinco condados: Condado de Nova York (Manhattan), Condado de Kings (Brooklyn), Condado de Bronx (The Bronx), Condado de Richmond (Staten Island) e Condado de Queens (Queens). Os cinco condados infantis da cidade de Nova York não estão incluídos em nosso conjunto de dados porque as características ou dinâmicas da cidade de Nova York receberam mais atenção. Quaisquer propriedades do nó "US/New York/New York City" é uma agregação de seus cinco condados. Se você quiser reduzir o espaço, na minha opinião, a descrição acima relacionada à "cidade de Nova York" pode ser excluída, ou seja, de "Vale a pena mencionar isso" para "uma agregação de seus cinco condados".

Identificamos os nós de geolocalização e seus atributos geográficos básicos, ou seja, latitude e longitude, utilizando todos os recursos de dados disponíveis da propriedade de geolocalização ou conexão (mais detalhes nas subseções a seguir). Por exemplo, os nós do aeroporto são obtidos dos dados de voo e os pais multiescala de um aeroporto são obtidos dos recursos de dados da propriedade do aeroporto.¹⁶

Conexão de geolocalização por via aérea

Conforme demonstrado na Tabela 3, utilizamos dados de voos internacionais e dados de voos domésticos de condados para conectar localizações geográficas em uma determinada escala por via aérea. Obtemos os dados de voo do OpenFlight16 e do Bureau of Transportation Statistics (BTS) do Departamento de Transportes dos EUA.¹⁷ Cada voo é uma aresta direcionada conectando dois aeroportos, por exemplo, a aresta "(CN/Guangdong/Guangzhou/CAN, EUA/California/Los Angeles/LAX)" denota a conexão aérea do aeroporto "CN/Guangdong/Guangzhou/CAN" para o aeroporto "US/California/Los Angeles/LAX". Também denota a conexão de ar de "CN/Guangdong/

Guangzhou" para "EUA/Califórnia/Los Angeles" na escala 3, "CN/Guangdong" para "EUA/Califórnia" na escala 2 e "CN" para "EUA" na escala 1. Essa característica é uma das razões pelas quais o gráfico formulado por nosso conjunto de dados é um multigrafo.

Conexão de geolocalização por terra A borda, conexão de geolocalização por terra, indica que as duas geolocalizações estão compartilhando fronteiras. Nós fornecemos limites de fronteira que conectam países em todo o mundo, limites de fronteira que conectam os nós da escala 2 dos EUA (ou seja, estados) e limites de fronteira que conectam os condados dos EUA. As bordas das bordas também são bidirecionais porque acreditamos que a capacidade das bordas das bordas são assimétricas. Observe que é garantido que um nó de geolocalização não possui loops automáticos (ou seja, não possui uma aresta apontando para si mesmo).

Número diário de passageiros para voos Para entender como os voos afetam o movimento da população entre as regiões, obtemos o número diário de passageiros para cada par de portos aéreos que possuem voos. Para isso, são utilizados os dados do segmento T-100 (todas as transportadoras) do BTS (Departamento de Transportes dos Estados Unidos)¹⁷. O BTS fornece anualmente dados do segmento T-100 que contém dados de segmento sem escalas por tipo de aeronave, passageiros transportados, carga transportada e correio em libras, capacidade disponível (ou seja, assentos), informações do aeroporto de origem, informações do porto aéreo de destino, horas da aeronave e distância, até a escala mensal máxima. Para a simulação realista, dados muito recentes são preferíveis, mas os dados de março de 2019 são escolhidos e março de 2020 é estimado como o mesmo de março de 2019 porque os dados de 2020 não foram fornecidos quando nossos dados foram coletados pela primeira vez e o vírus teve um sério impacto sobre os EUA durante março de 2020. Dos dados fornecidos, apenas o código do aeroporto de origem, o código do aeroporto de destino e o número do passageiro são usados, e os dados mensais dos passageiros transportados são divididos por 30 para estimar o número diário de passageiros para cada par de aeroportos que possuem voos.

Como o número de passageiros fornecido para voos específicos para transporte de carga e correio (16,5% voos) é 0, adicionamos 1 para esses voos para ponderar os voos adequadamente. Resumindo, esses dados indicam, em dias normais, a média diária de passageiros para cada par de aeroportos que possuem voos.

Geolocalização População Os dados populacionais da região são do US Census Bureau,¹⁸ Nações Unidas (ONU), Departamento de Assuntos Econômicos e Sociais e Divisão de População. O US Census Bureau fornece dados estatísticos do censo todos os anos

para a população em nível de cidade para os Estados Unidos. Os dados originais contêm dados do censo de 2010 e população estimada de 2011 a 2018. A população estimada de 2018 é selecionada para simulação. Para as populações de outros países, são utilizados os dados de perspectiva de população mundial de 2019 da ONU. Isso fornece a estimativa da população mundial pela seguinte base de 1950 a 2020: localização, ano civil único, idade de um ano, faixa etária de cinco anos e sexo.

Intervenção não farmacêutica governamental geolocalizada no COVID-19 Para demonstrar um exemplo de aplicações

interessantes que nosso conjunto de dados pode fornecer, coletamos dados sobre a política de "ficar em casa" que os governos estaduais dos EUA emitiram para lidar com a pandemia do COVID-19 com o período de tempo efetivo correspondente (data inicial e data final). Obtemos os dados do The New York Times. Uma descrição de texto da política, bem como seu link da web, também são fornecidos em nosso conjunto de dados. Observe que nem todos os estados têm data de início e data de término. Alguns estados têm apenas uma data de início, mas nenhuma data final foi divulgada até o ponto da primeira coleta de nosso conjunto de dados.

