

<https://doi.org/10.1038/s41538-026-00764-0>

# Mapping regional disparities in discounted grocery products

Antonio Desiderio<sup>1,2,6</sup>✉, Alessia Galdeman<sup>3,4,6</sup>, Franziska Bäuerlein<sup>1,6</sup> & Sune Lehmann<sup>1,5</sup>✉

Food waste represents a major challenge to global climate resilience, accounting for almost 10% of annual greenhouse gas emissions. The retail sector is a critical player, mediating product flows between producers and consumers, where supply chain inefficiencies can shape which items are put on sale. Yet how these dynamics vary across geographic contexts remains largely unexplored. Here, we analyze data from Denmark's largest retail group on near-expiry products put on sale. We uncover the geospatial variations using a dual-clustering approach. We characterize multi-scale spatial relationships in retail organization by correlating store clustering – measured using shortest-path distances along the street network – with product clustering based on promotion co-occurrence patterns. Using a bipartite network approach, we identify three regional store clusters, and use percolation thresholds to corroborate the scale of their spatial separation. We find that stores in rural communities put meat and dairy products on sale up to 2.2 times more frequently than metropolitan areas. In contrast, metropolitan and capital regions lean toward convenience products, which have more balanced nutritional profiles but less favorable environmental impacts. By linking geographic context to retail inventory, we provide evidence that reducing food waste requires interventions tailored to local retail dynamics, highlighting the importance of region-specific sustainability strategies.

The growing field of food informatics—which draws on large-scale digital records to chart the complex system of human interactions with food products—has transformed our understanding of food processing, consumption, and diet<sup>1–4</sup>. For example, advances in computational approaches now allow us to predict the levels of food processing<sup>5</sup>, revealing that more than 73% of the U.S. food supply is ultra-processed<sup>6</sup>, and uncovering the vast, unmapped chemical diversity of our diets<sup>7</sup>. At the same time, integrating mobility data with food choice patterns<sup>8,9</sup> has shown that the food environments people encounter throughout their daily routines differ markedly from those near their homes, significantly shaping dietary decisions and consumption behaviors<sup>10</sup>. Yet, while this line of research has clarified the link between food and health through nutrition<sup>11–14</sup>, food systems also influence human health through another key pathway: food waste, which generates environmental burdens that exacerbate resource scarcity and associated health risks<sup>15–21</sup>. Globally, nearly one-third of all food produced for human consumption is lost or discarded each year, with profound environmental, economic, and social consequences<sup>22–25</sup>.

While food waste originates from inefficiencies throughout the supply chain<sup>26–29</sup>—from production through distribution to consumption – the retail sector represents a critical transition point where products shift from distribution networks to consumers<sup>30–32</sup>. As a complex adaptive system<sup>33,34</sup>, retail sales are shaped by both promotional interventions on perishable items and evolving consumer preferences<sup>30,31</sup>, creating heterogeneous spatial and temporal patterns in product distribution across stores<sup>35</sup>. For instance, large-scale analyses of grocery purchasing records have linked dietary behavior to health outcomes, revealing spatially structured differences within cities<sup>36</sup>, while lifestyle-based consumption behaviors form distinct clusters across urban environments<sup>37</sup>, indicating that food choices—and by extension food waste—are deeply embedded in cities' social and spatial fabric<sup>38,39</sup>. Given extensive research demonstrating how hierarchical spatial organization<sup>40</sup> fundamentally shapes urban mobility<sup>41,42</sup>, economic activity<sup>43</sup>, and overall human behavior patterns<sup>44–48</sup>, we were inspired to question whether or not similar spatial hierarchies also structure the spatial distribution of promotions of perishable foods—in other words, how food

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark. <sup>2</sup>ISI Foundation, Turin, Italy. <sup>3</sup>Data Science Section, IT University of Copenhagen, Copenhagen, Denmark. <sup>4</sup>Computer Science Department, University of Milan, Milan, Italy. <sup>5</sup>Copenhagen Center for Social Data Science (SODAS), University of Copenhagen, Copenhagen, Denmark. <sup>6</sup>These authors contributed equally: Antonio Desiderio, Alessia Galdeman, Franziska Bäuerlein. ✉e-mail: [antde@dtu.dk](mailto:antde@dtu.dk); [sljo@dtu.dk](mailto:sljo@dtu.dk)

waste dynamics and promotional strategies vary across different spatial scales and geographic contexts<sup>39</sup>.

