SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
APPLIED PHYSICS
STATISTICAL PHYSICS, THERMODYNAMICS AND NONLINEAR DYNAMICS
PHYSICS
THEORETICAL PHYSICS

Received
2 April 2012

Accepted
9 May 2012

Published
14 June 2012

Correspondence and requests for materials should be addressed to H.A.M. (hmakse@lev.ccny.cuny.edu)

Collective behavior in the spatial spreading of obesity

Lazaros K. Gallos[1], Pablo Barttfeld[2], Shlomo Havlin[3], Mariano Sigman[2] & Hernán A. Makse[1,2]

[1]Levich Institute and Physics Department, City College of New York, New York, NY 10031, USA, [2]Integrative Neuroscience Laboratory, Physics Department, FCEyN, Universidad de Buenos Aires, Buenos Aires, Argentina, [3]Minerva Center and Physics Department, Bar-Ilan University, Ramat Gan 52900, Israel.

Obesity prevalence is increasing in many countries at alarming levels. A difficulty in the conception of policies to reverse these trends is the identification of the drivers behind the obesity epidemics. Here, we implement a spatial spreading analysis to investigate whether obesity shows spatial correlations, revealing the effect of collective and global factors acting above individual choices. We find a regularity in the spatial fluctuations of their prevalence revealed by a pattern of scale-free long-range correlations. The fluctuations are anomalous, deviating in a fundamental way from the weaker correlations found in the underlying population distribution indicating the presence of collective behavior, i.e., individual habits may have negligible influence in shaping the patterns of spreading. Interestingly, we find the same scale-free correlations in economic activities associated with food production. These results motivate future interventions to investigate the causality of this relation providing guidance for the implementation of preventive health policies.

The World Health Organization has recognized obesity as a global epidemic[1]. Obesity heads the list of non-communicable diseases (NCD) like diabetes and cancer, for which no prevention strategy has managed to control their spreading[2–7]. Since the gain of excessive body weight is related to an increase in calories intake and physical inactivity[8–10] a principal aspect of prevention has been directed to individual habits[11]. However, the prevalence of NCDs shows strong spatial clustering[12–14]. Furthermore, obesity spreading has shown high susceptibility to social pressure[6] and global economic drivers[3–5,7]. This suggests that the spread and growth of obesity and other NCDs may be governed by collective behavior acting over and above individual factors such as genetics and personal choices[4,5].

To study the emergence of collective dynamics in the spatial spreading of obesity and other NCDs, we implement a statistical clustering analysis based on the physics of critical phenomena. We start by investigating regularities in obesity spreading derived from correlation patterns of demographic variables. Obesity is determined through the Body Mass Index (BMI) obtained via the formula weight(kg)/[height (m)]^2. The obesity prevalence is defined as the percentage of adults aged ≥ 18 years with a BMI ≥ 30. We investigate the spatial correlations of obesity prevalence in the USA during a specific year using microdata defined at the county-level provided by the US Centers for Disease Control (CDC)[14] through the Behavioral Risk Factor Surveillance System (BRFSS) from 2004 to 2008 (see Methods section). The average percentage of obesity in USA was historically around 10%. In the early 80's, an obesity transition in the hitherto robust percentage, steeply increased the obesity prevalence (Fig. 1a).

Results

Spatial correlations. The spatial map of obesity prevalence in the USA shows that neighboring areas tend to present similar percentages of obese population[14] forming spatial 'obesity clusters'[12,13]. The evolution of the spatial map of obesity from 2004 to 2008 at the county level (Fig. 1b) highlights the mechanism of cluster growth. Characterizing such geographical spreading presents a challenge to current theoretical physics frameworks of cluster dynamics[15–22]. The properties of such spatial arrangement are determined by the equal-time two-point correlation function, $C(r)$, measuring the influence of an observable x_i in county i (e.g., in this study: population density, prevalence of adult obesity and diabetes, cancer mortality rates and economic activity) on another county j at distance r[15]:

$$C(r) \equiv \frac{1}{\sigma^2} \frac{\sum_{ij} (x_i - \overline{x})(x_j - \overline{x}) \delta(r_{ij} - r)}{\sum_{ij} \delta(r_{ij} - r)}. \tag{1}$$