Padrão de movimento de geolocalização

Para demonstrar a capacidade de movimento das geolocalizações, enriquecemos nosso conjunto de dados com o Google COVID-19 Community Mobility Reports.¹⁹ Isso fornece padrões de tendência de movimento variáveis no tempo para algumas geolocalizações (começando em "15 de fevereiro de 2020"). Cada tendência de mobilidade é um conjunto de séries temporais apresentadas por uma comunidade que é classificada pelas funções sociais do espaço (ou seja, lugares, como mercearias e parques, dentro de uma área geográfica) e destaca a variação percentual nas visitas das pessoas. Especificamente, esses dados mostram como as visitas e o tempo de permanência em diferentes lugares mudam em comparação com os dias normais. O Google calculou essas alterações com base nos dados do mapa do Google de usuários que ativaram o Histórico de localização para sua conta do Google. As seis funções sociais são as seguintes:

- 1) mercearia e farmácia—lugares como mercearias, armazéns de alimentos, drogarias e farmácias; 2) parques—lugares como parques, praias públicas, marinas, praças e jardins públicos; 3) estações de trânsito—locais como centros de transporte público, como estações de metrô, ônibus e trem;
- 4) varejo e recreação – lugares como restaurantes, cafés, shopping centers, museus, bibliotecas e teatros;
- 5) residencial—lugares de residência; 6) locais de trabalho—locais de trabalho.

TABELA 4. Recursos calculados a partir das propriedades estáticas dos nós em cada escala para a análise de correlação de Pearson e Spearman.

Nome	Descrição
POP	O tamanho da população.
NUM_FLI	O número de arestas de voo conectadas ao nó.
WEI_FLI	O peso total (número diário de passageiros) nas bordas do voo que estão conectadas ao nó.
NUM_NEI	O número de nós vizinhos adjacentes à borda na mesma escala.
NUM_NEIF	O número total de voos que os nós vizinhos adjacentes à borda na mesma escala possuem.
DIS_CNA	A distância euclidiana das coordenadas (latitude e longitude) deste nó até as coordenadas da China.
FLI_CNA	Se este nodo tem voo direto da China.
ARRV	A hora de chegada do COVID-19 a este nó, calculada como o número de dias desde 31 de dezembro de 2020, que é a data oficial em que a China anunciou o primeiro caso de COVID-19.

ANÁLISE DE DADOS

Examinamos nossos dados e realizamos a seguinte análise:

1) comparação de correlação de recursos multiescala nas propriedades estáticas do nó (consulte a seção “Análise de correlação multiescala”), 2) visualização da relação entre a implementação da política “ficar em casa” data de início em relação ao número de pessoas infectadas naquele dia (consulte a seção “Análise do pedido de permanência em casa”) e 3) Análise das propriedades do nó dinâmico do ponto de vista da teoria da informação (consulte a seção “Nó dinâmico seção “Análise de propriedade”).

Análise de correlação multiescala Vários tipos de recursos podem ser calculados a partir de nosso conjunto de dados. Inspirados em Adiga et al.14, calculamos as características com base nas propriedades estáticas do nó, conforme apresentado na Tabela 4 da escala 1 a 3, e conduzimos a análise de correlação de Pearson na Figura 2 e a análise de correlação de Spearman na Figura 3.

Pela Figura 2, pudemos observar o seguinte.

- 1) O número de voos e os pesos dos voos são altamente e linearmente correlacionados em todas as escalas, com coeficiente de correlação de Pearson (r) acima de 0,92.
- 2) Na escala 1, o horário de chegada do COVID-19 está negativamente correlacionado linearmente com o fato de o país ter voo direto da China, com $r = 0,54$.
- 3) Na escala 2, a população do estado é alta e positivamente correlacionada linearmente com o número de voos ($r = 0,86$) e os pesos totais dos voos ($r = 0,92$). A população do estado também está positivamente correlacionada linearmente com o fato de o estado ter voos diretos da China ($r = 0,62$). Além disso, se um estado tem voo direto da China está positiva e linearmente correlacionado com o peso total do voo desse estado ($r = 0,57$). Por outro lado, o tempo de chegada do COVID-19 de um

estado é negativamente correlacionado com o peso total de voo deste estado ($r = 0,51$).

- 4) Na escala 3, apenas o número de voos e os pesos totais dos voos são altamente correlacionados com a existência de voos diretos da China.
- 5) Quando consideramos todas as três escalas, o padrão de correlação linear é menos óbvio do que olhar para cada escala individualmente, onde a escala de estado, ou seja, a escala 2, tem os padrões de correlação mais óbvios. Curiosamente, quando consideramos todas as escalas juntas, a distância da China torna-se a característica mais correlacionada com a hora de chegada do COVID-19, em vez da população na escala do condado ou da cidade e características relacionadas ao voo (especialmente se houver voo direto) da China) na escala do país e do estado.

Os padrões mostrados na Figura 2 podem ser resumidos da seguinte forma: 1) os padrões de correlação diferem distintamente entre as escalas, o que indica a necessidade de explorar estratégias de modelagem únicas para cada escala e a possibilidade de obter ganhos de desempenho ao permitir que as escalas auxiliem uns aos outros; 2) como um caso de uso potencial de nosso conjunto de dados, a hora de chegada do COVID-19 pode ser previsível a partir de um voo da China na escala do país e pode ser previsível a partir do peso total do voo na escala do estado. Outras características, como a população, são importantes e podem desempenhar papéis em diferentes níveis de importância em cada uma das escalas.

A correlação de Pearson indica o linear oculto relação entre as feições, enquanto a correlação de Spearman indica a existência de relação não linear. Em comparação com a correlação de Pearson, os coeficientes de correlação de Spearman são geralmente mais altos (consulte a Figura 3), indicando que um modelo não linear pode ser capaz de capturar melhor os padrões ocultos em nosso conjunto de dados. O número de voos e os pesos dos voos são