We address these questions by analyzing spatial patterns of promotional strategies for near-expiry products across Denmark. Even in a small country like Denmark, the impact of food waste is significant—contributing to approximately 2.3 million tonnes of CO<sub>2</sub>-equivalent greenhouse gas emissions annually<sup>49</sup>, and beyond economic loss and resource depletion, addressing it is essential for achieving sustainable development goals related to climate action, resource efficiency, and social equity<sup>24,28,35,50,51</sup>. Furthermore, a previous analysis of Danish retail data focusing on alcohol consumption demonstrated significant consumer heterogeneity in food purchasing patterns<sup>52</sup>, with distinct lifestyle-based clusters showing different preferences for healthy versus processed foods, highlighting how food patterns could vary across regions with fundamentally different consumer bases. Products are discounted by retailers as part of a complex optimization strategy that balances supply constraints, demand variability, and expiration pressures to minimize waste while maximizing revenue<sup>53–56</sup>. When a product is placed on sale repeatedly over time, this temporal pattern signals systematic inefficiencies in demand forecasting models, inventory optimization protocols, and product lifecycle management strategies<sup>56</sup>. Consequently, understanding the spatial heterogeneity of promotional patterns across geographic scales offers a diagnostic lens into how place-based variations in logistics, forecasting, and consumer behavior shape food waste trajectories in retail networks.

## Results

Salling Group, Denmark's largest retailer, has focused on food waste reduction with a goal to cut food waste by 50% by 2030<sup>57</sup>. As part of this strategy, Salling Group provides a public API (Application Programming Interface) that offers real-time information about discounted products approaching their expiration dates, such as meat, dairy products, and prepared meals. We collected data on discounted products across Netto stores, which operate 542 locations in Denmark and rank among the country's top discount supermarket chains (See Supplementary Materials, Section 1, Table S1–4 and Methods for data description). Our dataset captures daily information for each product, recording when it is on sale at a discounted price and where. This dataset allows us to examine the complex spatial-temporal patterns of food waste across Denmark over our 153-day period (See Methods for further details). Furthermore, we manually inspect the dataset and remove offers corresponding to high volumes of products (>50 units), representing the 95th percentile, as these tend to correspond to large stock quantities on sale for promotional campaigns rather than genuine near-expiry products. We combine store location data with OpenStreetMap to compute distances between stores, and each product has been mapped through its corresponding Nutri-Score and Environmental impact scores to enable comprehensive sustainability analysis using OpenFoodFacts (See Methods for further details).

### Spatial clustering and product composition patterns in Denmark

To investigate product distribution patterns across Denmark, we analyze data spanning the country's five administrative regions: Hovedstaden (Capital Region), Sjælland (Zealand), Syddanmark (Southern Denmark), Midtjylland (Central Jutland), and Nordjylland (North Jutland) (See Fig. 1a). First, we employ hierarchical spatial clustering of store locations using shortest-path distances on the street network to identify densely connected urban cores and their peripheral hinterlands<sup>40,58</sup>. We identify groups of stores that are densely connected in terms of street-network travel distance, while leaving more isolated locations unclustered. This approach allows us to distinguish compact urban cores from their (diffuse) peripheries (see Methods for full details). Figure 1b shows a distinct core-periphery structure<sup>59</sup>, as it is characterized by high silhouette scores (0.65), suggesting well-separated and coherent clusters of stores. The maximum persistence (0.36), which quantifies the robustness of clusters across multiple resolutions, is low-to-moderate, reflecting that while the Capital Region forms a stable dominant core, peripheral areas