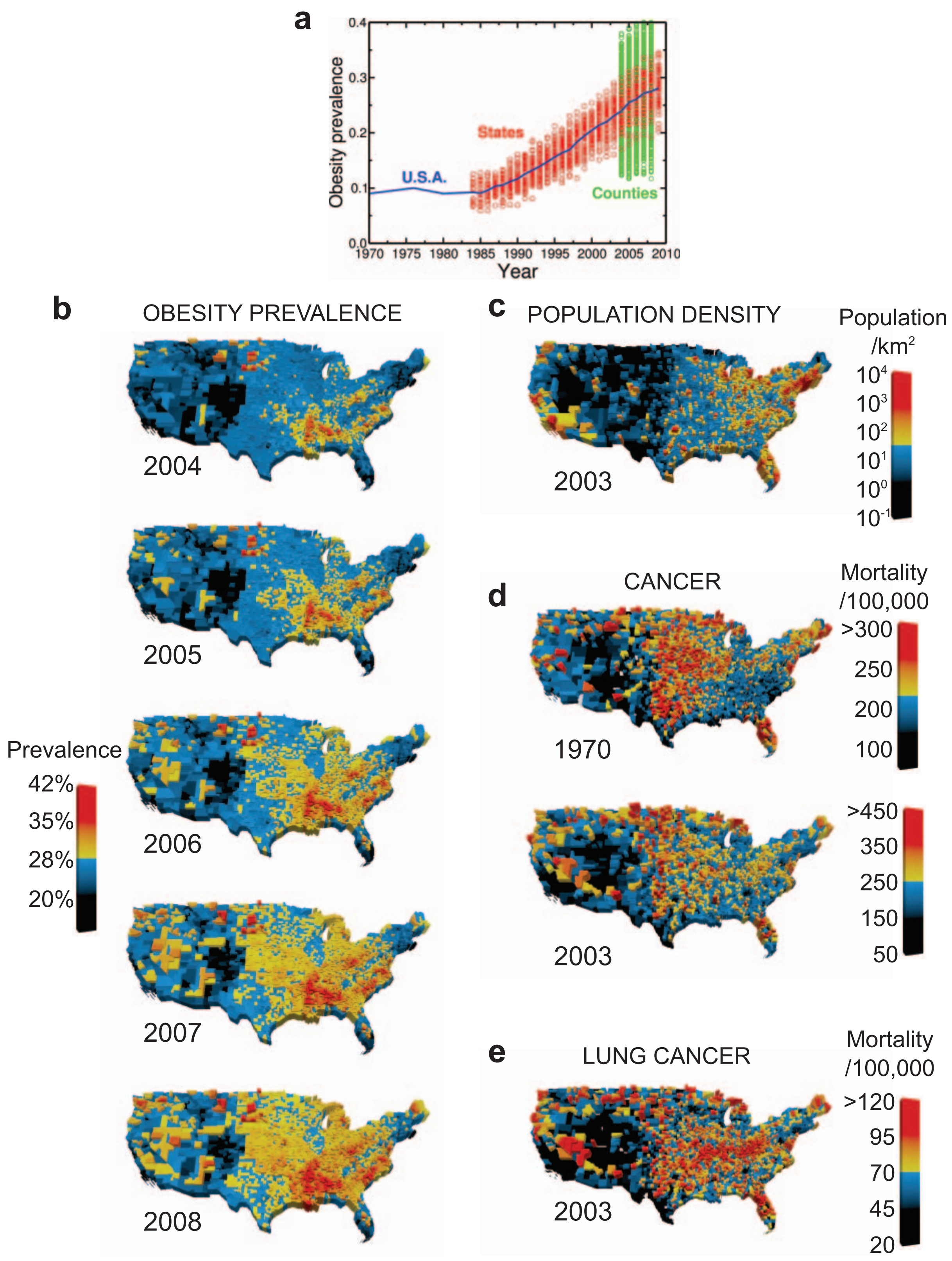

Figure 1 | The obesity transition. (a) CDC[14] provides an estimate of the number of obese adults, based on self-reported weight and height, country-wide since 1970 (blue line), at the state level from 1984 to 2009 (red symbols), and at the county level from 2004 to 2008 (green symbols). A transition is observed around 1980. We base our analysis on the micro-data at the county level. (b) Map of the spatial spreading of obesity prevalence evidencing clustering dynamics. (c) Map of the population density defined at the county level in 2003 showing correlated patterns albeit with less clustering than in obesity. (d) Map of cancer mortality rates per county in 1970 and 2003 visualizing the transition from high correlations and clustering to weak correlation and more uniformity in 2003. (e) Map of lung cancer mortality per county indicating large clustering properties similar to obesity.

Here, $\overline{x}$ is the average over $N = 3,092$ counties in the contiguous USA, $\sigma^2 = \sum_i (x_i - \overline{x})^2 / N$ is the variance, and r_{ij} is the euclidean distance between the geometrical centers of counties i and j. The delta function selects counties whose centers are at a distance r. Large positive values of $C(r)$ reveal strong correlations, while negative values imply anti-correlations, i.e., two areas with opposed tendencies relative to the mean in obesity prevalence (analogous to two domains with opposite spins in a ferromagnet[15]).

Spatial correlations in any indicator ought to be referred to the natural correlations of population fluctuations (Fig. 1c). To this aim, we first calculate $C(r)$ for the population in USA counties, p_i, by using the density: $x_i = p_i/a_i$ in Eq. (1), where a_i is the county area. Population density correlations show a slow fall-off with distance (Fig. 2a) approximately described by a power-law up to a correlation length ξ:

$$C(r) \sim r^{-\gamma}, \quad r \lesssim \xi, \qquad (2)$$

where γ is the correlation exponent. Correlations become short-ranged when $\gamma \geq d$ ($d = 2$ is the dimension of the map), and stronger as γ decreases[15,16]. An Ordinary Least Squares (OLS) regression analysis[23] on the population reveals the exponent $\gamma = 1.01 \pm 0.08$ (a value that is the average over the individual exponents for years 1969–2009, Fig. 2a, error bars denote 95% confidence interval [CI]). For the fitting, we adopt standard procedures for functional forms like Eq. (2)[24] where we vary the minimum and maximum values of the fitting interval, monitoring the value of R^2 that optimizes the fitting area (see details in the Methods section and SI) in order to calculate the exponent γ. The same plot in linear axes, Fig. 2b reveals a distance where correlations vanish, $C(\xi) = 0$ with $\xi = 1050$ km, representing the average size of the correlated domains[25]. As we increase r larger than ξ, we consider correlations between areas in the East and West which are anti-correlated since $C(r) < 0$ for $r > \xi$.

In a typical analysis of empirical data, the possible extent of correlations is restricted by the finite system size. Even when long-range correlations are known to be present, a cut-off value will eventually emerge. We call this cut-off value the correlation length, ξ. It is expected that the value of ξ is related to the system size. A stringent test for the existence of scale-free correlations, such as those appearing in critical systems, is through finite-size scaling analysis, where we test the behavior of ξ as a function of the system size. If ξ is fixed and does not change when the system size increases, then any correlations that exist cannot be scale-free. The idea of scale-free correlations implies that, for finite systems, correlations are of the order of the system size and the value of ξ increases monotonically as we move to larger systems.

The finite-size scaling analysis requires the study of independent systems of different sizes. Here, we use high-resolution population data for 50 countries and calculate the value of ξ in each case (see results in Supplementary Table S1). To determine whether population correlations are scale-free, we calculate $C(r)$ for geographical systems of different sizes using a high resolution grid of 2.5 arc-seconds, available for several countries from Ref. [26] (see Methods section). The resulting correlations (Fig. 2d) reveal the same picture as for the USA at the county-level (Fig. 2a), i.e., a power-law up to a correlation length. We then measure ξ for every country, and investigate whether, as expected with the laws of critical phenomena[27], it increases with the country size, L. Indeed, we obtain (Fig. 2e and Supplementary Table S1),

$$\xi(L) \sim L^{\nu}, \qquad (3)$$

where $\nu = 0.9 \pm 0.1$ is the correlation length exponent[15]. This result implies that the fluctuations in human agglomerations are scale-free, i.e., the only length-scale in the system is set by its size and the correlation length becomes infinite when $L \to \infty$[15,25,27].

