npj | urban sustainability

Article

Published in partnership with RMIT University

https://doi.org/10.1038/s42949-024-00148-x

Megacities are causal pacemakers of extreme heatwaves

Xueli Yang [1], Zhi-Hua Wang [1] ✉, Chenghao Wang [2,3] & Ying-Cheng Lai [4,5]

Global climate change has been shown to cause longer, more intense, and frequent heatwaves, of which anthropogenic stressors concentrated in urban areas are a critical contributor. In this study, we investigate the causal interactions during heatwaves across 520 urban sites in the U.S. combining complex network and causal analysis. The presence of regional mediators is manifest in the constructed causal networks, together with long-range teleconnections. More importantly, megacities, such as New York City and Chicago, are causally connected with most of other cities and mediate the structure of urban networks during heatwaves. We also identified a significantly positive correlation between the causality strength and the total populations in megacities. These findings corroborate the contribution of human activities e.g., anthropogenic emissions of greenhouse gases or waste heat, to urban heatwaves. The emergence of teleconnections and supernodes are informative for the prediction and adaptation to heatwaves under global climate change.

The progress of development of modern human societies is driven by urbanization. Or, as emphatically put by Spengler, "*world history is city history*"[1]. The sustainable future of human societies is closely linked to urban sustainability. Today, urban areas accommodate 56% of the world's population, consume over two thirds of the world's energy, and produce about 70% of global carbon emissions[2,3]. Critical environmental challenges faced by cities due to global climate change include excessive heat stress, environmental pollution, infrastructure vulnerability, public health risks, and degraded ecosystems, to name a few[4–6].

As a particular type of climatic extremes, heatwaves incur widespread adverse impacts on human health, infrastructure, and natural ecosystems[7]. Global climate change has led to an increase in duration, intensity, and frequency of heatwaves in contiguous United States (CONUS) since the mid-1960s[8–10]. To understand the formation and development of extreme heatwaves, both natural contributors[7] and anthropogenic drivers[10] were studied. Natural contributors of heatwaves include, but are not limited to, high-pressure synoptic systems[11], land surface and temperature couplings such as soil moisture memory[12], climate variability and large-scale teleconnections such as El Niño-Southern Oscillation (ENSO)[13] and the Pacific Decadal Oscillation (PDO)[14]. Meanwhile, anthropogenic emissions have contributed approximately to two-thirds of the observed global temperature rise due to climate change for the period 1971-2010[15].

The contribution of human activities to global climate change is mainly associated with the combustion of fossil fuels and the concomitant increases of greenhouse gases (GHGs) in the atmosphere. The warming effects caused by increasing GHGs have been detected in most continents[16,17]. For example, the observed increasing trends in U.S. temperature extremes for the last few decades were mainly due to changes in anthropogenic forcings[18,19]. The projected near-term increases in GHG forcing could result in warm-season drying and intensification of heat extremes throughout most U.S. regions[20]. For example, a recent study revealed that almost half of the world's population will likely be exposed to heatwaves threatening human thermoregulatory capacity by 2100 under an intermediate emission scenario (Representative Concentration Pathways RCP 4.5), and nearly third-quarters of the world's population will be threatened by heatwaves under a higher emission scenario (RCP 8.5)[21].

Various urban environments are intrinsically connected and co-evolve together. For example, some major U.S. cities exhibit striking similarity (or "analog") in their urban thermal environments[22]. This can be partly explained by the long-range connectivity in urban areas, or "teleconnection"[23]. U.S. cities are clustered as connected groups under different environment stressors associated with human activities, such as extreme heat, air pollution, and precipitation climatology[24]. To gain insights into the intrinsic dynamics of heatwave propagations, it is useful to take a

[1]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA. [2]School of Meteorology, University of Oklahoma, Norman, OK 73072, USA. [3]Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73019, USA. [4]School of Electricity, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA. [5]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. ✉e-mail: zhwang@asu.edu

network approach by treating the various urban areas in the U.S. as the interconnected nodes in the network.

In recent years, the techniques of complex networks have been applied in climate studies, ranging from the observations and predictions of weather extremes[25–28] to early warning schemes[29,30]. The climate system can be treated as a network of many dynamical systems that enables us to investigate collective behavior of climate dynamics from network theory[26]. A climate network is characterized as a collection of components in climate that are meaningfully connected. In the network, spatial variables in the climate system, such as precipitation or temperature, can be viewed as nodes whose connectivity (by edges) is determined by a critical threshold of strength of interactions based on linear or nonlinear interactions[27,28]. The connectivity among different locations (cities) in the climate network, i.e., the hub-periphery structure (clustering in network structure), can be quantified by network analysis[24]. Previous efforts in identifying various atmospheric teleconnections were largely based on linear Pearson's or Spearman's correlations[31]. However, the so-constructed climate networks may not capture the underlying physical processes in the complex climate system[32–34]. To overcome the deficiency, nonlinear-dynamics based methods have been used to investigate the interactions of large-scale patterns in climate, such as event synchronization[26,27,35,36], mutual information[32], distance metric[37], and causality[33,34,38]. Causality networks are especially useful for distinguishing direct dependencies from indirect ones or from common drivers in complex climate systems[33,34,38,39]. Compared with numerical models, network-based analysis is less influenced by measurement errors and initial and/or boundary conditions, and are relatively more accurate for weather predictions[29].

In this work, we aim to investigate causal interactions across the built environment of the U.S. during summertime (May-September) heatwaves. This is primarily motivated by the fact that the causal analysis of urban heatwaves has been hitherto underexplored, especially from the perspective of complex urban system dynamics. In fact, quantifying the impact of human activities and their connection with extreme heatwaves remains an open challenge, especially in urban areas where the impact of heatwaves can be intensified through, e.g. the synergistic interactions with urban heat islands. Here we evaluate the spatial patterns of U.S. extreme heatwaves for the last four decades (1979-2021) and identify various causal relationship among 520 U.S. urban areas during the detected extreme heatwaves. The inferred causality allows us to construct *directed* causal networks and to evaluate the network topology, viz. the structure and connectivity matrices that quantify the network. In addition, we study the emergence of hubs/clusters and estimate how large cities contribute to the propagation of extreme heatwaves in the causal network. The presence of hub-periphery structures in urban networks enables us to identify supernodes in heatwaves propagations, especially in Northwest, Northeast, and the Great Lakes. Further, megacities (with population totals greater than or around 10 million in metropolitan areas), such as New York City and Chicago, are manifested as *pacemakers* during heatwaves, in the sense that they are causally connected with most other urban areas and mediate the occurrences and propagations of heatwaves. It is also shown that human activities, roughly represented by the total population size, largely impact the causal interactions of cities in urban heatwaves.

Results

Major heatwaves in CONUS

In the U.S., warm extremes have increased more rapidly in recent decades compared to cold extremes[18]. In fact, warm season (May-September) heatwaves across the U.S. exhibited a substantial increase during the last four decades[40,41]. We use daily maximum temperature of the gridded data during warm season to detect extreme heatwaves during the warm seasons of 1979-2021. From these 43 years, we select 12 extreme heatwaves (Supplementary Figure 1) that impacted most of the U.S. states. Of all the heatwaves detected, we found some regions such as the Great Plains (Texas, Oklahoma, and Kansas) and Southwest (Arizona, California) are more susceptible to extreme heatwaves, which is consistent with findings in prior studies that the central and southern U.S. is an "warming hole" region[42]. In Fig. 1, four out of