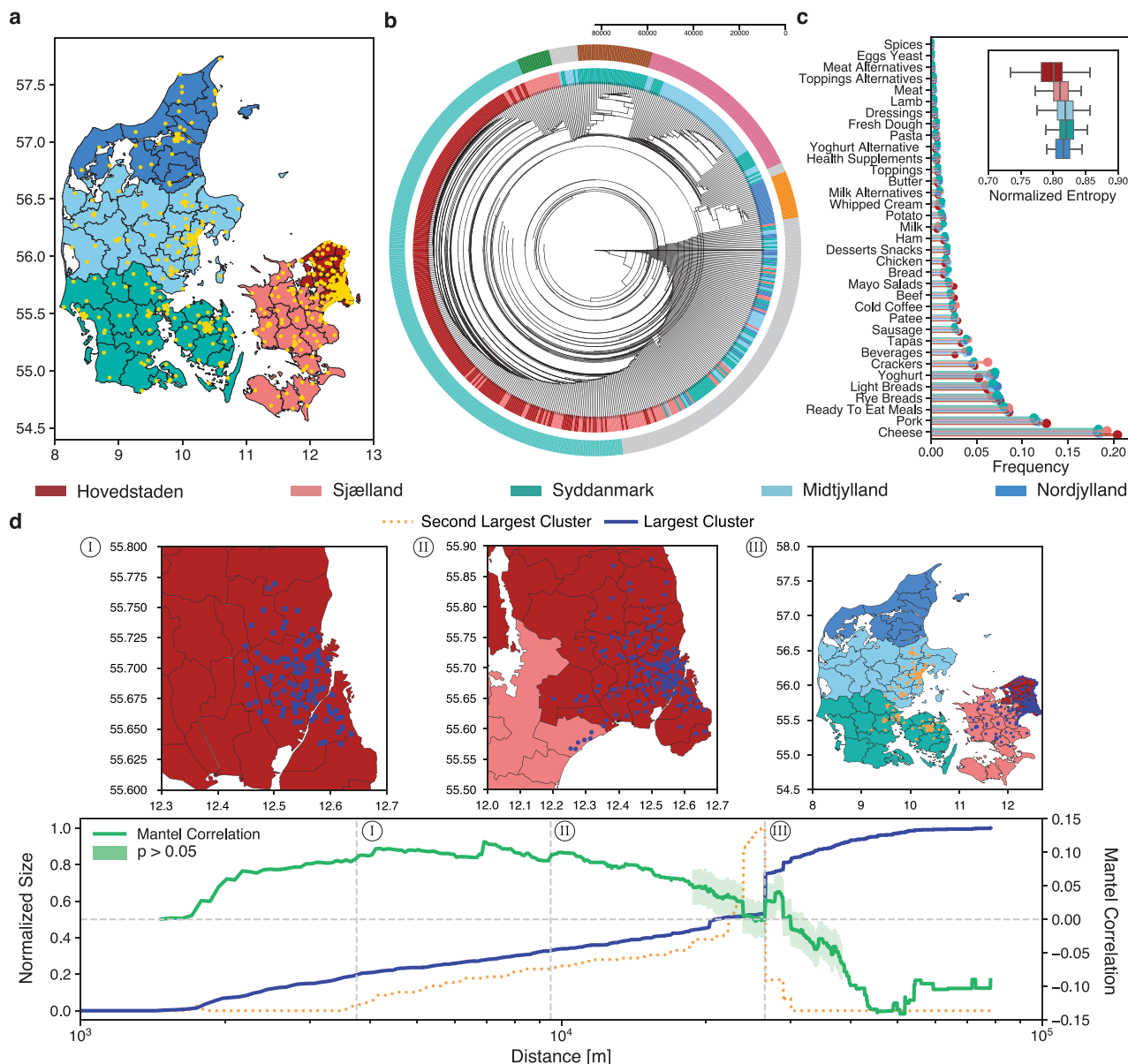
display fragmented, low-persistence clustering that aligns with their scattered population distributions.