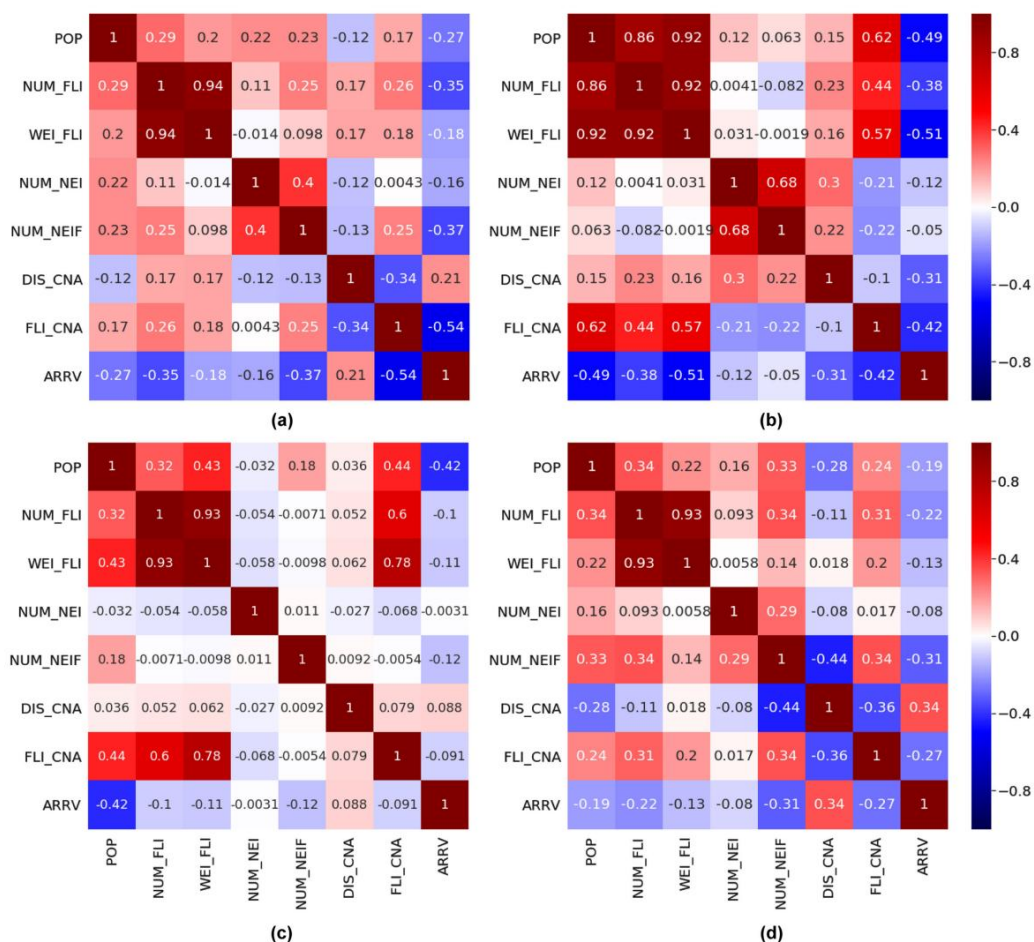


FIGURA 2. Análise de correlação de Pearson em recursos extraídos das propriedades estáticas fornecidas por nosso conjunto de dados. (a) Escala 1: País. (b) Escala 2: Estadual. (c) Escala 3: Concelho. (d) Escalas 1–3.

ainda altamente correlacionados em todas as escalas, no entanto, ao usar a correlação de Spearman.

- 1) Na escala 1, a população está altamente correlacionada com o número de voos deste país e o número de vizinhos adjacentes à fronteira que o país possui. O número de voos está altamente correlacionado com o fato de o país ter voos da China. A hora de chegada do COVID-19 é altamente negativa e não linearmente correlacionada com o número de voos e o número total de voos que o vizinho adjacente à fronteira deste país possui.
- 2) Na escala 2, o horário de chegada do COVID-19 está positivamente correlacionado com a população, altamente e negativamente correlacionado com as características relacionadas ao voo (número de voos e peso total do voo). A população do estado também está alta e positivamente correlacionada com características relacionadas ao voo. Além disso, o número de vizinhos adjacentes à fronteira deste país possui.

altamente correlacionados positivamente com o total de voos que esses vizinhos têm.

- 3) O padrão de correlação é muito menos óbvio na escala 3. Somente a população está fortemente correlacionada negativamente com a hora de chegada do COVID-19 (coeficiente de correlação de 0,72).

Em resumo: 1) os padrões de correlação ainda diferem distintamente entre as escalas, e a escala 3 é muito diferente das outras duas escalas, e 2) existem altas relações não lineares entre população, características relacionadas ao voo e a hora de chegada do COVID -19.

Análise do pedido de permanência em casa
Com base nos recursos da Tabela 5 que calculamos a partir de nosso conjunto de dados coletados, realizamos uma análise de permanência em casa na Figura 5. Especificamente,