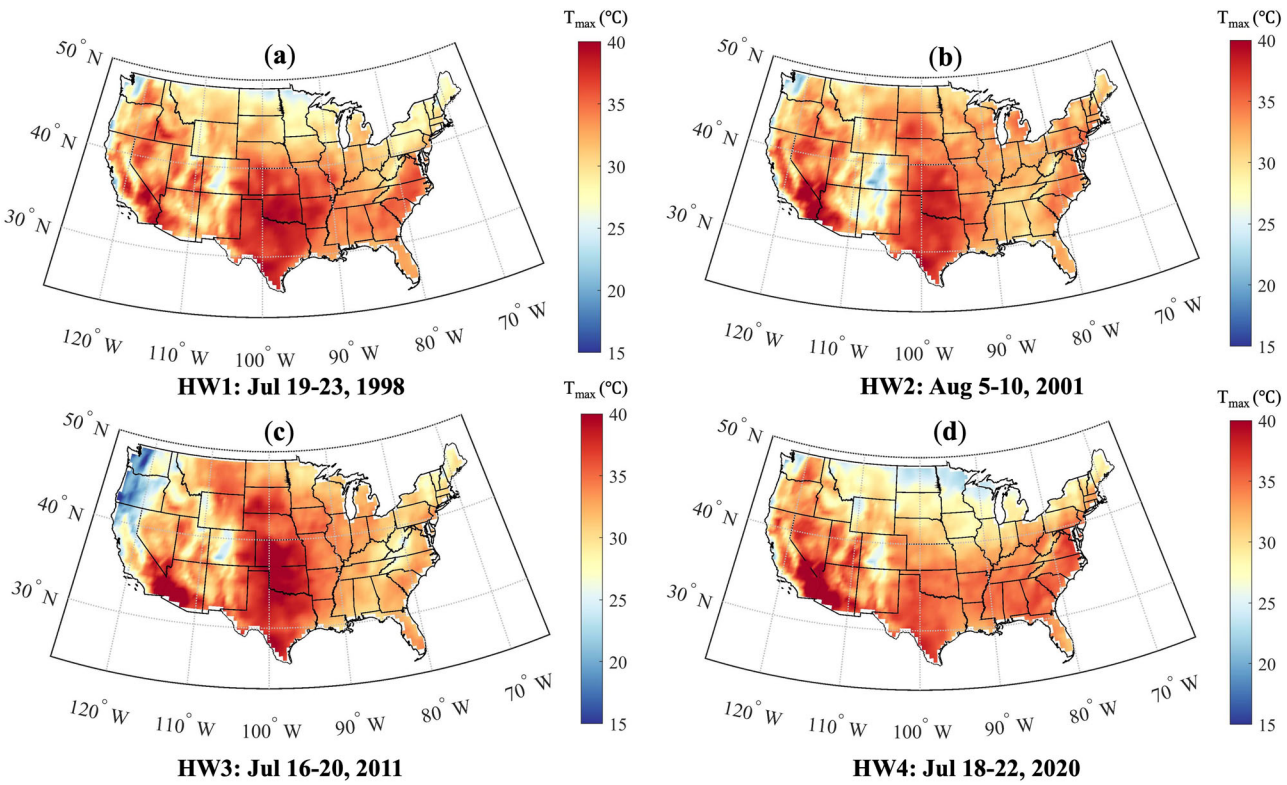

Fig. 1 | Identified major heatwave events based on maximum daily temperature (°C) across the CONUS. Four major heatwaves (**a**) July 19-23, 1998, **b** August 5-10, 2001, **c** July 16-20, 2011, and **d** July 18-22, 2020, are presented here for illustration purposes (all 12 heatwaves can be found in Supplementary Figure 1).

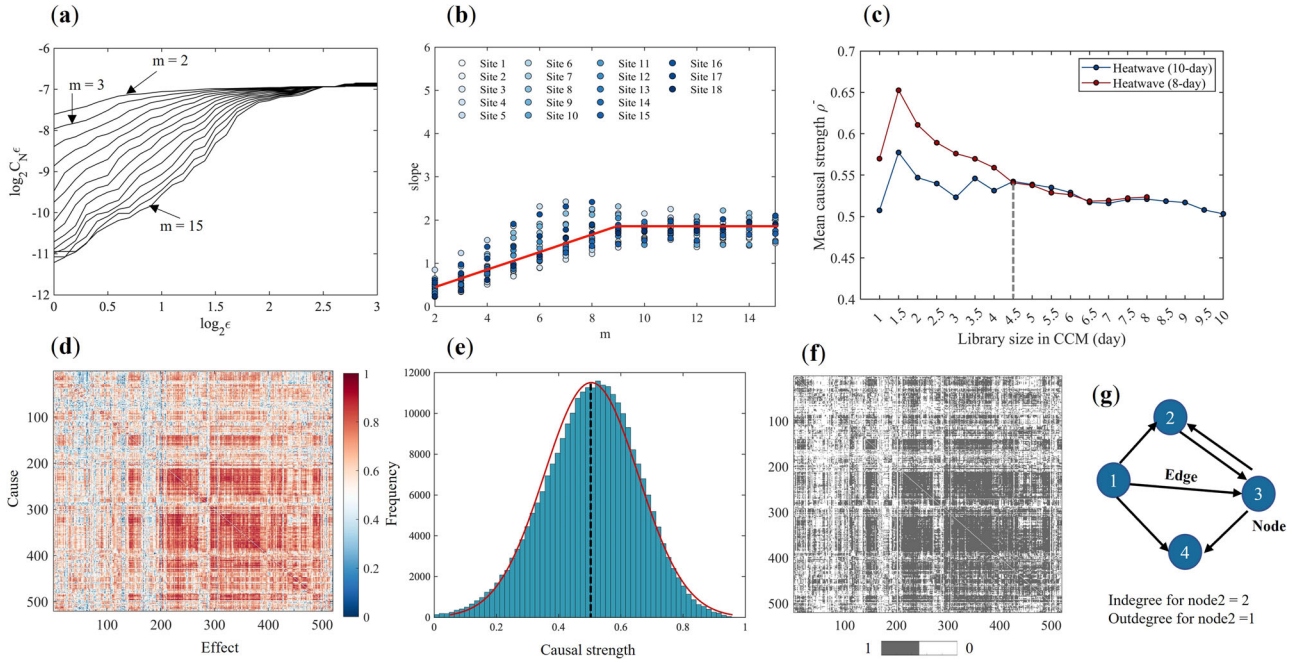

Fig. 2 | An illustration of causal network construction using the CCM algorithm during a U.S. heatwave event. a The relationship between $\log C_N(\varepsilon)$ and the neighborhood size ε with different embedding dimension m (ranging from 2 to 15) (site 1 is shown here for illustration). **b** The slope values in curve (**a**) for different embedding dimension m values at 18 sites (randomly selected from the 520 sites). The embedding dimension for the reconstruction of phase-space manifold in CCM is taken as the value of m where the slope plateaus (shown as the red solid line). **c** Robustness test of the library size L in the calculation of CCM causality. **d** The directed causality network (520 urban sites as nodes) for the heatwave in 2001 (results for other heatwaves can be found in Supplementary Fig. 6). **e** Distribution of causality and the threshold chosen (black dotted line) for constructing the adjacency matrix (0-1 matrix), as shown in **f**. **g** An illustration of a directed network structure (nodes, ingoing, and outgoing links).

the 12 heatwave events are selected for illustration purposes, where each heatwave event lasted for more than four days (see Methods for the detection of heatwave events), which captures the characteristic patterns of heatwaves in the U.S. For example, the Texas heatwave of 2011 (Fig. 1c) was centered in Texas and extended to adjacent southern plains states accompanied by an extreme drought (characterized as the record-breaking driest consecutive 12 months based on precipitation). The principal physical processes contributing to this record-breaking heatwave were associated with the coupling of soil moisture and high temperature, such as the severe rainfall deficit and antecedent dry conditions[43]. This region has also experienced other notable heatwaves, including the Texas-Oklahoma heat wave in 1998 (Fig. 1a), which have strong connections to antecedent droughts[44]. Since the focus of our study is on the heatwaves of the U.S. urban areas, we extract the corresponding hourly air temperature during the same time (1979–2021). The distribution of temperature and the anomalies after removing the long-term trend for the 520 urban sites can be found in Supplementary Figs. 2 and 3.