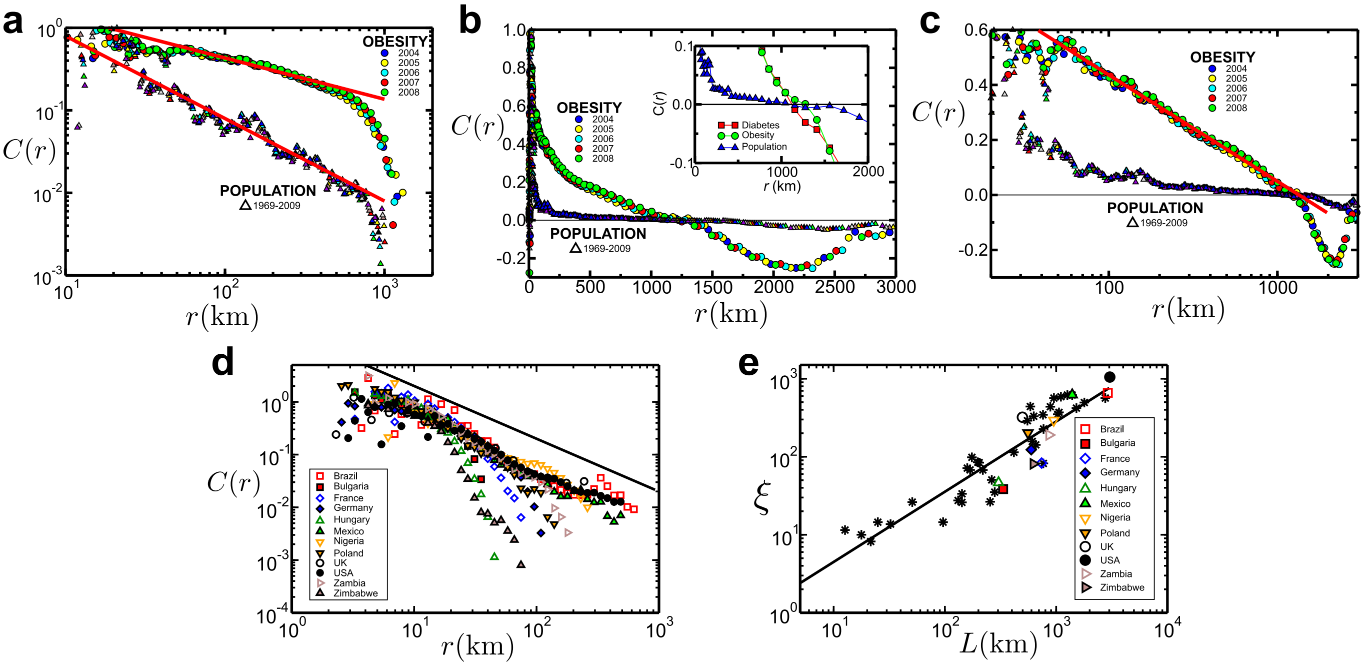

Figure 2 | Long-range correlations in spreading phenomena. (a) Correlation function, $C(r)$, averaged over counties at distance r for population density from 1969–2009 and obesity prevalence from 2004–2008. The lines are fittings based on OLS regression analysis[23,24]. (b) Correlation function, $C(r)$, for population density and obesity as in (a) above, but in linear axes. The plot shows the correlation length, ξ, at $C(\xi) = 0$ and highlights the fact that ξ is approximately the same for the population, obesity and diabetes prevalence (data for 2004). The plot also highlights the anticorrelations for $r > \xi$. The inset zooms in the area around $C(\xi) = 0$. (c) Correlation function, $C(r)$, for population density and obesity as in (a) above, but in log-linear axes. The plot is compatible with logarithmic decay for the obesity correlation function $C(r) \sim \ln(r_0/r)$, where $r_0 = 1307$ km (the continuous line indicates this fitting). The population density decays faster than that and cannot be described by a similar function. (d) Population density correlation function, $C(r)$ vs r, for different countries in 2009 as indicated. (e) Correlation length ξ vs linear country size L for different countries. The symbols indicate the same countries as in Fig. 2d. The remaining star symbols are for other countries as indicated in Supplementary Table S1. L is the square root of the total area of the country.

We interpret any departure from $\gamma = 1$ as a proxy of anomalous dynamics beyond the simple dynamics related to the population growth. When we calculate the spatial correlations of obesity prevalence ($s_i \equiv o_i/p_i$, o_i is the number of obese adults in county i) in USA from 2004 to 2008 we also find long-range correlations (Fig. 2a). The crux of the matter is that the correlation exponent for obesity ($\gamma = 0.50 \pm 0.04$, averaged over the individual exponents for years 2004–2008 with an average $R^2 = 0.96$) is smaller than that of the population, signaling anomalous growth. Since smaller exponents mean stronger correlations, the increase in obesity prevalence in a given place can eventually spread significantly further than expected from the population dynamics.

The small γ exponent of obesity (in comparison with $\gamma = 2$, the uncorrelated value) indicates a very slow decay of obesity correlations. In such cases the exact value of γ may not be very accurate. This is common behavior for systems with correlation exponent close to $\gamma = 0$; we notice that a similar scale-free correlation function with exponent $\gamma \approx 0$ was found in the velocity fluctuations in bird flocks[25]. Furthermore, the limiting case of $\gamma \to 0$ is equivalent to a slow logarithmic decay: both cases, small γ and logarithmic decay imply the existence of long-range correlations. Indeed, Fig. 2c suggests that a slow logarithmic dependence can also describe the variation of correlation with distance in obesity prevalence. In fact, a fitting to a logarithmic function $C(r) \sim \ln(r_0/r)$ gives $r_0 = 1307$ km with $R^2 = 0.99$, similar to the R^2 value obtained by a power-law fitting. The value of r_0 is another estimation of the obesity correlation length, ξ, which is of the same order of magnitude as the population correlation length. The natural noise in the empirical data and the small system size do not allow to accurately distinguish between power-law with small exponent and logarithmic fittings for obesity. In either case, though, both a power-law exponent of $\gamma = 0.5$ and a logarithmic decay (which represents the limit of $\gamma = 0$) indicate the presence of strong and long-range scale-free correlations. These are in sharp contrast to the exponent $\gamma = 1$ of population density correlations, as is evident from Fig. 2c where $C(r)$ for population approaches zero much faster than logarithmically. In what follows, we report the correlations in terms of exponents rather than the equivalent logarithmic decay.

We also calculate fluctuations in variables which are known to be strongly related to obesity[8,9,12,28]: diabetes and physical inactivity prevalence (fraction of adults per county who report no physical activity or exercise, see Methods section). The obtained γ exponents are anomalous with similar values as in obesity (Fig. 3a). The system size dependence of ξ for obesity and diabetes cannot be measured directly, since there is no available micro-data for other countries, analogous to the ones in the USA. However, we find that the value of