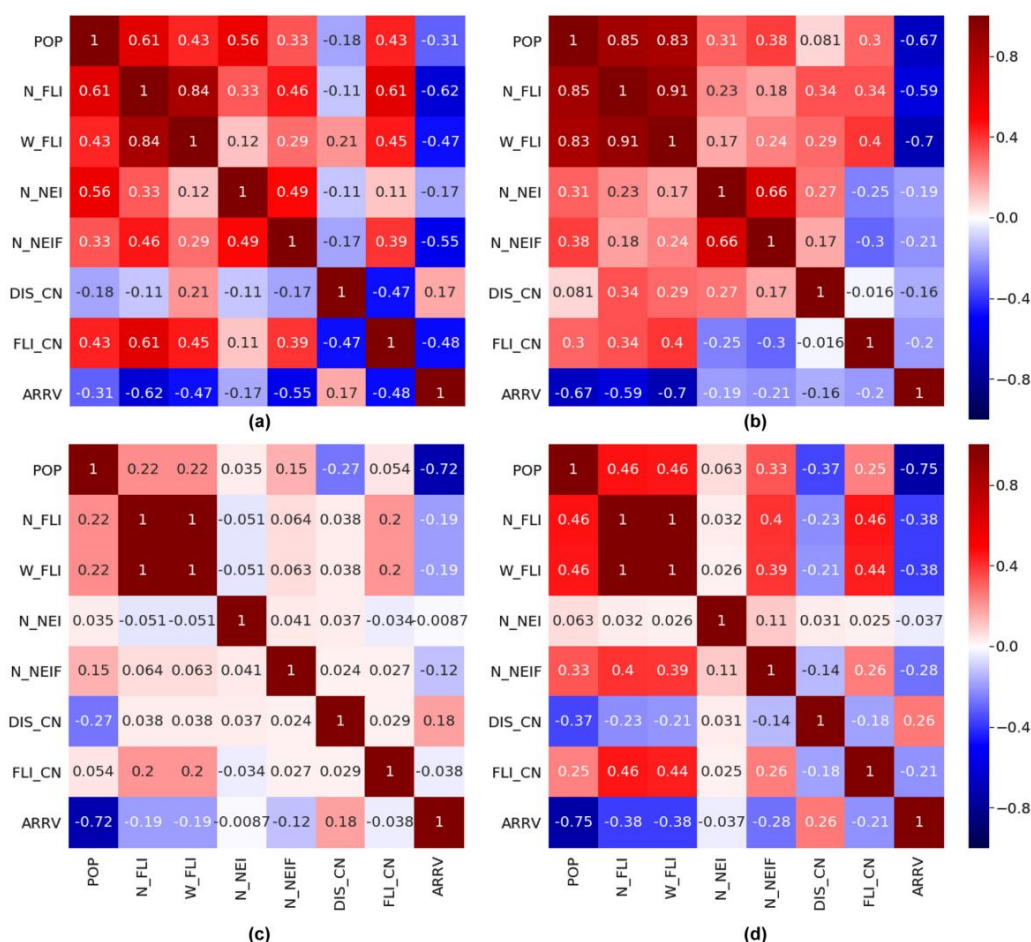


FIGURA 3. Análise do coeficiente de correlação de Spearman em características extraídas das propriedades estáticas fornecidas pelo nosso conjunto de dados. (a) Escala 1: País. (b) Escala 2: Estadual. (c) Escala 3: Concelho. (d) Escalas 1-3.

plotamos a relação entre a data de início da implementação da política de permanência em casa (eixo x) em relação à proporção da população infectada com COVID-19 naquele dia (eixo y) no nível estadual dos EUA na Figura 5(a). Como mostra a figura, quanto mais tarde a data de implementação, maior a taxa de infectados pode ser observada, exceto para Nova York e o Distrito de Columbia como dois outliers, o que sugere alta taxa anormal de infecção para esses dois estados. Na Figura 5(b), a análise de correlação de Pearson é realizada com as características da Tabela 5; enquanto isso, na Figura 5(c), a análise de correlação de Spearman é realizada. Podemos observar correlação alta e positiva entre a população e o número de infectados no estado. Correlação alta e positiva também pode ser observada entre infectados

proporção e o número de pessoas infectadas, bem como a proporção de infectados e os dias desde que a China anunciou o primeiro caso de COVID-19 no dia da implementação da política de permanência em casa.

Análise dinâmica de propriedades de nós

Calculamos e plotamos as informações mútuas na Figura 6(a) e a entropia condicional na Figura 6(b) entre oito séries temporais de diferentes tipos de propriedades dinâmicas de nós fornecidas por nosso conjunto de dados coletado. Entre as oito séries temporais, seis são tendências de mobilidade classificadas por funções sociais do espaço e duas são séries temporais relacionadas ao COVID-19, ou seja, o número diário de infectados acumulados e o número diário de infectados líquidos. Em ambos os gráficos, o eixo diagonal mostra a entropia da série temporal correspondente. Para interpretar os plots, por exemplo, a entropia de "Varejo" é 5,3, conforme mostra a Figura 6(a) e (b). A informação mútua entre "Varejo" e "Mercearia" é de 3,8, conforme mostra a Figura 6(a). A entropia condicional, "Varejo" condicionada a "Mercearia" é 1,5. A entropia de "Varejo" é a soma da informação mútua entre "Varejo" e "Mercearia" e a entropia condicional, ou seja, "Varejo" condicionada a "Mercearia" (basicamente $H(X|Y)$).

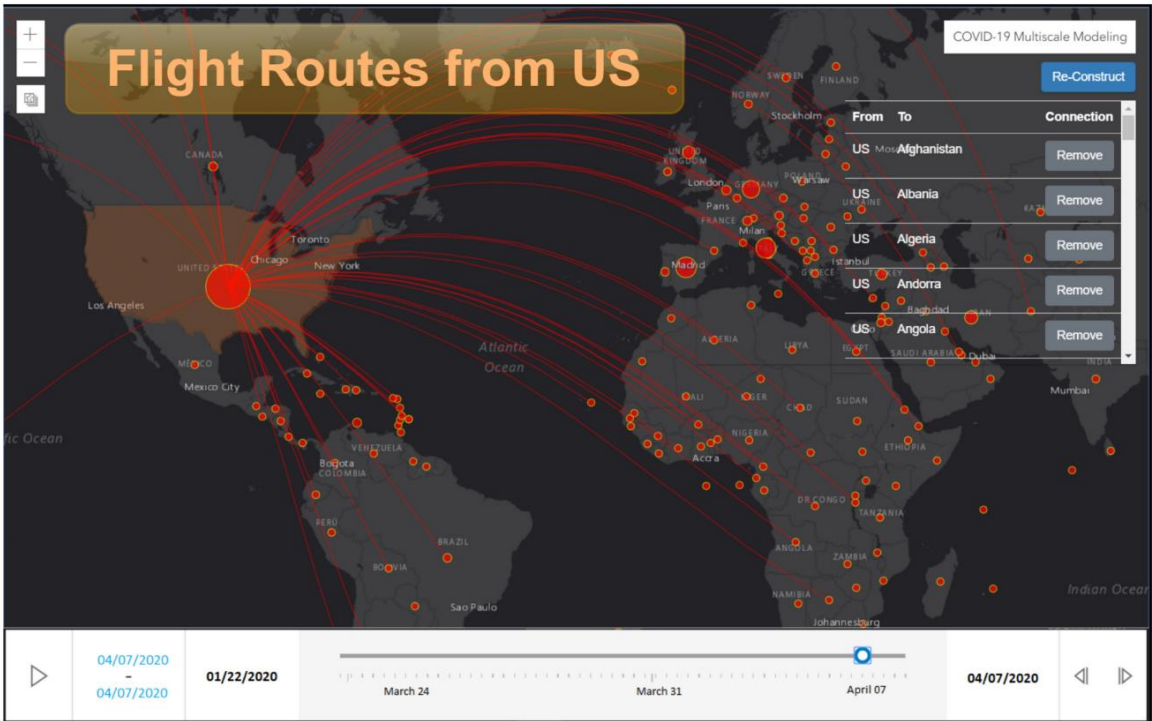


FIGURA 4. Rotas de voos de ida do nó dos EUA para outros países. A área de rolagem no lado direito mostra em pares ligações aéreas entre os EUA e outros países. Na frente de cada linha, há um botão para remover aquela conexão específica do gráfico multiescala. Os discos circulares vermelhos são os nós com tamanhos variados que refletem o número de casos ativos de COVID-19 naquele nó (ou seja, quanto maior o tamanho do disco, maior o número de casos ativos de COVID-19). O controle deslizante na parte inferior permite mover entre os dias para visualizar a disseminação do COVID-19 para um nó específico ou todos os nós.