To quantify both the magnitude of store-product promotional dynamics—reflecting inventory management decisions and consumer behavior patterns—and the temporal persistence of discounting activities that signal systematic surplus conditions, we aggregate products by category and compute cumulative daily promotion frequencies across the observation period. Figure 1c shows that Cheese, Pork, Ready-to-Eat Meals, Breads, and Yoghurt emerge as the most commonly offered items nationwide. These categories represent consumer staples with varying shelf lives, ranging from perishable items such as Pork and Bread, which typically remain viable for approximately one week, to longer-lasting products like Cheese and Yoghurt. Although overall patterns of on sale items are similar across regions, normalized entropy—a measure of diversity where higher values indicate a more uniform distribution of promotional activity across product categories—highlights regional distinctions: the Capital Region (Hovedstaden)—comprising the urban core of Copenhagen and Frederiksberg alongside adjacent rural municipalities such as Fredensborg, Hillerød, and Gribskov—have a low mean entropy (0.79) and greatest variability across stores within the region. This means that, on average, fewer products tend to be on sale, and the high variability suggests substantial heterogeneity in promotional strategies within the region. On the other end of the scale Syddanmark (0.820), Midtjylland (0.818), Nordjylland (0.815), and Sjælland (0.811) display higher and more stable diversity. These patterns suggest regional differences with respect to on sale product diversity<sup>33,53–56</sup>.

### Scale-dependent retail product similarity

Building upon these entropy-based findings, we characterize inter-store differences in promotional composition by computing the Bray-Curtis<sup>60</sup> dissimilarity matrices across all store pairs, using category-specific promotion frequencies as abundance proxies that reflect underlying supply-demand dynamics (See Methods and Supplementary Materials, Section 2, Fig. S1). The Bray-Curtis dissimilarity index quantifies compositional differences between stores by comparing the relative frequencies of promotional activities across product categories, in direct analogy to ecological analyses that assess community dissimilarity based on species abundance distributions. Product categories with high promotional frequencies—indicative of persistent surplus conditions or systematic demand forecasting errors—contribute disproportionately to inter-store dissimilarity, with the index ranging from zero (identical promotional profiles) to one (completely distinct promotional patterns), thereby allowing us to understand the spatial heterogeneity of supply chain inefficiencies across retail networks. Then, we examine the relationship between geographic proximity and compositional similarity using Mantel correlations<sup>61</sup> computed across increasing distance thresholds (see Methods). These correlations are summarized in a cumulative correlogram that characterizes how promotional composition similarity varies with spatial separation (Fig. 1d).

We observe a positive spatial-product correlations at smaller scales, peaking at around 4000 m (Mantel Correlation  $\sim 0.1$ ,  $p < 0.05$ ), driven primarily by the cohesive Capital Region cluster (See Fig. 1dI). This correlation diminishes upon inclusion of the Sjælland region (See Fig. 1dII) and becomes zero and nonsignificant ( $p > 0.05$ ) when peripheral regional clusters merge with the largest connected component ( $\sim 26500$  m). Beyond this critical distance, the correlation starts to be statistically significant, but in the negative direction, indicating divergence in composition among more distant stores (See Fig. 1dIII). This scale-dependent pattern suggests that while local retail environments in the capital region exhibit modest product homogeneity, regional differentiation strategies emerge at larger spatial scales, indicating hierarchical organization in Denmark's retail landscape<sup>40</sup>. This hierarchical organization is statistically confirmed by PERMANOVA analysis ( $F = 6.9$ ,  $p = 0.001$ ), which tests for differences in multivariate composition using permutation-based comparisons of Bray-Curtis dissimilarities, demonstrating that spatial clusters correspond to genuinely