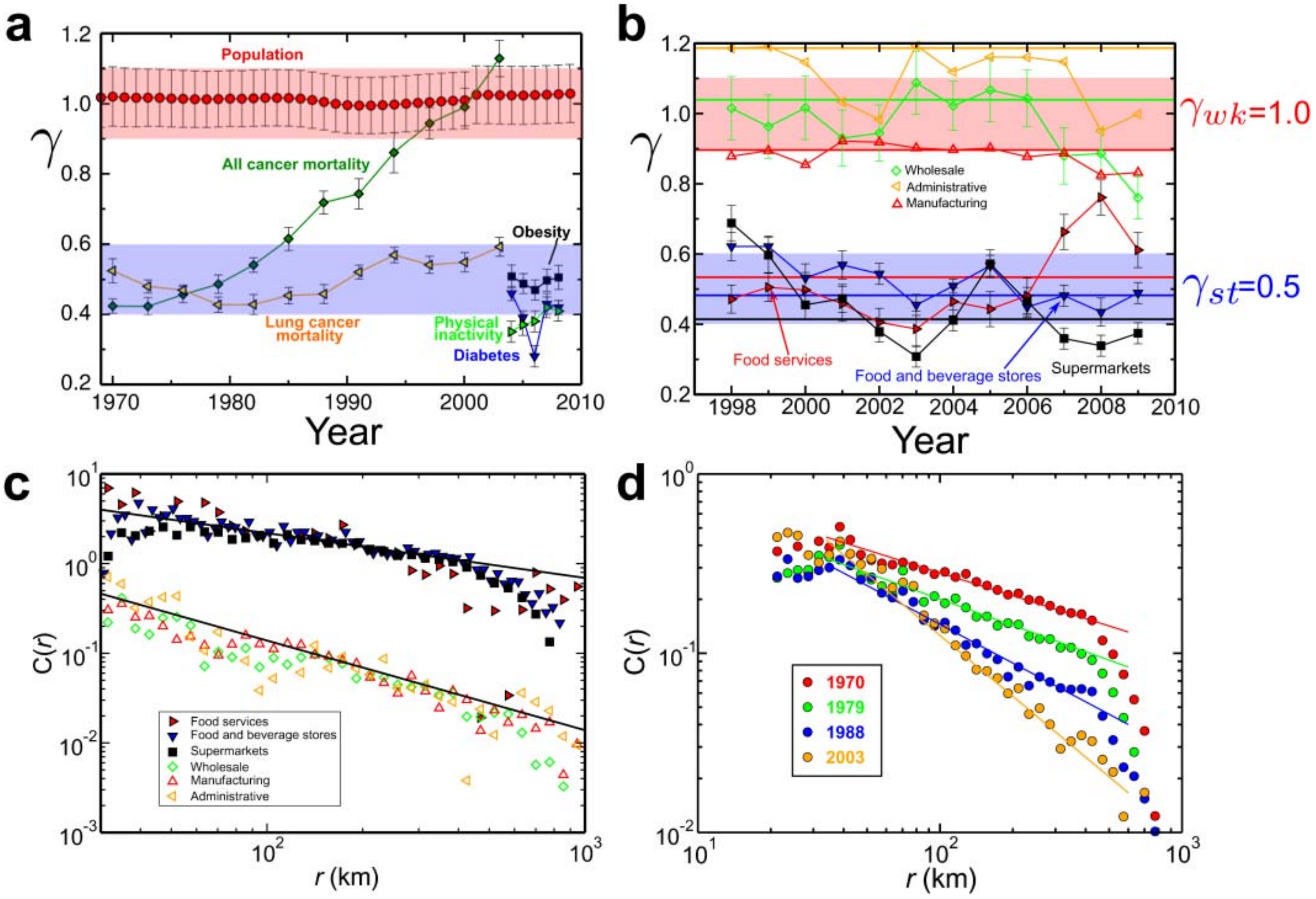

Figure 3 | Correlation exponents. (a) Temporal evolution of γ for population distribution, obesity, diabetes, physical inactivity, all cancer mortality, and lung cancer mortality per county. The diagram displays the classes of strong correlations, $\gamma_{\rm st} = 0.5$, and weak correlations, $\gamma_{\rm wk} = 1$. Additionally, theory predicts $\gamma_{\rm rnd} \geq 2$ for uncorrelated systems. We did not observe any human activity or indicators whose correlations fall within this class, unless the data of different counties is shuffled. (b) Evolution of γ for different economic indicators describing the food industry and generic economic sectors as indicated. We quantify economic activity by the total number of employees of a given sector per county population. Horizontal lines represent the fitted exponent value of a global correlation curve, averaged over all years. (c) Correlation functions for the economic activities indicated in the figure. The plot shows the segregation of the data into two classes. For clarity, the curves for food industry have been vertically shifted by a decade. The solid lines indicate $\gamma_{\rm wk} = 1$ and $\gamma_{\rm st} = 1/2$. (d) Change in $C(r)$ for cancer mortality rates in the period 1970–2003.

ξ for obesity and diabetes in USA is very close to ξ of the population distribution, as shown above (inset of Fig. 2b). Assuming that the equality of the correlation lengths holds also for other countries, then obesity and diabetes should satisfy Eq. (3) as well. Thus, we expect that the correlations in obesity and diabetes may become scale-free in the infinite system size limit.

The form of the correlations in obesity are reminiscent of those in physical systems at a critical point of second-order phase transitions[15,25,27]. Physical systems away from criticality are uncorrelated and fluctuations in observables, e.g., magnetization in a ferromagnet or density in a fluid, decay faster than a power-law, e.g., exponentially[15,27]. Instead, long-range correlations appear at critical points of phase transitions where fluctuations are not independent and, as a consequence, fall-off more slowly. The existence of long-range correlations with $\gamma = 0.5$ — rather than the noncritical exponential decay — may signal the emergence of strong critical fluctuations in obesity and diabetes spreading. The notion of criticality, initially developed for equilibrium systems[15,27], has been successfully extended to explain a wide variety of dynamics away from equilibrium ranging from collective behavior of bird cohorts, biological and social systems to city growth, just to name a few[25,27,29,30] (it is interesting to note that the shape of the correlation function in obesity is similar to the scale-free correlations found in the velocity fluctuations in starling flocks, see Fig. 2 in Ref. [25]). Its most important consequence is that it characterizes a system for which local details of interactions have a negligible influence in the global dynamics[15,27]. Following this framework, the clustering patterns of obesity are interpreted as the result of collective behavior which may not merely be the consequence of fluctuations of individual habits.

It should be noticed that criticality is not the only possible dynamics leading to power-law correlations. A system at criticality will necessary develop scale-free correlations which allow all system subparts to feel the influence of far-away system areas. The existence of a power-law correlation function, though, does not necessarily imply the existence of criticality. For example, in the two-dimensional XY model, power-law correlations exist below the critical temperature with a temperature dependent exponent, i.e. in a non-critical phase[31]. The idea of criticality can be tested more stringently by showing the existence of a number of critical properties. The critical length should diverge with increasing system size, which we already showed to be true, or the susceptibility should diverge, i.e. external perturbations should lead to a diverging response function at the critical point. In the case of obesity such perturbations are very difficult to observe, but it is still possible for a future study to monitor changes in obesity spreading under particular perturbations. For instance, the introduction of a new health policy or a food industry regulation may allow the study of how these external factors influence obesity levels. Currently, we can only suggest that the present analysis is compatible with the idea of criticality, and further studies are needed to actually prove the existence of criticality in obesity spreading.

This finding is in analogy with the behavior of bird flocks[25] or brain dynamics[27,32]. In these studies, long-range correlations were found in the velocity fluctuations of bird flocks and in the activity of the brain obtained via fMRI, respectively. The correlations were attributed to the presence of enough noise to drive the system to a critical phase. For instance, the noise in bird flocks could be a result of random errors or computational mistakes in the calculation of directionality by individual birds, with the resulting total error finely tuned to bring the system at criticality. Criticality in the brain might be related to an optimization of information transfer. In obesity spreading, the order parameter of the system is the obesity prevalence, but there is no obvious method to control this parameter. Similarly, in neuron networks[27,32] an analogy was found with the Ising model, where the main parameters of the model, such as the exchange interaction, could be directly calculated experimentally. The Ising model allowed the study of properties, such as the divergence of the heat capacity, that provide strong evidence in favor of criticality. In the case of obesity spreading, the indications that we have for criticality are based on Eq. (3), where ξ increases with L, i.e. on the existence of scale-free correlations that diverge as the system size increases. Further studies may be needed to explore analogies with statistical models, similarly with the above referenced works.

The underlying hypothesis is that the correlations of fluctuations observed in the obesity prevalence may be inherited by specific demographic and economic variables which are thought to be related to the rise of obesity[4,5]. As a tentative way of addressing which elements of the economy may be related to the obesity spread, we calculate γ in economic indicators related to obesity[4,5]. Except for transient phenomena, all studied indicators yield exponents that fall around $\gamma_{\rm wk} = 1$ or $\gamma_{\rm st} = 1/2$, representing two universality classes of weak and strong correlations, respectively (Figs. 3a and 3b).

We begin by studying the correlations in generic sectors of the economy (measured through the number of employees in an economic sector per county population, see Methods section). We find γ close to $\gamma_{\rm wk} = 1$ (over the period 1998–2009, Fig. 3b and c) for sectors which are not related to obesity, e.g., wholesalers, administration, and manufacturing. This suggests that generic sectors of the economy inherit the correlations in the population (Figs. 3b and c).

Interestingly, analysis of the spatial fluctuations in the economic activity of sectors associated to food production and sales (supermarkets, food and beverages stores and food services such as restaurants and bars) gives rise to the same anomalous value as obesity and diabetes ($\gamma_{\rm st} = 1/2$, 1998–2009, Fig. 3b and c). Although these results cannot inform about the causality of these relations, they show that the scaling properties of the obesity patterns display a spatial coupling which is also expressed by the fluctuations of sectors of the economy related to food production.