TABELA 5. Características que são calculadas para conduzir a análise de correlação de Pearson e Spearman na ordem “ficar em casa”.

Nome	Descrição
Infetado	O número de pessoas infectadas com COVID-19 no dia de implementação da política de permanência em casa.
População	O tamanho da população.
Delta_dias	No dia de implementação da política de permanência em casa, o número de dias desde 31 de dezembro de 2020, que é a data oficial em que a China anunciou o primeiro caso de COVID-19.
Infected_Ratio	Infected dividido pela população.

Informação mútua é uma quantidade que mede uma relação entre duas variáveis aleatórias que são amostradas simultaneamente. Em particular, mede quanta informação é comunicada, em média, em uma variável aleatória sobre outra. Alta informação mútua indica uma grande redução na incerteza, baixa informação mútua indica uma pequena redução e zero informação mútua entre duas variáveis aleatórias significa que as variáveis são independentes. Como podemos ver nos gráficos, as oito séries temporais compartilham alta informação mútua, indicando relacionamentos não independentes.

DADOS INTERATIVOS
SERVIÇO DE VISUALIZAÇÃO

Apresentamos um painel interativo baseado na Web para visualização e download do conjunto de dados geoespaciais multiescala (visite este site para obter mais detalhes). A plataforma permite aos usuários expandir para diferentes níveis de abstração das escalas de nível de continente, país,

^d[On-line]. Disponível: <https://usmannn.github.io/COVID-19-Multiscale-Simulation/>

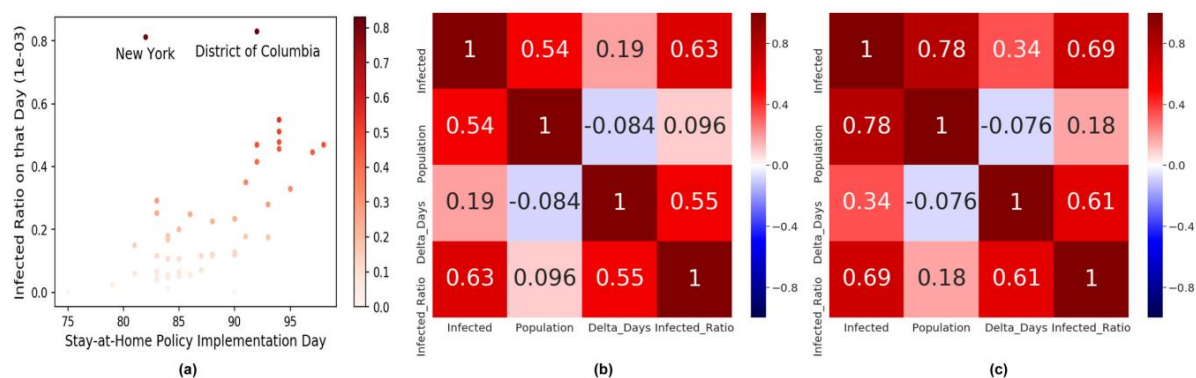


FIGURA 5. Análise do pedido "Fique em casa". (a) Taxa de permanência em casa versus proporção de infectados. (b) Correlação de Pearson. (c) Correlação de Spearman.

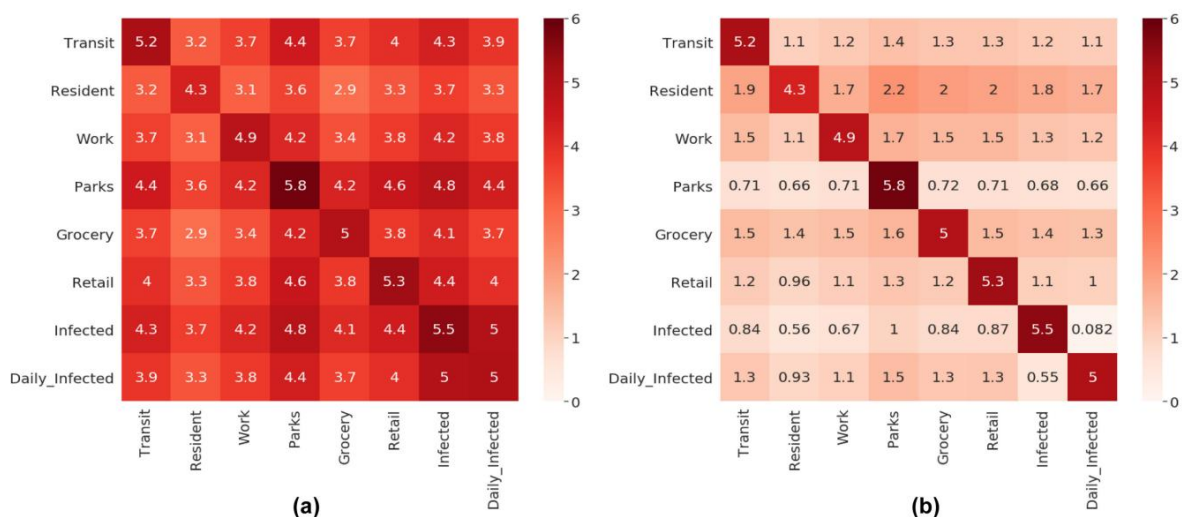


FIGURA 6. (a) Informação mútua e (b) entropia condicional entre as oito séries temporais de diferentes tipos de dinâmica de nós propriedades.

idades e províncias para bairros individuais, para visualizar interativamente o impacto do COVID-19.