Causal network of urban heatwaves

Network theory has been increasingly applied to climate sciences in recent years. Mathematically, a climate network is a graph that consists of nodes and edges that connect nodes in meaningful way[26,30,32]. In this study, each of the 520 urban sites are represented as nodes of the urban network, where the edges are determined by the causal strength between a pair of nodes (see **Methods**). Long range connections (edges linking geographically distant vertices) enhance the information transfer (heat or moisture) within the network[32]. Identifying the evolution of teleconnections can provide a pathway to analyze or even predict large-scale extreme climate events (heatwaves). Those atmospheric teleconnections presented in data can be well captured through network analysis[29]. Besides, the directions of the teleconnections are key to deciphering the spatial propagation of heatwaves. Causality is capable of revealing the true interactions and connecting them with the underlying physical mechanisms of the climate dynamics[33,34,38].

Here we construct directed climate networks based on the causality for all 520 stations in urban areas during each of the 12 heatwaves (the details of the causality calculation using the convergent cross mapping (CCM) method can be found in **Methods**). The network results for all heatwaves are shown in Supplementary Figure 6. Figure 2d shows the asymmetrical causal matrix of the 520 urban locations (no self-connection) for the August 5-10, 2001 heatwave event, revealing directed interactions among most city pairs. Some pairs show a high causality over 0.8, suggesting strong causal interactions during heatwaves. To determine the locations of cities having strong causal connection with others during heatwaves, we introduce three parameters: Indegree, Outdegree, and PageRank centrality in the (directed) network analysis to uncover the structures of causal networks. The indegree and outdegree of a node measure the ingoing and outgoing links from and to other nodes in a network, respectively[45]. Clustering of homogeneous sites with high inward strength (similar to indegree here) are vulnerable to heatwaves occurring at other various locations[27], while centrality can be interpreted heuristically as the energy or matter flow[28]. In our study, a city with a higher indegree and outdegree has more connections with others in two-way directions during heatwaves spreading, and the PageRank centrality of a city can be viewed as its importance in propagating extreme heat flow in the network. The distribution of indegree, outdegree, and PageRank centrality are shown in Figs. 3–5, along with the directed ingoing and outgoing spatial links, including the long-range teleconnections. The causal network of heatwaves across the CONUS possesses hub- or core-periphery structures, where some cities stand out as groups with high values of network metrics. These regions are highlighted in zoom-in boxes (**a**-**g**) in Figs. 3–5, such as urban clusters in the Northeast (box **g**), the Great Lakes (box **e**), and the Ohio valley region (box **f**). This signifies the vital roles these urban regions play in the spreading of heat during extreme heatwaves, during which a large area of the U.S. becomes vulnerable to extreme thermal conditions.

In a climate network, certain nodes pose more connections than others: a cluster of such nodes functions as hubs or supernodes[46]. Previous studies

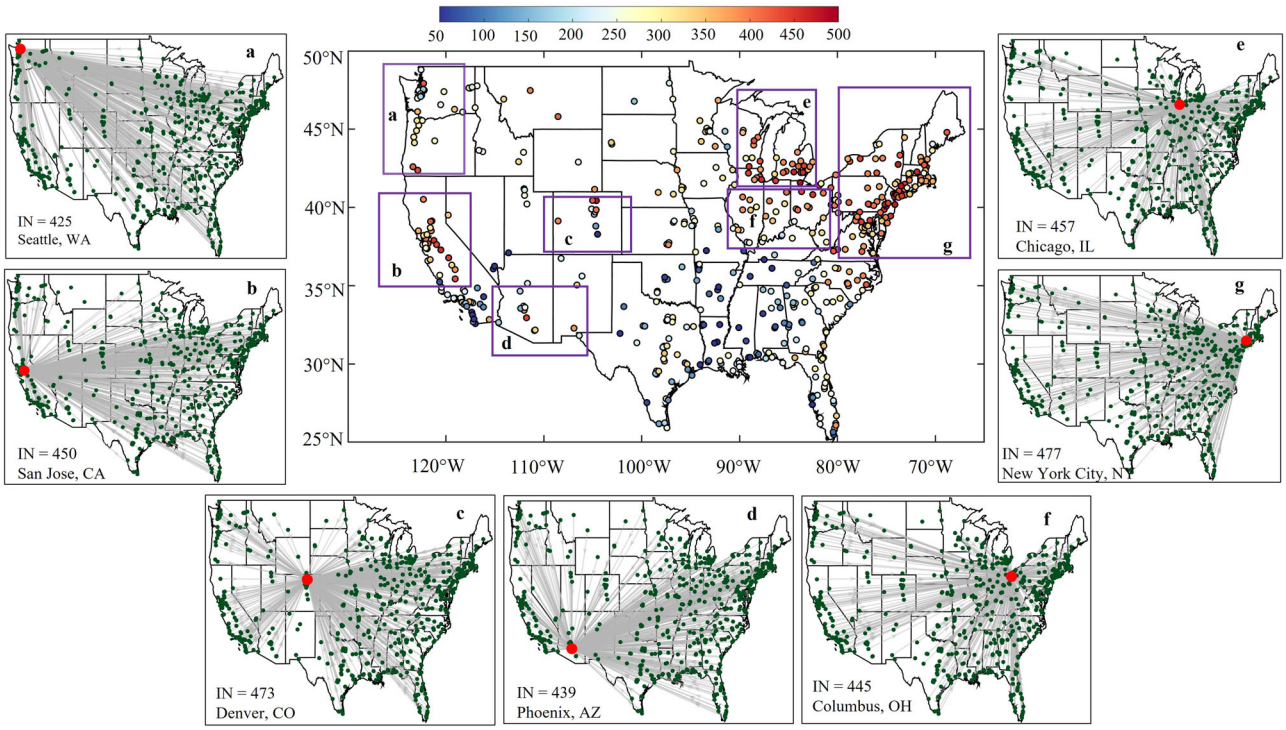

Fig. 3 | Indegree of 520 urban sites in causal network during a heatwave event. Subplots from (**a**) to (**g**) are the "hub" cities that show the greatest values of indegree or *ingoing* edges (denoted by the gray directed links with ingoing arrows), inlcuding (**a**) Seattle, WA, **b** San Jose, CA, **c** Denver, CO, **d** Phoenix, AZ, **e** Chicago, IL, **f** Columbus, OH, and **g** New York, NY. Inset values are for the highest indegree value (IND) among urban areas for each sub-region ((**a**)-(**g**)). Those cities are also metropolitan areas with large population, such as New York City, Chicago, Seattle, San Jose. Green nodes in each subplot are the urban areas that are causally connected with the considered metropolitans (marked by red solid dots), and the gray links with ingoing arrows indicate where the causal information of heat comes from. A city with high indegree indicates that it is venerable to extreme thermal conditions from large number of other urban areas, and it serves as a heat *sink* experiencing simultaneous heatwave propagation from others.

revealed that regions such as Northwest and Southwest behave as drought hubs that synchronize with others via teleconnections across different spatial scales, possibly attributable to the upper-tropospheric ridge over the Rockies and the North Pacific High[28]. Another study suggested that regions with high connectivity in the northeast U.S. and the Great Plains exhibit a high coherence of extreme precipitation[37]. In addition, causality studies on the regional scale found that the Ohio Valley region acts as a regional gateway of long-term heat transfer and moisture transport across the U.S., regulated by atmospheric uplifts[34,38]. For heatwaves, the locations with the highest centrality are clustered near Lake Michigan and Northeast, while those with the lowest centrality are mainly concentrated near the coastal regions of CONUS, including regions in southern U.S. such as Florida and Southern California[27].