It is of interest to study other health indicators for which active health policies have been devoted to control the rate of growth. We apply the correlation analysis to lung cancer mortality defined at the county level and compare with mortality due to all types of cancer (see Methods section). The spatial correlations of cancer mortality per county show an interesting transition in the late 70′s from anomalous strong correlations, $\gamma_{\rm st} = 1/2$, to weak correlations, $\gamma_{\rm wk} = 1$, (Fig. 3a and 3d). This transition is visualized in the different correlated patterns of cancer mortality in 1970 and 2003 in Fig. 1d, i.e., the clustering of the data is more profound in 1970, while in 2003 it spreads more uniformly. This behavior raises the intriguing possibility that the anomalous strongly-correlated dynamics of the past have been smoothed out with time. The current status of all-cancer mortality fluctuations is close to the natural one, inflicted by population correlation. Conversely, fluctuations in the mortality rate due to lung cancer from 1970 to 2003 have remained highly correlated and close to the obesity value, $\gamma_{\rm st} = 1/2$ (Fig. 3a and 1e), while the other types of cancer have become less correlated. This is an interesting finding since lung cancer prevalence, similarly to obesity, is affected by a global factor (smoking) and has been growing rapidly during the studied period. A question for future research is whether the strong scale-free correlations in indicators like obesity, diabetes and lung cancer may explain the fast growth of the indicators in comparison with the population. Studies of scale dependence of the growth rates might shed light to this question[33].

Evolution of obesity clusters near percolation. The most visible characteristic of correlations is the formation of spatial clusters of obesity prevalence. To quantitatively determine the geographical formation of obesity clusters, we implement a percolation analysis[16–22,34]. The control parameter of the analysis is the obesity threshold, s. An obesity cluster is a maximally connected set of counties for which s_i exceeds a given threshold s: $s_i \geq s$. By decreasing s, we

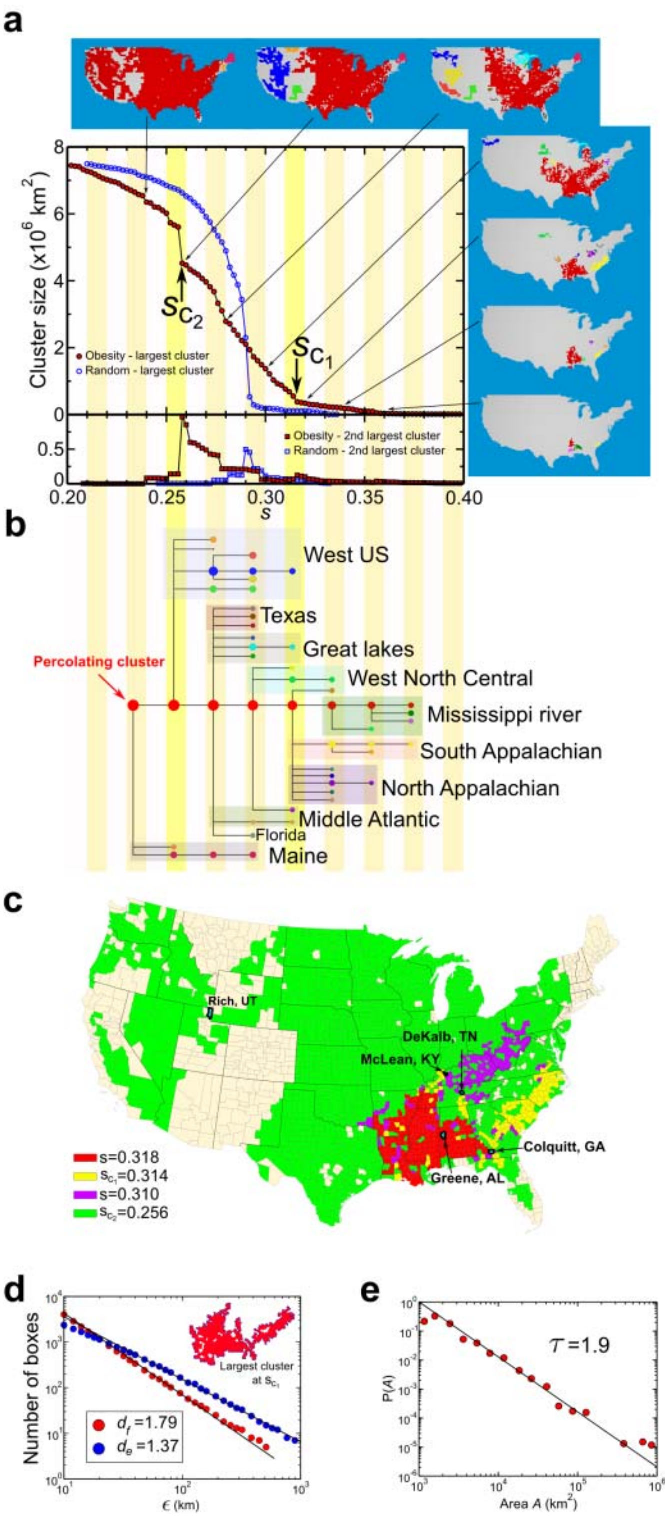

Figure 4 | Percolation picture of obesity. (a) Size of the first (circles) and second (squares) largest components as a function of the obesity prevalence threshold *s* in 2008. As we lower *s*, the largest component increases abruptly indicating absorption of whole clusters, as also evidenced by the peaks in the second largest cluster[16]. We observe two main transitions at s_{c_1} and s_{c_2} in the real data (red) and a single second-order transition in the randomized data (blue). The maps show the progression of the obesity clusters with at least 5 counties for a given *s*. (b) Percolation tree representing the hierarchical formation, growth and merging of obesity clusters. Each dot represents a cluster at a given *s* with a size proportional to the logarithm of the cluster's area. Cluster colors follow Fig. 4a and we indicate their geographic regions. As we lower *s* from right to left, regions of high obesity prevalence appear first in the tree. The main percolating cluster starts in the lower Mississippi basin (red) at high *s* and absorbs clusters until percolating through all US. In particular, we note the two main transitions at s_{c_1}, where it absorbs the two Appalachian clusters, and at s_{c_2}, where it absorbs the West US cluster. (c) Detail of the evolution of obesity clusters near percolation as indicated. The map shows the shape of the first (red), second (yellow), and third (violet) clusters around s_{c_1}, and the largest (green) cluster at s_{c_2}, together with the location of the red bonds responsible for the transitions. The epicenter is Greene county, AL with 43.7% obesity prevalence. (d) Box fractal dimension of percolating cluster in the inset measured by the number of boxes of size ε needed to cover the cluster: $N_B(\epsilon) \sim \epsilon^{-d_f}$, and fractal dimension of the boundary measured by the number of boxes needed to cover the hull: $N_h(\epsilon) \sim \epsilon^{-d_e}$. (e) Probability distribution of the area of the obesity clusters, $P(A) \sim A^{-\tau}$, at percolation s_{c_1} averaged from 2004–2008. This scaling law generalizes Zipf's law[29] from urban to obesity clusters.

monitor the progressive formation, growth and merging of obesity clusters.