Esforços semelhantes foram relatados na literatura. Por exemplo, um extenso repositório de visualização de dados é apresentado para análise exploratória de dados da pandemia de COVID-19.²⁰ Um painel interativo é desenvolvido para rastrear o COVID-19 nos estados, condados e cidades dos EUA.²¹ Esses trabalhos, no entanto, geralmente por forma em um único nível de abstração (por exemplo, escala do país), e não leva em conta nenhum dado de mobilidade para modelar o fluxo de informação entre e dentro de diferentes escalas, e conta apenas com os valores estimados para incorporar implicitamente a transmissão. Eles também restringem os usuários a interagir com diferentes políticas integradas na modelagem da propagação da doença para visualizar como as diferentes políticas podem ou não impactar a epidemiologia.

Em contraste, nossa plataforma permite que os usuários interajam com diferentes políticas (por exemplo, fechamento de fronteiras e bloqueios), que são modeladas na rede multiescala, para visualizar seu impacto na disseminação ou controle da doença. Mais especificamente, a plataforma permite:

- 1) adicionar/remover conexões entre diferentes nós (por exemplo, arestas na rede multiescala);
- 2) adicionar/remover certas geolocalizações (por exemplo, nós em a rede multiescala);
- 3) atualizar a taxa de transmissão de doenças em cada conexão (por exemplo, a largura das arestas entre os nós);
- 4) visualizar tendências epidemiológicas temporais;
- 5) mover-se entre diferentes níveis de abstração (por exemplo, para diferentes escalas - país, estado, cidade,

etc.) visualizar como uma política afeta a propagação ou controle da doença em diferentes escalas;

- 6) baixar o conjunto de dados geoespaciais multiescala da epidemiologia para as políticas exploradas pelo usuário.

A Figura 1 mostra uma visão geral de nossa plataforma multiescala interativa baseada na Web. As diferentes imagens da figura são resultado da interatividade do usuário com o sistema. No início (por exemplo, exibição padrão), a plataforma mostra todos os nós da escala do país. Um usuário pode então selecionar interativamente um ou mais nós para visualizar suas conexões com outros nós, atualizar a taxa de transmissão da doença nas bordas selecionadas, adicionar ou remover os nós e/ou bordas da rede multiescala e pode reconstruir o gráfico com base sobre as últimas mudanças para observar seu impacto na propagação da doença. Além disso, o usuário pode se mover entre as escalas (por exemplo, do nível do país para os estados com o país) para visualizar como uma mudança de política feita em uma escala maior pode ter impactado as tendências (por exemplo, propagação da doença) na parte inferior escalas.

A Figura 4 mostra um caso de uso específico em que um usuário passou o mouse e clicou no nó dos EUA. Ao fazer isso, a plataforma de visualização mostrou as rotas ativas de voos de ida dos EUA para outros países em diferentes continentes. Para cada rota ativa (ou seja, aresta no gráfico multiescala), o painel permite que o usuário a remova (por exemplo, imponha uma política de fechamento de fronteira para um determinado país) e depois reconstrua o gráfico multiescala passando o mouse e clicando no botão reconstruir. Isso agora permitiria ao usuário observar as implicações após a aplicação, por exemplo, da política de fechamento de fronteiras para um ou mais países sobre a disseminação do COVID-19. O controle deslizante na parte inferior permite que o usuário vá e volte entre dias e meses para visualizar a disseminação do COVID-19 ao longo do tempo, com ou sem a presença de quaisquer políticas aplicadas.

DETALHES DE IMPLEMENTAÇÃO

Selecionamos o Python como a principal linguagem de implementação para processamento e análise de dados por causa de suas vantagens, como versatilidade, extensibilidade e acessibilidade conveniente. O HTML5 foi usado para o painel de visualização interativa. Os pacotes Pandas e Numpy foram usados para processar arquivos de dados (CSV e JSON), e Scipy e Numpy foram usados para análise de dados. O recurso computacional que utilizamos para a coleta e análise do conjunto de dados, bem como para o desenvolvimento do dashboard, é uma única máquina com 256 GB de memória e 48 núcleos de CPU a 2,30 GHz.

APLICABILIDADE DOS DADOS

Nesta seção, descrevemos algumas aplicações e questões de pesquisa exemplares (RQ) que futuros pesquisadores podem conduzir e responder usando nosso conjunto de dados.

RQ 1. Modelo de “propagação” entre regiões: Nosso conjunto de dados pode ser usado para pesquisa científica para modelar e simular “propagação” entre regiões em qualquer nível. Em última análise, a “propagação” é afetada pela ação (por exemplo, movimento) das pessoas. Por exemplo, juntando nosso conjunto de dados com os dados dos casos de COVID-19, “propagação” seria a propagação da doença. O grafo geoespacial multiescala que nosso conjunto de dados fornece pode ser considerado como entradas estruturais adicionais para previsão de propagação. O RQ é, como o movimento entre as regiões afetaria a propagação da doença em todas as regiões?

RQ 2. Identifique agrupamentos populacionais interessantes: Como nosso conjunto de dados fornece informações demográficas populacionais ricas, os pesquisadores podem usar nosso conjunto de dados para identificar agrupamentos populacionais interessantes que podem revelar alguns fatos ocultos, mas úteis. Por exemplo, uma região com população idosa acima da média pode estar sob alto risco de COVID-19.

RQ 3. Prever as necessidades de internação de uma geolocalização: Pode-se levantar o RQ, como e como as propriedades de uma região determinam as necessidades de atendimento médico-hospitalar.

RQ 4. Entenda o impacto multiescalar da política: E se a cidade de Nova York proibisse estritamente as viagens aéreas, e como isso afetaria outras regiões? RQ, como e como a política de uma região afeta outra região dentro ou entre escalas, pode ser levantada.