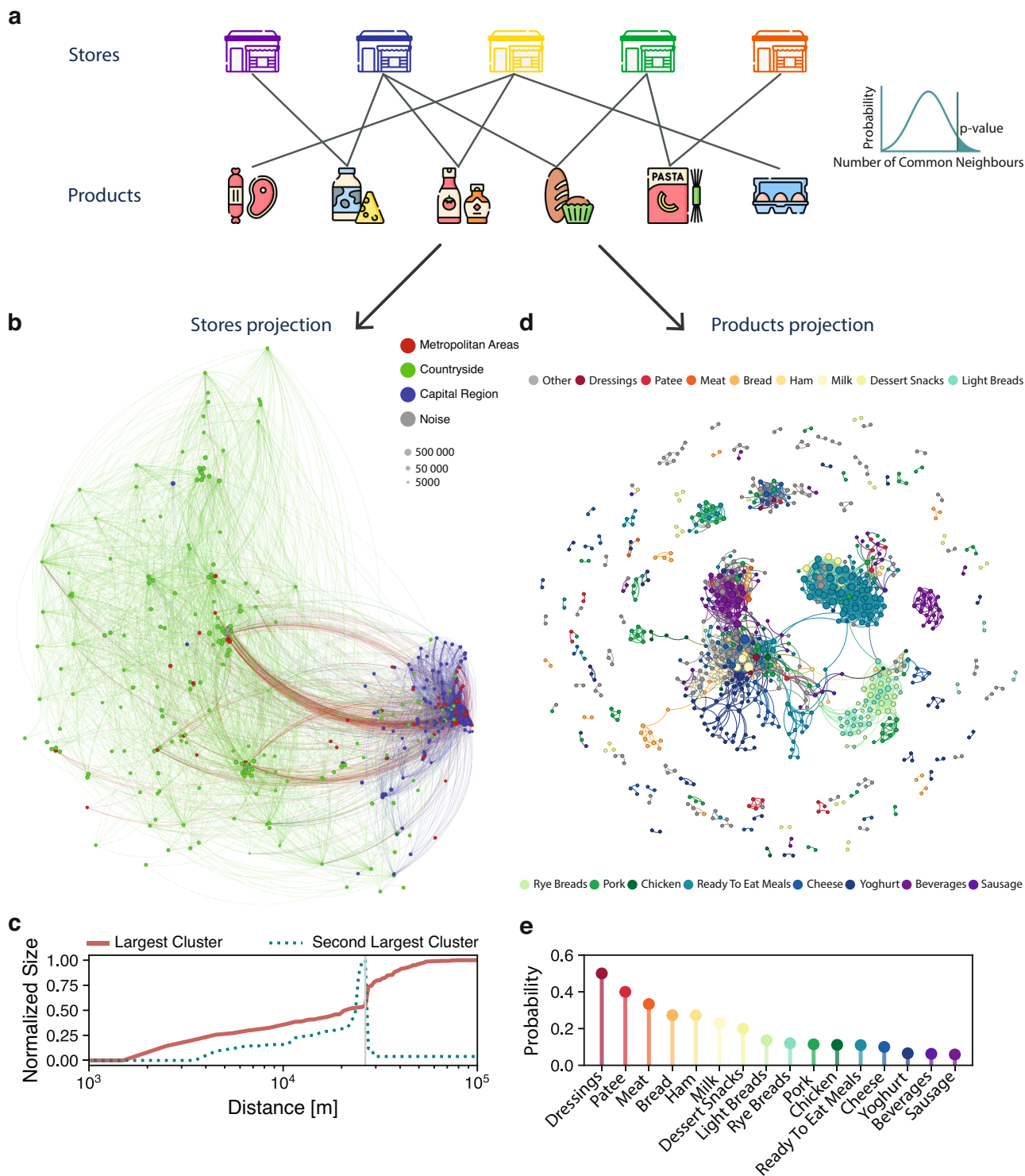
**Fig. 1 | Spatial patterns of retail product on sale distribution across Denmark's administrative regions.** **a** Map of Denmark showing the five administrative regions analyzed: Hovedstaden (red), Sjælland (pink), Syddanmark (teal), Midtjylland (light blue), and Nordjylland (blue), with store locations and spatial clustering results overlaid. Map data copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org>. **b** Hierarchical spatial clustering analysis showing the core-periphery structure of retail locations based on shortest-path distances along the street network. Clusters are identified by the colors in the second outer ring, with noise shown in grey, whereas regions are identified by the colors in the first ring. **c** Regional analysis of product-on sale category frequencies and diversity patterns. The lollipop chart, corresponding to a rank plot, shows commonly on sale product categories nationwide (Cheese, Pork, Ready-to-Eat Meals, Breads, and