In random uncorrelated percolation[16], small clusters would be formed in a spatially uniform way until a critical value, s_c, is reached, and an incipient cluster spans the entire system. Instead, when we analyze the obesity clusters we observe a more complex pattern exemplified in Fig. 4a and 4b for year 2008. At large *s*, the first cluster appears in the lower Mississippi basin (red in Fig. 4a) with epicenter in Greene county, AL. Upon decreasing *s* to 0.32, new clusters are born including two spanning the South and North of the Appalachian Mountains, which acts as a geographical barrier separating the second and third largest clusters (yellow and violet in Fig. 4a, respectively). Further lowering *s*, we observe a percolation transition in which the Appalachian clusters merge with the Mississippi cluster. This point is revealed by a jump in the size of the largest component and a peak in the second largest component at $s_{c_1} = 0.314$ (Fig. 4a) as features of a percolation transition[16]. As a comparison, when we randomize the obesity data by shuffling the values between counties, a single critical point at $s_c = 0.29$ appears as a signature of an uncorrelated percolation process (blue symbols in Fig. 4a).

Obesity clusters in the West persist segregated from the main Eastern cluster avoiding a full-country percolation due to low-prevalence areas around Colorado state. Finally, the East and West clusters merge at $s_{c_2} = 0.256$ by a red bond (Rich county, Utah) producing a second percolation transition; this time spanning the whole country (see Fig. 4a and c, where the whole spanning cluster is green). This cluster-merging process is a hierarchical percolation progression represented in the tree model in Fig. 4b.

The shape of the main obesity clusters and location of the red bonds and obesity epicenter are depicted in Fig. 4c overlayed with a US map showing the boundaries of states and counties. Figure 4c shows the obesity clusters obtained at $s = 0.318$, $s_{c_1} = 0.314$, $s = 0.310$, and $s_{c_2} = 0.256$, depicting the process of percolation. At $s = 0.318$, we plot the largest red cluster which is seen in the lower Mississippi basin. The highest obesity prevalence is in Greene county, AL, which acts as the epicenter of the epidemic. At s_{c_1}, we plot in yellow the second largest cluster in the Atlantic region south of the Appalachian Mountains, and at $s = 0.310$ we plot the third largest cluster (violet), which appears north of the Appalachian Mountains. We mark with black the three red bonds that make the Mississippi cluster to grow abruptly by absorbing the clusters in the Appalachian range. The red bonds are DeKalb county, TN, McLean county, KY, and Colquitt county, GA. This transition is reflected in the jump in the size of the largest cluster in Fig. 4a. The same process is observed in the second percolation transition at s_{c_2}, when the red bond, Rich county, UT, joins the Eastern and Western clusters for a whole-country percolation.

Scaling exponents of percolation clusters. To further inquire whether the spreading of obesity has the features of a physical system at the critical point, we examine the geometry and distribution of obesity clusters. For long-range correlated critical systems percolating through nearest neighbors in two dimensional maps, the geometrical structure[16,19–22] gives rise to three critical exponents: the fractal dimension of the spanning cluster, d_f, the fractal dimension of the hull, d_e, and the cluster size distribution exponent, τ, analogous to Zipf's law[29]. These exponents can be calculated through the following methods:

(i) The scaling of the number of boxes N_B to cover the infinite spanning cluster versus the size of the boxes ϵ:

$$N_B(\epsilon) \sim \epsilon^{-d_f}, \tag{4}$$

defines the fractal dimension of the spanning cluster, d_f.

(ii) The number of boxes, N_h, of size ϵ covering the perimeter of the infinite cluster:

$$N_h(\epsilon) \sim \epsilon^{-d_e}, \tag{5}$$

defines the hull fractal dimension, d_e.

(iii) The probability distribution of the area of clusters at percolation:

$$P(A) \sim A^{-\tau}, \tag{6}$$

is characterized by the critical exponent τ. Additionally, there is a scaling relation between the fractal dimension and the cluster distribution exponent[16]: $\tau = 1 + 2/d_f$. This scaling law (6) is a generalization of Zipf's law[29] for urban populations to obese populations.

For the percolating obesity cluster at s_{c_1} displayed in the inset of Fig. 4d, we confirm critical scaling with exponents: $(d_f, d_e, \tau) = (1.79 \pm 0.08, 1.37 \pm 0.06, 1.9 \pm 0.1)$ (Fig. 4d, e).

The exponents (d_f, d_e, τ) for percolation with long-range correlations have been calculated numerically in Refs. [19–22] as a function of the correlation exponent γ using standard percolation analysis. There exists also a theoretical prediction based on Renormalization Group in Ref. [18] for the correlation length exponent. A direct computer simulation of long-range percolation[19–22] for $\gamma = 0.5$ finds the values of the three geometric exponents to be $(d_f, d_e, \tau) = (1.9 \pm 0.1, 1.39 \pm 0.03, 2.05 \pm 0.08)$, consistent with those reported here.

We notice that the exponent τ is expected to be larger than 2. This is due to mass conservation, assuming that the power-law Eq. (6) extends to infinity at percolation in an infinite system size. The fact that we find a value slightly smaller than 2 for the obesity clusters, might be due to a finite size effect. We also notice that the values of the exponents obtained from correlated percolation at $\gamma = 0.5$ are not too far from those of uncorrelated percolation[19]. Therefore, the values of the exponents may not be enough to precisely compare the obesity clusters with long-range percolation clusters. However, they serve as an indication that the obesity clusters have the geometrical properties of clusters at a critical point, such as scaling behavior. Furthermore, it could be possible that long-range correlated percolation may capture only part of the dynamics of the clustering epidemic. It could be, for instance, that higher order correlations, beyond the two-point correlation captured by $C(r)$, are also relevant in determining the value of the exponents. In this case, our analysis should be supplemented by studies of n–point correlation functions, beyond $C(r)$.

Covariance. The present approach is based on critical phenomena and attempts to classify dissimilar indicators (from health to economy) with universal scaling exponents (γ, ν, d_f, d_e, τ). Thus, our approach supplements covariance analyses[7,35] which are routinely done in social sciences. Here, we have used physics concepts to shed a different view on the spreading of obesity. Our analysis can be extended to study the geographical spreading of any epidemic: from diabetes and lung cancer, as shown here, to the spreading of viruses or real estate bubbles, where the spatial spreading plays an important role.

Population correlations are naturally inherited by all demographic observables. Even variables whose incidence varies randomly from county to county would exhibit spatial correlations in their absolute values, simply because its number increases in more populated counties and population locations are correlated. Indeed, the absolute number of obese adults per county is directly proportional to the population of the county[33]. Our aim is to measure spatial fluctuations on the frequency of incidence, independent of population agglomeration. Thus, spatial correlations of all indicators ought to be calculated on the density defined, in the case of obesity, as $s_i = o_i/p_i$, rather than on the absolute number of obese people, o_i, itself. The spatial correlations of the fluctuations of s_i from the global average captures the collective behavior expressed in the power-law described in Eq. (2).

While the understanding of covariance between obesity and other factors is out of the scope of the present study, we can still tentatively study the covariance of obesity and economic factors, such as income. We calculated the covariance between the obesity fraction at the county level with the per capita personal income in this county. The result (shown in SI-Fig. S2) indicates that there is a generally broad dependence of higher obesity in counties with lower income. This is indicated by the running average curve, which decreases as a function of the income. However, this covariance is not very strong as can be seen by the wide spreading of the counties in this plot. For instance, the county with the highest obesity prevalence (43.7%) in 2008 has an income of $31908, which is very close to the median income value. Consequently, the personal income indicator may not be reliably used to predict the obesity level at a given county.