RQ 5. Compreender o impacto de uma região para outra região: Por exemplo, a tendência de mobilidade do estado de Nova York mostra que as pessoas tendem a ficar mais em casa. Isso afetaria e como isso afetaria a futura tendência de mobilidade da região adjacente do estado de Nova York, como Nova Jersey?

RQ 6. Capture e modele a relação complexa de regiões em escalas: Por exemplo, como o estado de Nova York afetaria um condado em Nova Jersey que está mais próximo de Nova York? Como o número total de casos de COVID-19 nos EUA afetaria um condado no Alasca?

RQ 7. Avaliação para modelo baseado em grafo: Uma vez que nossa funcionalidade de banco de dados fornece um grafo com 1) tipos heterogêneos de nós e arestas bidirecionais e 2) atributos estáticos e dinâmicos de nós e arestas. Novos modelos baseados em grafos, como Graph Convolutional Neural Network, podem ser avaliados em nosso conjunto de dados.

RQ 8. Descubra padrões (geo)espaciais-sociais: uma vez que nosso conjunto de dados fornece um instantâneo da população

paisagem de cada geolocalização, e a região pode ser país, província, condado, cidade e vila, nosso conjunto de dados é útil para aplicativos para descobrir padrões (geo)espaço-sociais. Por exemplo, determinar populações de votos para campanha eleitoral, compreender os fluxos comerciais bilaterais entre regiões, prever a qualidade do ar modelando conjuntamente o impacto da população, planejamento de rotas de transporte, seleção de locais de hotéis, previsão de volume de tráfego móvel entre cidades, etc. que o planejamento de rotas de transporte e a seleção do local do hotel são semelhantes ao Problema do Caixeiro Viajante (TSP). A natureza multiescala do TSP o torna uma tarefa gráfica desafiadora que requer raciocínio sobre as vizinhanças de nós locais e também sobre a estrutura global do gráfico.

Nosso conjunto de dados está de acordo com a natureza multiescala do TSP.

CONCLUSÃO

O surgimento e rápido surto de COVID-19, infelizmente, tornou-se uma grave preocupação global. Os pesquisadores estão respondendo ativamente para combater essa pandemia, dando origem a vários conjuntos de dados e painéis de visualização relacionados. Para abordar a limitação dos recursos existentes e permitir pesquisas mais abrangentes sobre epidemiologia computacional, oferecemos um conjunto de dados geoespaciais multiescala e um painel de visualização interativo. Nossa análise do conjunto de dados geoespaciais selecionados indica a necessidade de modelagem multiescala e fornece informações valiosas, como a existência de altas relações não lineares entre a população, recursos relacionados ao voo e a hora de chegada do COVID-19. Formulamos nosso conjunto de dados como um multigrafo multiescala com tipos heterogêneos de nós e arestas bidirecionais e com atributos estáticos e dinâmicos residindo em ambos os nós e arestas. Devido a essa formulação geral, nosso conjunto de dados tem amplas aplicações. Além disso, para auxiliar a pesquisa em epidemiologia computacional, nosso conjunto de dados considera em conjunto a geografia da localização, demografia da população, intervenção não farmacêutica do governo (ou seja, política de permanência em casa) e padrões de movimento da população com registro de data e hora. Como nosso conjunto de dados fornece um amplo conjunto de informações, certos atributos estão presentes apenas para nós parciais; aumentar o conjunto de dados para resolver essa limitação pode ser um caminho que vale a pena explorar.

Além disso, usar o conjunto de dados coletados para responder a várias perguntas exemplares descritas na seção “Aplicabilidade de dados” pode ser uma direção de trabalho futura interessante.

AGRADECIMENTOS Esta publicação é

baseada no trabalho financiado pela King Fahd University of Petroleum & Minerals (KFUPM).

O(s) autor(es) do KFUPM reconhecem o Centro de Pesquisa Interdisciplinar para Sistemas Seguros Inteligentes pelo

apoio. Este trabalho também foi apoiado em parte pelos prêmios NSF: IIS-1703883, IIS-1955404, IIS-1955365, RETTL 2119265 e EAGER-2122119, e em parte pelos EUA

Departamento de Segurança Interna sob Grant 22STES00001 01 01.

REFERÊNCIAS

1. OMS, “Relatórios de situação da COVID-19”, 2020. Acesso: junho de 2022. [Online]. Disponível: <https://rb.gy/qxukpf>
2. E. Chen, K. Lerman e E. Ferrara, “COVID-19: The first public coronavirus twitter dataset”, 2020, arXiv:2003.07372.
3. E. Dong, H. Du e L. Gardner, “Um painel interativo baseado na Web para rastrear o COVID-19 em tempo real”, *Lancet Infect. Dis.*, vol. 20, não. 5, pp. 533–534, 2020.
4. B. Xu et al., “Dados epidemiológicos do COVID-19 surto, informações de casos em tempo real”, *Sci. Dados*, vol. 7, não. 106, pp. 1–6, 2020.
5. HW Hethcote, “A matemática da infecção doenças”, *SIAM Rev.*, vol. 42, não. 4, pp. 599–653, 2000.
6. CC Ku, T.-C. Ng, e H.-H. Lin, “Epidemiológica benchmarks do controle do surto de COVID-19 na China após o bloqueio de Wuhan: um estudo de modelagem com uma abordagem empírica”, 2020. Acessado em junho de 2022. [Online]. Disponível: <https://ssrn.com/abstract%43544127>
7. P. Patlolla, V. Gunupudi, AR Mikler e RT Jacob, “Ferramentas de simulação baseadas em agentes em epidemiologia computacional”, em *Proc. Int. Workshop Inov. Internet Community Syst.*, 2004, pp. 212–223.
8. PF Gorder, “Epidemiologia computacional,” *Comput. Sci. Eng.*, vol. 12, não. 1, pág. 4–6, 2009.
9. W. Garira e F. Chirove, “Um método geral para modelagem multiescala de sistemas de doenças transmitidas por vetores,” *Interface Focus*, vol. 10, não. 1, 2020, Arte. não. 20190047.
10. L. Wang, J. Chen e M. Marathe, “Tdefsi: Theory previsão epidêmica baseada em aprendizagem profunda guiada com informações sintéticas”, *ACM Trans. Spatial Algorithms Syst.*, vol. 6, não. 3 de 2020, art. não. 15.
11. C. Yang et al., “EpiMob: análise visual interativa de restrições de mobilidade humana em toda a cidade para controle de epidemias,” *IEEE Trans. Vis. Comput. Gráficos*, acesso antecipado, 6 de abril de 2022, doi: [10.1109/TVCG.2022.3165385](https://doi.org/10.1109/TVCG.2022.3165385).
12. K. Samant, E. Memeti, A. Santra, E. Karim e S. Chakravarthy, “Cowiz: Visualização interativa de COVID-19 baseada em análise de rede multicamadas”, em *Proc. IEEE 37^a Int. conf. Data Eng.*, 2021, pp. 2665–2668.
13. M. Chinazzi et al., “O efeito das restrições de viagem na disseminação do novo surto de coronavírus de 2019 (COVID-19)”, *Science*, vol. 368, nº. 6489, pp. 395–400, 2020.