Surprisingly, besides the climate conditions associated with high connectivity of these seven subregions, we find that the urban areas with the highest values of network metrics within each sub-region are almost coincident with major populous cities (megacities), as shown by the red dot in each subplot in Figs. 3–5. For example, in Fig. 3, cities such as New York City (NY), Chicago (IL), San Jose (CA) are causally connected with others, as represented by the number of ingoing links (IN number in each subplot) over 450, suggesting those cities are impacted by about 90% of the 520 urban sites during the heatwave event. From a network perspective, these large cities serve as heat *sinks* during heatwaves. The outgoing links of cities represented by outdegree in Fig. 4 also show some hubs in those seven subregions, and large cities such as Detroit (MI), New York City (NY), San Jose (CA), Portland (OR), Columbus (OH) exert outgoing influence on approximately 400 other urban areas, acting as heat *sources* during the extreme heatwaves. It is worth noting that both ingoing links and outgoing links come to or from those large cities that contain long-range teleconnections across large spatial areas of the U.S. Furthermore, we find some large cities, such as Denver (CO) and Phoenix (AZ) are causally impacted by many urban areas even when they are "isolated" cities located in less dense urban clusters (box **c** and **d**). The distribution of the PageRank centrality in Fig. 5 represents the overall importance of cities on regulating heatwave propagations, and large cities with high value of PageRank centrality such as Chicago (IL), Seattle (WA), New York City (NY), San Jose (CA) serving as *pacemakers* on propagation of extreme thermal conditions. Similar information can be inferred from the results for other heatwaves (Supplementary Figures 9-11). These results suggest that the propagation of extreme thermal conditions among urban areas is likely due to the interplay of synoptic (climate) conditions with anthropogenic stressors such as emissions of GHGs and waste heat.

The role of megacities in regulation of heatwaves

Extreme heatwaves can influence building performance and waste heat discharge in several ways. The waste heat emission in inland and dense urban districts is more sensitive to extreme events than in coastal or suburban areas[47]. To further investigate the roles that large cities play in the extreme heat propagation across the U.S. we choose some of the large cities (53 in total) with populations over 200,000 according to population totals in 2020 from the U.S. Census Bureau, Population Divisions (See **Methods**). Here we adopted two population metrices, viz. the population totals and density, to represent the extent of human-induced activities in large U.S. cities. Figure 6 shows the comparison of both population metrics in large cities, and their correlation with the causality (outdegree) for a selected heatwave event in 2020. The results in Fig. 6a and Fig. 6c suggest that most of those large cities have more than 230 outgoing links to other urban areas, meaning that they exert influence on more than half of 520 urban sites. Similar patterns can be observed in other extreme heatwaves (See Supplementary Figures 12 and 14), signifying that large cities are more conducive

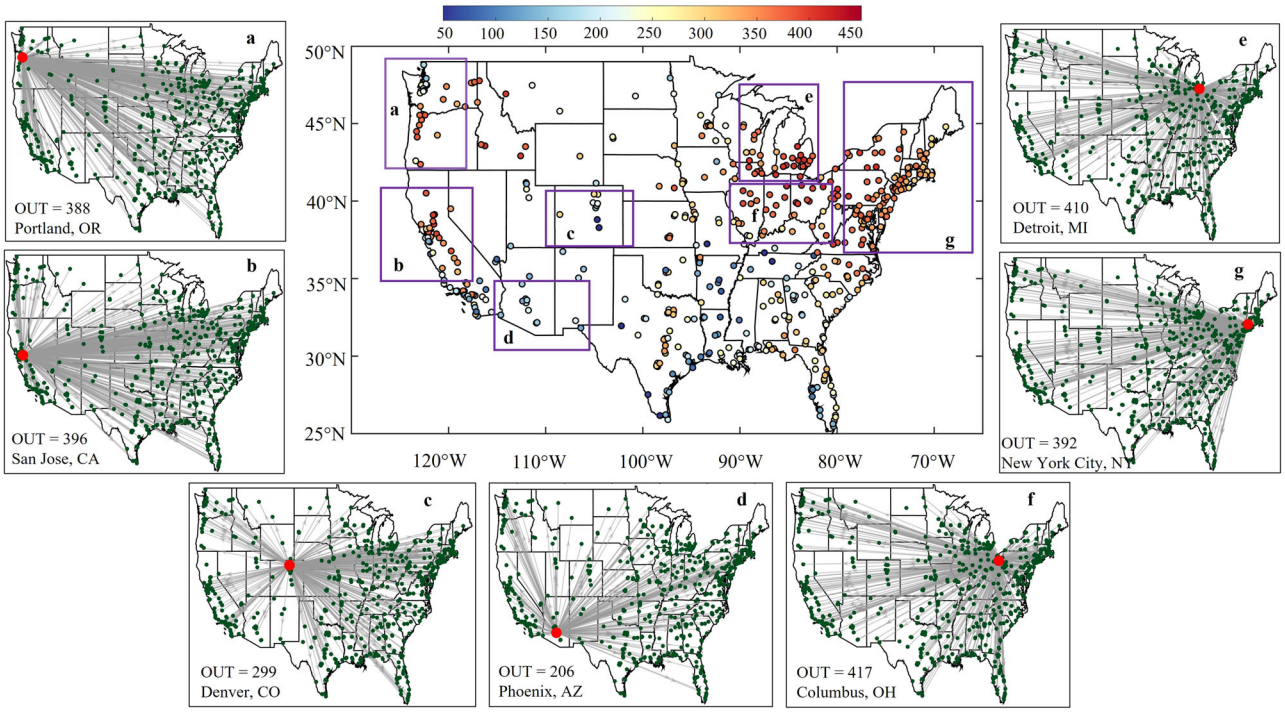

Fig. 4 | Outdegree of 520 urban sites in causal network during a heatwave event. The seven subregions inlcude (**a**) Seattle, WA, **b** San Jose, CA, (**c**) Denver, CO, **d** Phoenix, AZ, **e** Chicago, IL, **f** Columbus, OH, and **g** New York, NY, same as those in Fig. 3. The city with the highest outdegree (OUT) is marked by a red dot in each zoom-in box, where the *outgoing* edges (denoted by the gray directed links with outgoing arrows) means the connections that this target large city made to other urban areas (green nodes) during the extreme heatwave. Those cities are metropolitan regions with large populations, such as New York City, Portland, and Detroit. A city with high outdegree indicates that it serves as a heat *source* simultaneously dispersing heatwave to others.