In general, our approach attempts to go beyond this kind of covariance estimations by studying quantities such as the long-range correlated exponent γ, which may provide an alternative form of classification of dissimilar factors into universality classes, as done in Figs. 2a and 2b.

Discussion

Taken together, these results show that obesity spreading behaves as a self-similar strongly-correlated scale-free system. In particular, a note of caution has to be raised since, even if the highest prevalence of obesity is localized to the South and Appalachia, the scaling analysis indicates that the obesity problem is the same (self-similar) across all USA, including the lower prevalence areas.

Interestingly, the indicators that undergo a significant growth in short time intervals, such as lung cancer, diabetes, and obesity, fall in the universality class with strong long-range correlations ($\gamma_{\rm st} = 0.5$), although the inverse is not necessarily true. This finding leads us to the surprising conjecture that the static properties expressed by the exponent γ may be related to the growth rates[36], which is a dynamic quantity.

In Ref. [33] a model has been proposed where the population growth rate is characterized by a static exponent β that measures the scaling of resources or social activities with the population of a given city. The indicators related to the economic growth of the cities were found to increase faster than linear ($\beta > 1$) while the resources of the cities increase sub-linearly ($\beta < 1$). Thus, the population growth eventually depends on the value of β and different population estimates are predicted when switching from economies of scale ($\beta <$ 1, population growth asymptotically stops) to innovation-driven economies ($\beta > 1$, exponential population growth). This model is an attempt to classify different social and economic indicators according to human activity in cities, similar in scope with our study here. The relation between our results and Ref. [33] remains an open problem, since that study was a mean-field consideration and spatial correlations in activity were not taken into account.

Finally, we note that our results cannot establish a causal relation between obesity prevalence and economic indicators: whether fluctuations in the food economy may impact obesity or, instead, whether the food industry reacts to obesity demands. However, the comparative similarities of statistical properties of demographic and economical variables serves to identify possible candidates which shape the epidemic. Specifically, the observation of a common universality class in the correlations of obesity prevalence and economic activity of supermarkets, food stores and food services — which cluster in a different universality class than simple population dynamics — is in line with studies proposing that an important component of the rise of obesity is linked to the obesogenic environment[3,37] regulated by food market economies[4,5,9]. This result is consistent with recent research that relates obesity with residential proximity to fast-food stores and restaurants[7,35]. The present analysis based on clustering and critical fluctuations is a supplement to studies of association between people's BMI and food's environment based on covariance[7,35]. In sum, we have detected potential candidates in the economy which relate to the spreading of obesity by showing the same universal fluctuation properties. Eventually, these tentative relations ought to be corroborated by future intervention studies.

Methods

Datasets. Obesity is determined through the Body Mass Index (BMI) which compares the weight and height of an individual via the formula weight(kg)/height(m^2). A BMI value of 30 is considered the obesity threshold. Overweight but not obese is 25 <BMI< 30, and underweight is BMI<18.5. Our main measure in this work is the adult obesity prevalence of a county, $s_i = o_i/p_i$, defined for a given year as the number of obese adults o_i (BMI> 30) in a county i over the total number of adults in this county, p_i. We use the data from the USA Center for Disease Control (CDC) downloaded from Ref. [14]. CDC provides an estimate of the obesity country-wide since 1970, at the state level from 1984 to 2009, and at the county level from 2004 to 2008. The study of the correlation function $C(r)$ requires high resolution data. Therefore, we use data defined at the county level and restrict our study of obesity and diabetes to the available period 2004–2008. Other indicators are provided by different agencies at the county level for longer periods.

The datasets analyzed in this paper were obtained from the websites as indicated below. They can be downloaded from http://jamlab.org. The datasets consist of a list of populations and other indicators at specific counties in the USA at a given year. A graphical representation of the obesity data can be seen in Fig. 1b for USA from 2004 to 2008, where each point in the maps represents a data point of obesity prevalence directly extracted from the dataset.

The datasets that we use in our study have been collected from the following sources:

(a) Population

– US Census Bureau. We downloaded a number of datasets at the county level from http://www.census.gov/support/USACdataDownloads.html.

– For the population estimates we used the table PIN030. For the years 1969–2000 we use data supplied by BEA (Bureau of Economic Analysis) and for years 2000–2009 we use the file CO-EST2009-ALLDATA.csv from http://www.census.gov/popest/data/counties/totals/2009/files/CO-EST2009-ALLDATA.csv.

(b) Health indicators

– Data downloaded from the Centers for Disease Control and Prevention (CDC). http://apps.nccd.cdc.gov/DDT_STRS2/NationalDiabetesPrevalenceEstimates.aspx

The center provides county estimates between the years 2004–2008 for:

–Diagnosed diabetes in adults.
–Obesity prevalence in adults.
–Physical inactivity in adults.

The estimates for obesity and diabetes prevalence and leisure-time physical inactivity were derived by the CDC using data from the census and the Behavioral Risk Factor Surveillance System (BRFSS) for 2004, 2005, 2006, 2007 and 2008. BRFSS is an ongoing, state-based, random-digit-dialed telephone survey of the U.S. civilian, non-institutionalized population aged 18 years and older. The analysis provided by the BRFSS is based on self-reported data, and estimates are age-adjusted on the basis of the 2000 US standard population. Full information about the methodology can be obtained at http://www.cdc.gov/diabetes/statistics.

(c) Economic indicators

– We downloaded data for economic activity through http://www.census.gov/econ/. The economic activity of each sector is measured as the total number of employees in this sector per county in a given year normalized by the population of the county. The North American Industry Classification System (NAICS) (http://www.census.gov/eos/www/naics) assigns hierarchically a number based on the particular economy sector. The NAICS is the standard used by US statistical agencies in classifying business establishments across the US business economy.

In this study we have used the following economic sectors with their corresponding NAICS:

- 31. Manufacturing. Broad economic sector from textiles, to construction materials, iron, machines, etc.
- 42. Wholesale trade. Very broad sector including merchants wholesalers, motors, furniture, durable goods, etc.
- 56. Administrative jobs and support services.
- 445. Food and beverage stores. Including all the food sectors, from supermarkets, fish, vegetables meat markets, to restaurants and bars and other services to the food industry.
- 44511. Supermarkets and other grocery (except convenience) stores. This is a subsection of NAICS 445.
- 722. Food services and drinking places. A sub-sector of NAICS 72 which includes restaurants, cafeterias, snacks and nonalcoholic beverage bars, caterers, bars and drinking places (alcoholic beverages).

(d) Mortality rates

–We use data from the National Cancer Institute SEER, Surveillance Epidemiology and End Results downloaded from http://seer.cancer.gov/data/.

The Institute provides mortality data from 1970 to 2003, aggregated every three years. We analyze the mortality of a specific form of cancer per county normalized by the population of the county. Here, we use mortality data for the following causes of death:

–All cancer, independently of type.
–Lung cancer.