Yoghurt), while the inset of normalized entropy highlights regional differences in product diversity. **d** Mantel correlation analysis (green) examines the relationship between spatial proximity and promotional composition similarity across multiple distance scales. The correlogram reveals three distinct phases (grey vertical lines): (I) positive spatial-product correlations at smaller scales peaking around 4000 km, driven by the cohesive Hovedstaden cluster; (II) correlation diminishment upon inclusion of the Sjælland region, with the inversion of the Mantel correlation (and becomes statistically non-significant as highlighted by the shaded area) occurring at the point where the second-largest cluster (orange) merges with the first, as tracked by the blue line (fraction of nodes in largest cluster divided by the maximum value); and (III) negative correlations at larger scales (~26,500 km) indicating compositional divergence among distant stores.

distinct product compositions (see Methods for further details). To assess the robustness of our findings, we perform a sensitivity analysis by excluding random days from the dataset, incorporating an alternative check that measures total product availability as the sum of individual volumes offered, and normalizing volumes by dividing it by the corresponding population for each store within its respective municipality (See Methods and Supplementary Materials, Section 3, Figure S2–4). The analysis is unchanged in the light of these perturbations.

**Statistical validated projection reveals regional retail communities and key products**

Building on the fact that the spatial clustering results reveal scale-dependent patterns in Denmark's retail geography, we use a bipartite network approach to identify which products drive the regional differences (See Fig. 2a). The bipartite network framework avoids circular reasoning by allowing product-store associations to emerge organically from data rather than relying on geographical distance, which risks



**Fig. 2 | Bipartite network analysis identifies product-driven regional retail differentiation in Denmark.** **a** The infographic summarizes the bipartite network with stores and products on separate layers. **b** Store-layer projection after validation showing three well-defined communities corresponding to Metropolitan Areas (red), Capital Region (blue), and Countryside (green) clusters. The nodes are positioned according to their geographic coordinates (latitude and longitude), while avoiding overlaps. Node size proportional to the logarithm of the population. **c** Fraction of nodes in the largest cluster (red) and the second-largest cluster (teal)

identified via percolation analysis, with clusters merging at the same distance indicated by the hierarchical clustering (grey vertical line ~ 26,500 km). **d** Products-layer projection using the Yifan Hu layout. Node colors highlight different categories and size proportional to the number of neighbors. The network contains 124 communities, most of which are actually disconnected components (111). **e** Rank-ordered probability of each product category appearing among the top 10% of nodes by betweenness centrality.