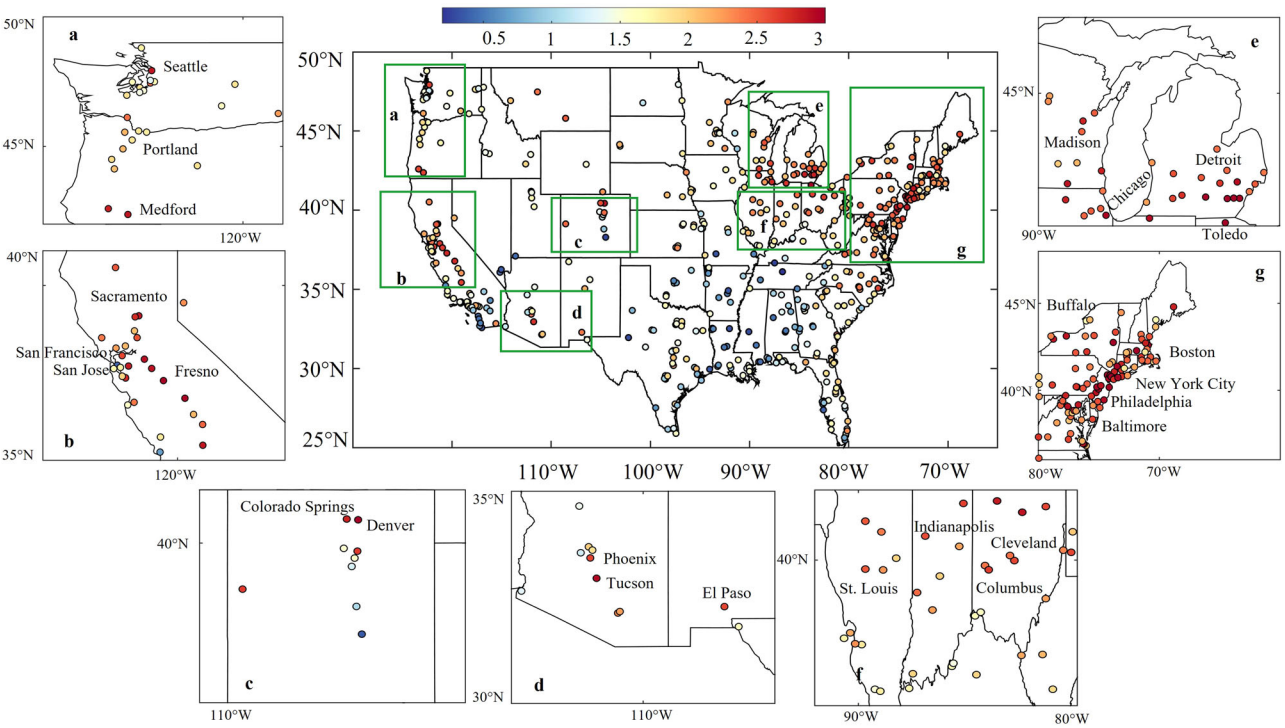

Fig. 5 | PageRank centrality of 520 urban sites in causal network during a heatwave event. The seven subregions inlcude (**a**) Seattle, WA, **b** San Jose, CA, **c** Denver, CO, **d** Phoenix, AZ, **e** Chicago, IL, **f** Columbus, OH, and **g** New York, NY, same as in Fig. 3 with relatively large values of centrality. Metropolitan regions with the largest population size in the subregions are also labeled (approximate locations). The PageRank centrality can be interpreted as a measure to characterize the importance of 520 urban sites in heatwave spreading in the causal network during a heatwave event.

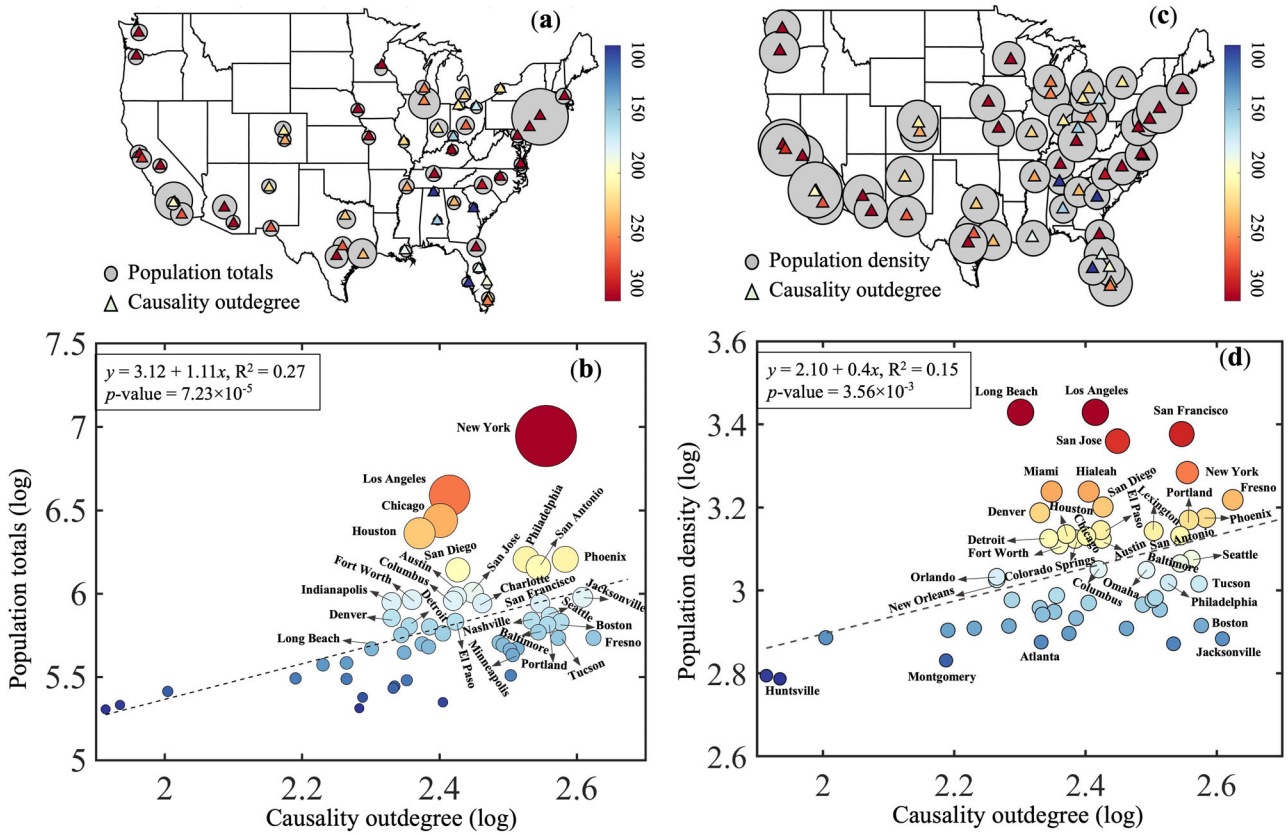

Fig. 6 | Comparison of causality (outdegree) with two population metrics for 53 large U.S. cities with a population over 200,000 during the 2020 heatwave event (July 18-22). a, **b** For population totals, and **c** and **d** are for population density. The top panel shows comparison maps of the causal outdegree (triangular markers) and population metrics: population totals in **a** and population density in (**c**), with sizes represented by gray shaded circles. The bottom panel represents the log-log correlation between the causal outdegree and population metrics: population totals in **b** and population density in **d** and causality, with sizes represented by colored circles.

to heat propagation to other (ambient and distant) areas and act as heat *sources* during these events. Interestingly, the results in Fig. 6b indicate positive correlation between the causality and population totals during many (10 out 12) heatwaves (See Supplementary Figure 13). For the attribution of population density, the results exhibit similar, though apparently weaker, positive correlation with the causal outdegree in Fig. 6d, than its counterpart of popular totals in Fig. 6b; both results are statistically significant. There are 8 out of 12 heatwaves that show a positive correlation between population density and causality (See Supplementary Fig. 15). This is understandable because the population totals have much larger variability among large U.S. cities (Fig. 6a), than the population density (Fig. 6c). Megacities with huge total population do not necessarily have a higher population density; an example is that the population density of the New York city is less than San Jose (Fig. 6d). Hence, the population totals may better represent the compounding impacts of human activities, such as the vehicular emission of waste heat whose density is more sensitive to population totals than density. Nevertheless, it is caveated that the relationship between causality and population metrics can be far more complex than a simple log-linear relationship[48]. Other measures of human activities such as landuse types (e.g., coverage of paved or vegetated areas) or management practices (e.g., urban irrigation) have the potential to significantly contribute to causal interaction with the thermal environment in cities during heatwaves.

Discussion

Causally interpretable networks of climatology such as precipitation[38], temperature[34], or mean sea-level pressure[33] facilitate a data-driven approach to investigate regional atmospheric pathways and gateways. Extreme climatic events are characterized by behaviors distinct from those of long-time climatology in terms of duration and magnitude, especially at the city-scale where the anthropogenic activities are concentrated. This work explores the patterns of urban heatwave propagation through a complex network analysis combined with causal inference. Our main finding is that the spreading of extreme thermal conditions among the urban areas exhibits clear hub-periphery structures, along with long-distance teleconnections via causal pathways. A key point is that the topological pattern of the heatwave networks is largely modulated by some city hubs, especially megacities, such as New York City and Chicago. Those megacities emerge as *pacemakers* of the occurrence of extreme heatwaves. Our study also unravels a positive relationship between the population and causality during heatwaves. This is likely due to the concentrated population and the associated extensive anthropogenic emissions of waste heat and greenhouse gases, such as increased use of air-conditioning during the heatwaves[47] or extensive anthropogenic emissions of air pollutions[24].

Our study quantifies the importance of urban areas across the CONUS in the propagations of extreme heatwaves. The findings can provide practical guidelines on extreme heatwave mitigation for urban planners, policymakers, and stakeholders. For instance, judicious strategies to reduce extreme heat stress could be implemented based on assessments of how vulnerable cities are (indegree) and which cities have a greater chance to potentially exert heat impact on others (outdegree). It has been known that the characteristics of long-range teleconnections in the climate network before extreme events can be helpful for the early prediction of climate drivers such as El Niño event[29]. As the frequency of extreme ENSO is projected to increase with GHG concentration[49], the long-distance teleconnections among urban areas uncovered in our study during extreme heatwaves can provide useful information for predicting extreme events. In addition, our study offers a new insight to inter-municipal and inter-

regional coordination of urban mitigation/adaptation strategies among different cities for better efficacy of sustainable urban development plans.