Gridded data of population from CIESIN

We take advantage of the available data of population distribution around the globe defined in a square grid of 2.5 arc-seconds obtained from[26]. These data allow to study the correlation functions of the population distribution for many countries. By using these data we are able to test the system size dependence of our results. We find that the correlation length ξ is proportional to the linear size of the country, L. The linear size is calculated as Total Area $= L^2$. We find that the correlation scales with the system size as discussed in the text. For instance, for the USA population distribution we find $\xi = 1050$ km, while a smaller country like UK has $\xi = 321$ km.

Supplementary Table S1 shows a list of countries used in Figs. 2d and e to determine the correlation length ξ of the correlation function of population density.

Fitting Methods

The fact that the correlation length diverges with the system size is an indication of critical behavior, and, thus, we search for power-law scaling, even though our system is finite.

The geographical analysis imposes constraints to the maximum possible scale of observing a power law, while there is a lot of noise in the datasets due to the complexity of acquiring and filtering the empirical data in their source. To improve the quality of the data we started by averaging the correlation functions over all years, for the cases where the powerlaw exponent seemed stable with time. We then calculated the running average with a window of 50 points along the x-axis. The resulting curve was fitted with standard OLS methods[24] in the range $[r_{min}, r_{max}]$, where r_{min} was in the range of 30–50 km, and r_{max} was in the range of 100–1000 km. We assessed the goodness of fitting in each interval through the coefficient of determination R^2, which can take values between 0 and 1. Here, we generally accept fittings where $R^2 \gtrsim 0.9$. The best fittings in almost all cases were in the range [40, 400]. The reported values of γ in the manuscript are obtained in this interval. We then used this interval to fit the individual correlation functions for each year.

1. World Health Organization. Obesity: Preventing and managing the global epidemic. *WHO Obesity Technical Report Series* **894** (World Health Organization Geneva, Switzerland 2000).
2. Butland, B. *et al. Foresight tackling obesities: future choices-project report, 2nd edn. London: Government Office for Science* (2007).
3. Hill, J. O. & Peters, J. C. Environmental contributions to the obesity epidemic. *Science* **280**, 1371–1374 (1998).
4. Swinburn, B. A. *et al.* The global obesity pandemic: shaped by global drivers and local environments. *Lancet* **378**, 804–814 (2011).
5. Nestle, M. *Food politics: How the food industry inuences nutrition and health.* (University of California Press, revised edition, 2007).
6. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *New. Engl. J. Med.* **357**, 370–379 (2007).

7. Block, J., Christakis, N. A., O'Malley, A. J. & Subramanian, S. Proximity to food establishments and body mass index in the Framingham Heart Study Offspring Cohort over 30 years. *Am. J. Epidemiol.* **174**, 1108–1114 (2011).
8. Caballero, B. & Popkin, B. M. editors. *The Nutrition transition: diet and disease in the developing world* (Academic Press, 2002).
9. Cutler, D. M., Glaeser, E. L. & Shapiro, J. M. Why have Americans become more obese? *J. Econ. Perspect.* **17**, 93–118 (2003).
10. Haslam, D. W. & James, W. P. Obesity. *Lancet* **366**, 1197–1209 (2005).
11. Department of Health and Human Services (USA). *Healthy People 2010: understanding and improving health. Conference edition.* (Washington, Government Printing Office, 2000).
12. Schuurman, N., Peters, P. A. & Oliver, L. N. Are obesity and physical activity clustered? A spatial analysis linked to residential density. *Obesity* **17**, 2202–2209 (2009).
13. Michimi, A. & Wimberly, M. C. Spatial patterns of obesity and associated risk factors in the conterminous U.S. *Am. J. Prev. Med.* **39**, e1–e12 (2010).
14. Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. http://apps.nccd.cdc.gov/DDT_STRS2/NationalDiabetesPrevalenceEstimates.aspx (accessed 15 Nov 2011).
15. Stanley, H. E. *Introduction to phase transitions and critical phenomena.* (Oxford University Press, Oxford, 1971).
16. Bunde, A. & Havlin, S. (editors). *Fractals and Disordered Systems* (Springer-Verlag, Heidelberg, 2nd edition, 1996).
17. Coniglio, A., Nappi, C., Russo, L. & Peruggi, F. Percolation and phase transitions in the Ising model. *J. Phys. A* **10**, 205–209 (1977).
18. Weinrib, A. Long-range correlated percolation. *Phys. Rev. B* **29**, 387–395 (1984).
19. Prakash, S., Havlin, S., Schwartz, M. & Stanley, H. E. Structural and dynamical properties of long-range correlated percolation. *Phys. Rev. A* **46**, R1724–1727 (1992).
20. Araujo, A. D., Moreira, A. A., Costa, R. N. & Andrade, J. S. Statistics of the critical percolation backbone with spatial long-range correlations. *Phys. Rev. E* **67**, 027102 (2003).
21. Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
22. Makse, H. A., Andrade, J. S., Batty, M., Havlin, S. & Stanley, H. E. Modeling Urban Growth Patterns with Correlated Percolation. *Phys. Rev. E* **58**, 7054–7062 (1998).
23. Montgomery, D. C. & Peck, E. A. *Introduction to linear regression analysis* (Wiley, New York, 1992).
24. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Review* **51**, 661–703 (2009).
25. Cavagna, A. *et al.* Scale-free correlations in starling flocks. *Proc. Natl. Acad. Sci. USA* **107**, 11865–11870 (2010).
26. Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World, Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University.Available at http://sedac.ciesin.columbia.edu/gpw (accessed 15 Nov 2011).
27. Mora, T. & Bialek, W. Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268–302 (2011).
28. Mokdad, A. H. *et al.* Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA - J. Am. Med. Assoc.* **289**, 76–79 (2003).
29. Rozenfeld, H. D., Rybski, D., Gabaix, X. & Makse, H. A. The area and population of cities: New insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 560–580 (2011).
30. Stanley, H. E. *et al.* Scale-invariant correlations in the biological and social sciences. *Philos. Mag. B* **77**, 1373–1388 (1998).
31. Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* (Cambridge University Press, 2000).
32. Schneidman, E., Berry, M. J., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
33. Bettencourt, L. *et al.* Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA* **104**, 7301–7306 (2007).
34. Weinrib, A. & Halperin, B. I. Critical phenomena in systems with long-range-correlated quenched disorder. *Phys. Rev. B* **27**, 413–427 (1983).
35. Chang, V. W. & Christakis, N. A. Income Inequality and Weight Status in U.S. Metropolitan Areas. *Soc. Sci. Med.* **61**, 83–96 (2005).
36. Rozenfeld, H. D. *et al.* Laws of population growth. *Proc. Natl. Acad. Sci. USA* **105**, 18702–18707 (2008).
37. Swinburn, B. & Figger, G. Preventive strategies against weight gain and obesity. *Obes. Rev.* **3**, 289–301 (2002).

Acknowledgments

LKG and HAM are supported by NSF-0827508 Emerging Frontiers, and PB and MS by Human Frontiers Science Program. We are grateful to H. Nickerson for bringing our attention to this problem, and S. Alarcón-Díaz and D. Rybski for valuable discussions. We thank Epiwork, ARL under Cooperative Agreement Number W911NF-09-2-0053 and the Israel Science Foundation for support.

Author contributions

All authors contributed equally to the work presented in this paper.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gallos, L.K., Barttfeld, P., Havlin, S., Sigman, M. & Makse, H.A. Collective behavior in the spatial spreading of obesity. *Sci. Rep.* **2**, 454; DOI:10.1038/srep00454 (2012).