conflating spatial proximity with functional relationships. The raw network is dense—comprising 542 stores, 1854 products, and over 418,000 weighted links, where each edge weight represents the total number of days a specific product appears on sale in a given store—many of which connections may be redundant or statistically uninformative. To prune the network we apply two filtering strategies: Revealed Comparative Advantage (RCA) and then the Bipartite Configuration Model (BiCM) to extract the underlying structure from the network<sup>62–64</sup>. RCA measures whether a store offers products more frequently than expected given the overall market<sup>65</sup>, with values above 1 indicating relative specialization (See Methods and Supplementary Materials, Section 4, Figure S6 for the RCA distribution). After removing weaker connections through RCA filtering, we employ BiCM to statistically validate the observed network structure. In a nutshell, BiCM generates an ensemble of randomized bipartite networks that preserve the degree sequences of stores (total number of unique products connected) and products (total number of unique stores connected) while randomizing other structural features (See Methods for further details). This approach constitutes a null model used to distinguish genuine patterns from artifacts of store size or product popularity (defined here as the number of stores in which a product is on sale)<sup>64</sup>.

When projecting the validated bipartite structure onto the store layer, we identify three well-defined communities (modularity  $Q = 0.49$ , indicating strong internal connectivity within clusters compared to random connections). These clusters correspond closely to the three phases previously revealed as part of our purely spatial analysis: Metropolitan Areas, Capital Region, and Countryside (See Fig. 2b). Notably, the Metropolitan Areas cluster extends beyond Copenhagen, encompassing cities such as Odense and Aarhus<sup>59</sup>, while the Capital Region cluster reflects the Finger Plan<sup>66</sup>, with urban development radiating from central Copenhagen along five suburban ‘fingers’ separated by green wedges. These communities are robust and emerge from two different community detection methods, Louvain and Infomap<sup>67,68</sup> (Adjusted Rand Index  $ARI = 0.82$ ,  $p < 0.001$ , See Methods for further details). We incorporate distance weights into the store-store network and apply percolation analysis<sup>40</sup> to identify the hierarchical connectivity structure of this network. This percolation profile captures how densely connected store clusters emerge as we vary the distance threshold parameter. Notably, this distance-weighted percolation profile closely mirrors that of our purely spatial percolation analysis, indicating that spatial proximity information is inherently encoded within the product-based similarity network (See Fig. 2c). Collectively, the convergence reveals that a) spatial proximity and b) product on sale similarity offer complementary perspectives on the same underlying organizational principle: stores that operate within similar geographical contexts face comparable market demands and operational constraints. These, in turn, generate convergent promotional patterns. The convergence also highlights potential shared inefficiencies that point to a sharp contrast between rural and urban supply-demand dynamics, and reflecting shared supply chain challenges and consumer preference behaviors<sup>33,37,39</sup>. For instance, the consistent discounting of meat in rural areas versus convenience foods in the capital reflect distinct local inefficiencies. These disparities suggest that a path to sustainable waste reduction can lie in shifting from reactive uniform discounting to proactive, region-specific inventory adjustments<sup>69</sup>. As in the preceding section, we perform a sensitivity analysis on the bipartite network weights, as detailed in Supplementary Materials, Section 4 (Fig. S5), which further underscores the robustness of our results.

On the other hand, the product layer projection captures similarities in aggregate promotional patterns across stores, revealing regionally characteristic product communities and identifying critical ‘bridge’ products<sup>34</sup>. This projection has a very high modularity ( $Q = 0.83$ ) and again we find agreement between Infomap and Louvain community assignments (Adjusted Rand Index  $ARI = 0.68$ ,  $p < 0.001$ ). The contrast between store-layer modularity ( $Q = 0.49$ ) and product-layer modularity ( $Q = 0.83$ ) points to moderate asymmetric organizational forces<sup>63</sup>: while