To the best of our knowledge, this study is a pioneering attempt to disentangle the causal mechanism behind extreme heatwaves and attribution of human impacts on urban heat extremes in the U.S., calling for future data-driven research endeavors on complex climate systems. For example, multidimensional measurements including frequency, magnitude, duration, and spatial extent could be considered to analyze heatwaves[40]. In addition, the role of atmospheric moisture can be included to determine heatwave impacts in urban areas, as humidity is a key variable in the measure of the human thermal comfort and heat-related morbidity and mortality[41]. Recent advances in machine learning techniques exhibit great potential for future studies of extreme urban heatwaves. For instance, a deep learning based Graph Neural Network was recently developed to predict regional extreme heatwaves, in terms of spatial extent and frequency, in U.S. cities[50]. Artificial neural networks (ANNs) in combination with complex network analysis have a great potential to capture the characteristics of the climate system for better prediction[29]. The framework proposed here can be integrated with deep learning tools for predicting urban heatwaves and their mechanistic causality.

While our analysis in this study is focused on the historical heatwaves, the proposed framework of causal network analysis, on the other hand, can be readily extended to enhance the predictability of future urban heatwaves. For example, it has been found that the structure of climate networks exhibits distinct patterns of evolution at the onset of critical transitions in climate dynamics (with the occurrence of heatwave as a special case), such as the increase of network clustering coefficient[30]. Therefore, by looking into the temporal variation of causal networks constructed using projected future evolution of urban temperatures in CONUS cities, it will enable the detection of possible occurrence of future heatwave and provide early-warning signals to such critical events. In addition, utilizing the framework proposed in this work, we can also identify and quantify potential causal drivers for the occurrence of extreme events, such as atmospheric forcing (e.g., blocking high pressure systems) and/or natural variability of climate system (e.g., ENSO and PDO).

Furthermore, it is also possible to extend the current work to look into the multiple spatiotemporal extents and scales involved in urban heatwaves and their dynamic evolutions under different background climate conditions. For instance, heatwaves spreading among arid/semi-arid cities susceptible to drought conditions likely have distinct and localized causal relationship with variables such as humidity or soil moisture. It remains intriguing, albeit challenging, to unravel causal heat-drought interactions in the context of a changing climate. It is also noteworthy that the causal network analysis used in this study is scalable, e.g. downscaled to neighborhood or upscaled to global scales, and transferable to urban networks in Europe or East Asia in the context of diverse urban planning/management practices. The key challenge that hinders the scalability and transferability of the proposed framework is the availability of archived urban dataset with sufficiently fine temporal resolution (to meet the minimum library size for CCM analysis) and/or adequately large spatial coverage (to render spatially complete network topology), which remains scarce especially for cities in developing countries.

Last but not least, the increasingly imperative urban environmental challenges, of which urban heatwaves are critical players, call for innovative adaptation strategies to be supported by cooperations among researchers, policy makers, and stakeholders. Nevertheless, such cooperation and/or resources for climate mitigation cannot be equally distributed spatially/temporally among stakeholders from all cities in the networks (such as the CONUS networks of cities), largely due to economic disparities, political priorities, and lack of regional and/or intercity agreements. The current study helps to provide a guidance of prioritizing the climate solutions temporally (e.g. towards extreme events such as heatwaves) and spatially (e.g. towards megacities that are hubs or pacemakers), so to optimize the efficiency and outcome of these solutions.

Methods

Detection of extreme heatwaves

In this study, extreme heatwaves are defined as major heatwaves in the study period with comparatively higher intensity, longer duration, and/or larger spatial coverage. For detection of extreme heatwave events, daily maximum temperature data covering 1979-2021 (43 years) are acquired from NOAA Climate Prediction Center (CPC) https://www.psl.noaa.gov/data/gridded/data.cpc.globaltemp.html, with 0.5 °× 0.5° spatial grids over the global domain starting from 1979. The CPC Global Unified Temperature data is provided by the NOAA PSL, Boulder, Colorado, USA. The dataset is based on global GTS data gridded using the Shepard algorithm[51]. In our study, we focus on the data for the warm season (May-September) over the CONUS. Heatwaves represent a continuous period of elevated temperature[7] and is defined as a period (at least 3 consecutive days) when the daily maximum temperature exceeds 99th percentile of the daily maximum during warm seasons for each year[24,27]. The 99th percentile is calculated based on the daily maximum temperature, centered on a 15-day moving window[41]. This relative threshold allows for a better representation of the physics behind the extremes for all locations with different climate conditions[7]. This criterion is applied for each grid during the study period from 1979 to 2021, and intense (with large areal extent) heatwaves across the country are chosen for analysis. For our analysis, 12 heatwaves during 1998-2021 are selected (Supplementary Figure 1).

Hourly air temperature data for CONUS urban areas

We retrieved hourly air temperature data of 520 urban weather stations from the Historical Comprehensive Hourly Urban Weather Database (CHUWD-H). This database has hourly air temperature from 1998 to 2021-the same period for network analysis of extreme heatwaves in this study. It is noteworthy that these weather stations constitute a subset of over 900 stations from the latest release of the National Renewable Energy Laboratory's typical meteorological year (TMY3) database, which ensures the quality of weather observations at these stations. To be qualified for an "urban station", the selected stations must be located within or nearby (within 55 km) urban areas. Among the 520 stations, 390 weather stations fall directly within urbanized areas (with at least 50,000 population) defined by the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) database (https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2010/TGRSHP10.pdf), while 130 weather stations are in close proximity to these urbanized areas (selected through manual inspection). The quality controlled hourly air temperature observations at all stations are retrieved from the Integrated Surface Database (ISD)[52]. Post-processing, gap filling, and quality control processes were then performed following our previous study. With respect to hourly temperature and its pre-processing (anomaly) for those 520 urban sites during the detected extreme heatwaves, we remove the long-term average value for each hour in each urban site to reduce the effect of seasonal/annual cycles and potential thermal analogue between adjacency cities on the causality results. The temperature data (original and anomaly) during heatwave events can be found in Supplementary Figures 2-3.

Causal inference using CCM method

For nonlinear dynamic system with weak or moderate coupling, the CCM method is used to detect causality from time series[53], which is based on the classic Taken's lag-coordinate embedding theory for reconstructing the shadow manifolds of the underlying nonlinear system[53–55]. A variable X with length L contains the corresponding time series $X = X(1), X(2), \ldots, X(L)$, to capture the dynamics of X, a manifold M_X needs to be reconstructed from lagged-coordinate vectors of X to estimate contemporaneous values of Y. The points in the manifold, denoted by $x(t)$, consists of a set of E-dimensional vectors: $x(t) = [X(t), X(t-\tau), X(t-2\tau), \ldots, X(t-(E-1)\tau)]$ can be formed for $t = 1 + (E-1)\tau$ to $t = L$, with $\tau > 0$ being the time lag. This set of vectors represents the reconstructed shadow manifold M_X. The calculation of causality in CCM lies in a simplex projection, a nearest-neighbor

algorithm that involves exponentially weighted distance from nearby points on a reconstructed manifold to do kernel density estimation[53].