14. A. Adiga et al., "Avaliando o impacto das suspensões de companhias aéreas internacionais na disseminação global inicial do COVID-19", Medrxiv, 2020.
15. A. Wilder-Smith e D. Freedman, "Isolation, quarentena, distanciamento social e contenção comunitária: papel fundamental para medidas de saúde pública antiquadas no novo surto de coronavírus (2019-nCoV)", J. Travel Med., 2020.
16. OpenFlight, "Openflight: Aeroporto, companhia aérea e rota data," 2017. Acesso: jun. 2022. [Online]. Disponível: <https://openflights.org/data.html>
17. USDOT, "Bureau of Transportation Statistics," 2019. Acesso: jun. 2022. [Online]. Disponível: https://www.transtats.bts.gov/databases.asp?Mode_ID%1&Mode_Desc%Aviation&Subject_ID2%0
18. USCB, "Totais da população da cidade e da cidade: 2010-2018," 2019. Acesso: jun. 2022. [Online]. Disponível: <https://rb.gy/x90vz4>
19. Google, "Mobilidade da comunidade COVID-19 do Google relatórios", 2019. Acesso: junho de 2022. [Online]. Disponível: <https://www.google.com/covid19/mobility>
20. SK Dey, MM Rahman, UR Siddiqi e A. Howlader, "Analyzing the epidemiological surto of COVID-19: A visual exploratory data analysis approach," J Med. Virol., vol. 92, nº. 6, pág. 632–638, 2020.
21. BD Wissel et al., "Um painel on-line interativo para rastrear o COVID-19 em condados, cidades e estados dos EUA em tempo real", J. Amer. Med. Informar. Assoc., vol. 27, não. 7, pp. 1121–1125, 2020.

MUHAMMAD USMAN é professor assistente no Departamento de Informação e Ciência da Computação, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Arábia Saudita, onde também trabalha no Center for Intelligent Secure Systems. Seus interesses de pesquisa incluem simulação de multidão e modelagem, IA centrada no ser humano, dinâmica de comportamento da multidão, análise e otimização de projetos arquitetônicos, visual espacializações e realidade virtual. Usman recebeu seu Ph.D. formado em ciência da computação pela York University, Toronto, ON, Canadá. Ele é o autor correspondente deste artigo. Entre em contato com ele em muhammad.usman@kfupm.edu.sa.

HONGLU ZHOU é Ph.D. estudante no Departamento de Ciência da Computação, Rutgers University (desde 2017 Fall), Piscataway, NJ, 08854, USA, sob orientação do Prof. Mubbasir Kapadia. Ela fez estágios na NEC Laboratories America e Google YouTube). Seus interesses de pesquisa incluem gráfico

aprendizagem de representação, dinâmica de multidão e visão computacional. Zhou recebeu seu diploma de Bacharel em Engenharia em ciência da computação e tecnologia pela Communication University of China. Entre em contato com ela em hz289@scarletmail.rutgers.edu.

SEONGHYEON MOON é Ph.D. candidato na Rutgers University, Piscataway, NJ, 08854, EUA, em ciência da computação. Ele é membro do Laboratório de Interfaces Visuais Inteligentes supervisionado pelo Dr. Mubbasir Kapadia. Seus interesses de pesquisa incluem simulação de multidões, análise de multidões, segmentação de objetos e rastreamento de objetos. Moon recebeu seu M.Sc. formado em engenharia mecânica pelo Instituto de Ciência e Tecnologia de Gwangju. Entre em contato com ele em sm2062@cs.rutgers.edu.

XUN ZHANG é Ph.D. candidato no Departamento de Computer Science, Rutgers University, Piscataway, NJ, 08854, EUA. Seus interesses de pesquisa incluem semântica, ambientes virtuais e narrativas de multidão. Vigarista contate-o em xz348@scarletmail.rutgers.edu.

PETROS FALOUTSOS é professor do Departamento de Electrical Engineering and Computer Science, York University, Toronto, ON, M3J 1P3, Canadá, e cientista afiliado do Toronto Rehabilitation Institute, University Health Net trabalho, Toronto, ON M5G 2A2. Faloutsos recebeu seu Ph.D. Bacharel em Ciência da Computação pela University of Toronto, Toronto. Entre em contato com ele em pfal@cse.yorku.ca.

MUBBASIR KAPADIA é diretor do Laboratório de Interfaces Visuais Inteligentes e professor associado do Departamento de Ciência da Computação, Rutgers University, Piscataway, NJ, 08854, EUA. Sua pesquisa está na interseção de inteligência, computação visual e interação humano-computador, com a missão de desenvolver interfaces visuais inteligentes para capacitar a criação de conteúdo para design arquitetônico com consciência humana, narrativa digital e jogos sérios. Kapadia recebeu seu Ph.D. formado em ciência da computação pela Universidade da Califórnia. Sua pesquisa é financiada pela DARPA e NSF e por meio do apoio generoso de parceiros industriais, incluindo Disney Research, Autodesk Research, Adobe Research e Unity Labs. É membro do IEEE. Vigarista contate-o em mk1353@cs.rutgers.edu.