To generate a cross-mapping estimation of $Y(t)$ from information of M_X, denoted as $\hat{Y}(t)|M_X$, we begin with locating the contemporaneous lagged-vector $x(t)$ on M_X and find its $E+1$ nearest neighbors, and then find the time indices of those $E+1$ nearest neighbors of $x(t)$, denoted from the closet to farthest by $t_1, t_2, \ldots, t_{E+1}$. Those time indices are used to identify the neighbors in Y to estimate $Y(t)$ from a locally weighted mean of the $E+1$ of $Y(t_i)$ values, the cross mapping from X to Y is defined as

$$\hat{Y}(t)|M_X = \sum_{i=0}^{E+1} w_i(t) \cdot Y(t_i) \tag{1}$$

where w_i is a weight based on the distance from its i^{th} nearest neighbor on M_X, and $Y(t_i)$ are the contemporaneous values of Y. The weights are determined by,

$$w_i(t) = \frac{u_i(t)}{\sum_{j=1}^{E+1} u_j(t)} \tag{2}$$

with

$$u_i = \exp\left\{-\frac{d[x(t), x(t_i)]}{d[x(t), x(t_1)]}\right\} \tag{3}$$

and $d[x(t), x(t_i)]$ is the Euclidean distance between two vectors. If X and Y are dynamically coupled, the nearest neighbors of M_X should identify the time indices of corresponding nearest neighbors on M_Y. The predictability measured by the correlation between the original $Y(t)$ and predicted values $\hat{Y}(t)|M_X$ from cross-mapping estimates, can be used to measure the directed dynamical influence from Y to X, which is defined as

$$\rho_{Y|M_X} = \frac{\mathbf{E}\left\{\left[Y(t) - \mu_Y\right] \cdot \left[\hat{Y}(t)|M_X - \mu_{\hat{Y}}\right]\right\}}{\sigma_Y \sigma_{\hat{Y}}} \tag{4}$$

where **E,** μ and σ are the statistical expectation, average, and standard deviation, respectively.

Larger values of $\rho_{Y|M_X}$ implies a stronger causal influence from Y to X, a non-positive value of $\rho_{Y|M_X}$ suggests that Y is not causal to X. Similar process can be applied to measure a causal influence from X to Y. In this study, X and Y denote the air temperature (anomaly) during heatwaves for each pair of cities.

To reconstruct the phase space of the underlying dynamical system, the delay-coordinate embedding algorithm in CCM requires a proper choice on the time delay τ and embedding dimension E value[56]. The method for choosing the parameters τ and E in our work is the correlation integral and dimension[57–59]. The correlation integral $C_N(\varepsilon)$ is the fraction of pairs of points on the attractor within a hypersphere of radius ε, it can be evaluated approximately by

$$C_N(\varepsilon) = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta\left(\varepsilon - \| x_i - x_j \|\right) \tag{5}$$

where N is the number of points inside the time series of X, Θ is the Heaviside step function, and $\| x_i - x_j \|$ is the distance between two vectors. The correlation dimension D_2 is given by

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\log C_N(\varepsilon)}{\log \varepsilon} \tag{6}$$

Results of calculations of the correlation integral from the time series of temperature records during two long-lasting heatwave events for some urban areas (18 randomly chosen) can be found in Supplementary Figs. 4 and 5. In the main text we present some results in Fig. 2: Slope values of $\log C_N(\varepsilon)$ versus $\log\varepsilon$ for increasing values of embedding dimension m are shown in Fig. 2a. The slope increases with the embedding dimension and reaches a plateau at m greater than 10, justifying $E = 10$ in CCM causality analysis, as illustrated in Fig. 2b. To ensure the data length or size of library L is sufficient to calculate causality, a sensitivity test for L is performed for two long-lasting heat wave events (see Supplementary Figs. 4 and 5) and the results for one case is illustrated in Fig. 2c, where the two different heatwave events show similar behavior as with the increase of library size: fluctuations at the beginning and stable values when L is large enough (around 4.5 days). All the major heatwaves chosen in this work with more than 4 days can be used to reconstruct the phase-space manifolds in the CCM method.

Network analysis

A *network* or graph is a collection of nodes (vertices) joined by edges. The fundamental mathematical representation of network is the *adjacency* matrix[45]. Since we consider causal interactions in two directions, the networks are directed, where each edge has a direction. The adjacency matrix **A** of the directed networks is defined to be the $n \times n$ matrix with elements A_{ij} such that:

$$A_{ij} = \begin{cases} 1 \text{if there is an edge from} i \text{to} j \\ 0 \text{otherwise} \end{cases} \tag{7}$$

In general, the adjacency matrix of a directed network is asymmetric, since the existence of an edge from i to j does not necessarily imply there is also an edge from j to i. In a directed network, each node has two degrees: the *indegree* is the number of ingoing edges connected to a node and the *outdegree* is the number of outgoing edges. The in- and outdegrees of node i can be written as

$$k_i^{in} = \sum_{j=1}^{n} A_{ij} \tag{8}$$

$$k_j^{out} = \sum_{i=1}^{n} A_{ij} \tag{9}$$

where n is the total number of nodes in the directed network. To measure the importance of nodes in the directed network, we use the *PageRank* centrality:

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta \tag{10}$$

For nodes with no outgoing edges, the contribution to the centrality of any other node is zero. Practically, k_j^{out} is set as one or any non-zero value as A_{ij} is always zero if j has no outgoing edges. The constant term α should be less than 1 and set to 0.85 in calculations based on experimentation, and β is set to 1[45]. For the construction of climate network, each urban site is treated as an individual node of the network ($n = 520$), and the causal connection between them represents the edge within the network. The causality by the CCM method is used to construct directed links/edges here: the entity of the adjacency matrix **A** is one when the causality strength exceeds a threshold, and zero otherwise. The determination of the threshold is done using the distribution of causality in statistics (details in Supplementary Figs. 7 and 8).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The 0.5 °× 0.5° CPC Global Unified Temperature is available from NOAA Physical Sciences Laboratory (PSL) at https://www.psl.noaa.gov/data/gridded/data.cpc.globaltemp.html. Boundaries of urban areas are

retrieved from the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) database available at https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html .The population data are obtained from the 100-m WorldPop dataset available at https://hub.worldpop.org/geodata/summary?id=24863. Population density is derived based on population counts and spatial extent of urban areas.

Code availability

All relevant computer codes are available from the authors upon request.

Received: 14 August 2023; Accepted: 18 February 2024;
Published online: 27 February 2024

References

1. Spengler, O. *The Decline of the West, Vol. II: Perspective of World-History*. (Rogue Scholar, 2020).
2. UN-Habitat. *World Cities Report 2020: The Value of Sustainable Urbanization*. (UN-Habitat, 2020).
3. United Nations. *World urbanization prospects: the 2018 revision*. (United Nations, 2019).
4. Antognelli, S. & Vizzari, M. Ecosystem and urban services for landscape liveability: A model for quantification of stakeholders' perceived importance. *Land Use Policy* **50**, 277–292 (2016).
5. Kumar, P. et al. The nexus between air pollution, green infrastructure and human health. *Environ. Int.* **133**, 105181 (2019).
6. Wang, Z.-H. Compound environmental impact of urban mitigation strategies: Co-benefits, trade-offs, and unintended consequence. *Sustain. Cities Soc.* **75**, 103284 (2021).
7. Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D. G. & Salcedo-Sanz, S. Heat waves: physical understanding and scientific challenges. *Rev. Geophys.* **61**, e2022RG000780 (2023).
8. Abatzoglou, J. T. & Barbero, R. Observed and projected changes in absolute temperature records across the contiguous United States. *Geophys. Res. Lett.* **41**, 6501–6508 (2014).
9. Mazdiyasni, O. & AghaKouchak, A. Substantial increase in concurrent droughts and heatwaves in the United States. *Proc. Nat. Acad. Sci.* **112**, 11484–11489 (2015).
10. Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**, 994–997 (2004).
11. Matsueda, M. Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.* **38**, L06801 (2011).
12. Lorenz, R., Jaeger, E. B. & Seneviratne, S. I. Persistence of heat waves and its link to soil moisture memory. *Geophys. Res. Lett.* **37**, L09703 (2010).
13. Arblaster, J. M. & Alexander, L. V. The impact of the El Niño-Southern Oscillation on maximum temperature extremes. *Geophys. Res. Lett.* **39**, L20702 (2012).
14. Kenyon, J. & Hegerl, G. C. Influence of modes of climate variability on global temperature extremes. *J. Climate* **21**, 3872–3889 (2008).
15. Hansen, G. & Stone, D. Assessing the observed impact of anthropogenic climate change. *Nat. Climate Change* **6**, 532–537 (2016).
16. Yang, J., Wang, Z. H. & Huang, H. P. Intercomparison of the surface energy partitioning in CMIP5 simulations. *Atmosphere* **10**, 602 (2019).
17. Wang, L., Wang, L., Li, Y. & Wang, J. A century-long analysis of global warming and earth temperature using a random walk with drift approach. *Dec. Anal. J.* **7**, 100237 (2023).
18. Meehl, G. A., Arblaster, J. M. & Tebaldi, C. Contributions of natural and anthropogenic forcing to changes in temperature extremes over the United States. *Geophys. Res. Lett.* **34**, L19709 (2007).
19. Cowan, T., Undorf, S., Hegerl, G. C., Harrington, L. J. & Otto, F. E. L. Present-day greenhouse gases could cause more frequent and longer Dust Bowl heatwaves. *Nat. Climate Change* **10**, 505–510 (2020).
20. Diffenbaugh, N. S. & Ashfaq, M. Intensification of hot extremes in the United States. *Geophys. Res. Lett.* **37**, L15701 (2010).
21. Mora, C. et al. Global risk of deadly heat. *Nat. Climate Change* **7**, 501–506 (2017).
22. Fitzpatrick, M. C. & Dunn, R. R. Contemporary climatic analogs for 540 North American urban areas in the late 21st century. *Nat. Commun.* **10**, 614–617 (2019).
23. Seto, K. C. et al. Urban land teleconnections and sustainability. *Proc. Nat. Acad. Sci.* **109**, 7687–7692 (2012).
24. Wang, C., Wang, Z.-H. & Li, Q. Emergence of urban clustering among U.S. cities under environmental stressors. *Sustain. Cities Soc.* **63**, 102481 (2020).
25. Boers, N., Bookhagen, B., Marwan, N. & Kurths, J. Spatiotemporal characteristics and synchronization of extreme rainfall in South America with focus on the Andes Mountain range. *Clim. Dyn.* **46**, 601–617 (2016).
26. Boers, N. et al. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* **566**, 373–377 (2019).
27. Mondal, S. & Mishra, A. K. Complex networks reveal heatwave patterns and propagations over the USA. *Geophys. Res. Lett.* **48**, e2020GL090411 (2021).
28. Mondal, S., K Mishra, A., Leung, R. & Cook, B. Global droughts connected by linkages between drought hubs. *Nat. Commun.* **14**, 144 (2023).
29. Ludescher, J. et al. Network-based forecasting of climate phenomena. *Proc. Nat. Acad. Sci.* **118**, e1922872118 (2021).
30. Yang, X., Wang, Z.-H. & Wang, C. Critical transitions in the hydrological system: early-warning signals and network analysis. *Hydrol. Earth Syst. Sci.* **26**, 1845–1856 (2022).
31. Wallace, J. M. & Gutzler, D. S. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Rev.* **109**, 784–812 (1981).
32. Donges, J. F., Zou, Y., Marwan, N. & Kurths, J. Complex networks in climate dynamics. *Eur. Phys. J. Special Topics* **174**, 157–179 (2009).
33. Runge, J. et al. Identifying causal gateways and mediators in complex spatio-temporal systems. *Natu. Commun.* **6**, 8502 (2015).
34. Yang, X., Wang, Z.-H., Wang, C. & Lai, Y.-C. Detecting the causal influence of thermal environments among climate regions in the United States. *J. Environ. Manag.* **322**, 116001 (2022).
35. Konapala, G. & Mishra, A. Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *J. Hydrol.* **555**, 600–620 (2017).
36. Wang, Z.-H., Wang, C. & Yang, X. Dynamic synchronization of extreme heat in complex climate networks in the contiguous United States. *Urban Climate* **38**, 100909 (2021).
37. Banerjee, A. et al. Spatial coherence patterns of extreme winter precipitation in the U.S. *Theor. Appl. Climatol.* **152**, 385–395 (2023).
38. Yang, X., Wang, Z.-H., Wang, C. & Lai, Y.-C. Finding causal gateways of precipitation over the contiguous United States. *Geophys. Res. Lett.* **50**, e2022GL101942 (2023).
39. Runge, J. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisc. J. Nonlinear Sci.* **28**, 075310 (2018).
40. Keellings, D. & Moradkhani, H. Spatiotemporal evolution of heat wave severity and coverage across the United States. *Geophys. Res. Lett.* **47**, e2020GL087097 (2020).
41. Schoof, J. T., Ford, T. W. & Pryor, S. C. Recent changes in U.S. regional heat wave characteristics in observations and reanalyses. *J. Appl. Meteorol. Climatol.* **56**, 2621–2636 (2017).
42. Kunkel, K. E., Liang, X.-Z., Zhu, J. & Lin, Y. Can CGCMs simulate the twentieth-century "warming hole" in the central United States? *J. Climate* **19**, 4137–4153 (2006).
43. Hoerling, M. et al. Anatomy of an extreme event. *J. Climate* **26**, 2811–2832 (2013).

44. Hong, S.-Y. & Kalnay, E. Role of sea surface temperature and soil-moisture feedback in the 1998 Oklahoma–Texas drought. *Nature* **408**, 842–844 (2000).
45. Newman, M. *Networks*. (Oxford University Press, 2018).
46. Tsonis, A. A., Swanson, K. L. & Roebber, P. J. What do networks have to do with climate? *Bull. Am. Meteorol. Soc.* **87**, 585–596 (2006).
47. Luo, X., Vahmani, P., Hong, T. & Jones, A. City-scale building anthropogenic heating during heat waves. *Atmosphere* **11**, 1206 (2020).
48. Good, P. et al. Nonlinear regional warming with increasing CO2 concentrations. *Nat. Clim. Change* **5**, 138–142 (2015).
49. Cai, W. et al. Increased frequency of extreme La Niña events under greenhouse warming. *Nat. Clim. Change* **5**, 132–137 (2015).
50. Li, P., Yu, Y., Huang, D., Wang, Z.-H. & Sharma, A. Regional heatwave prediction using graph neural network and weather station data. *Geophys. Res. Lett.* **50**, e2023GL103405 (2023).
51. Sun, Q. et al. A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev. Geophys.* **56**, 79–107 (2018).
52. Smith, A., Lott, N. & Vose, R. The integrated surface database: recent developments and partnerships. *Bull. Am. Meteorol. Soc.* **92**, 704–708 (2011).
53. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
54. Sugihara, G. & May, R. M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**, 734–741 (1990).
55. Takens, F. Detecting strange attractors in fluid turbulence. in *Dynamical Systems and Turbulence* (eds. Rand, D. & Young, L. S.) 898 366–381 (Springer-Verlag, 1981).
56. Jiang, J.-J., Huang, Z.-G., Huang, L., Liu, H. & Lai, Y.-C. Directed dynamical influence is more detectable with noise. *Sci. Rep.* **6**, 24088 (2016).
57. Grassberger, P. & Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349 (1983).
58. Grassberger, P. & Procaccia, I. Measuring the strangeness of strange attractors. *The Theory of Chaotic Attractors* (eds. Hunt, B. R., Li, T.-Y., Kennedy, J. A. & Nusse, H. E.) 170–189 (Springer, 2004).
59. Lai, Y.-C. & Ye, N. Recent developments in chaotic time series analysis. *Int. J. Bifurc. Chaos* **13**, 1383–1422 (2003).

Acknowledgements

Z.H.W. is supported by the U.S. National Science Foundation (NSF) under grant No. AGS-2300548. C.W. is supported by the U.S. NSF under grant No. OIA-2327435. Y.C.L. is supported by the Air Force Office of Scientific Research through Grant No. FA9550-21-1-0438.

Author contributions

Z.-H. Wang and Y.-C. Lai conceived the idea. Z.-H. Wang acquired the fund and supervised the research. C. Wang and X. Yang retrieved and processed the data. X. Yang conducted the simulations and drafted the manuscript. All authors revised the manuscript. Z.-H. Wang finalized the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42949-024-00148-x.

Correspondence and requests for materials should be addressed to Zhi-Hua Wang.