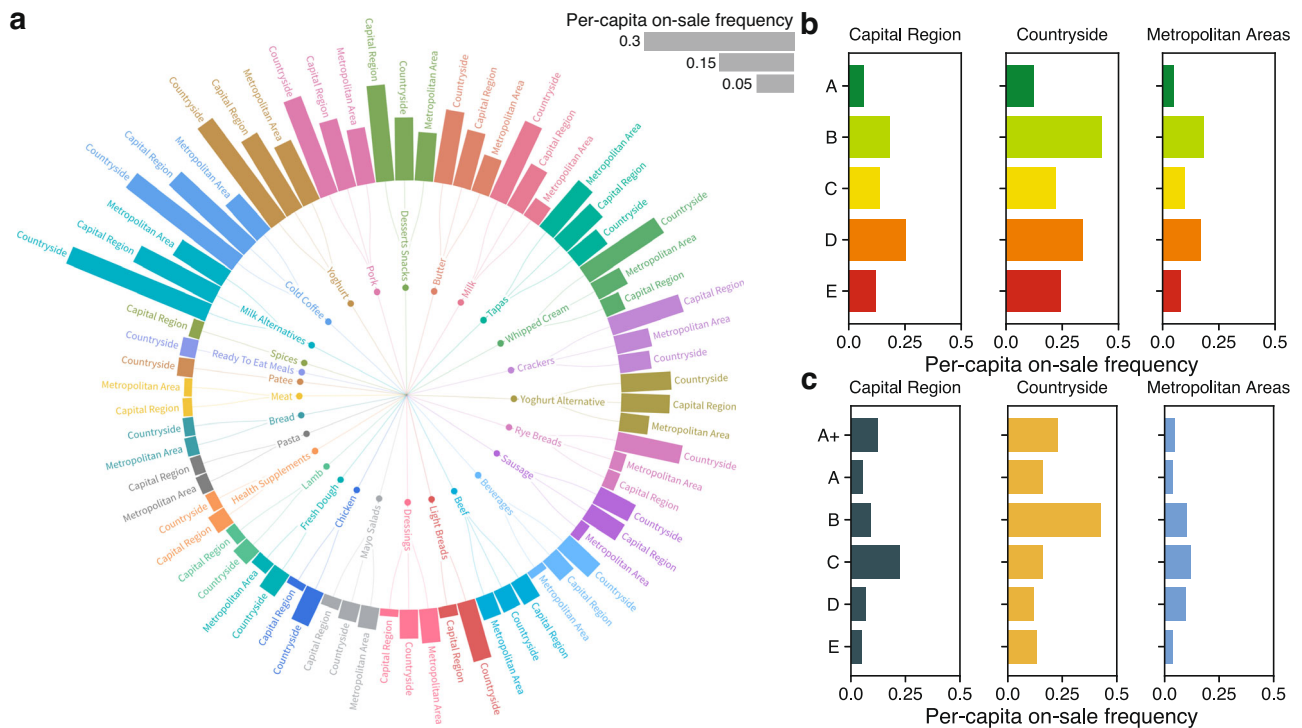
geographical constraints create moderate store clustering, product relationships exhibit much stronger functional cohesion. This cohesion implies that promotional co-occurrence patterns – driven by shared supply chain dynamics and consumer purchasing behaviors – create more rigid boundaries in product networks than spatial proximity generates in store networks. Although more than one hundred distinct communities emerge, this (large) number mainly reflects the fragmented structure of the network (See Fig. 2d). Many communities align closely with conventional categories—Light and Rye Breads, Ready-to-Eat Meals, Beverages, and various Meat assortments—while notable divergences appear, such as a central cluster that unites Yoghurt, Cheese, Milk, Tapas and Whipped Cream. Focusing on the largest connected component, which comprises 433 products, we compute node betweenness centrality (bc) to identify core ‘hub’ products that are frequently part of shortest paths between items. Unlike the categories in Fig. 1c, which are ranked by overall promotional volume, here we identify structural hub products that connect diverse promotional clusters. These clusters shape both regional uniformity and inter-regional differences through their network connectivity. Categories such as Dressings, Meat, Pâté, and Breads are most likely to include nodes with high betweenness centrality ( $bc > 0.03$ , top 90th percentile; see Fig. 2e), suggesting they serve as key intermediaries.

### Unhealthy products dominate peripheral communities

Building on the identification of store communities and product clusters, we now turn to the question of what differentiates these clusters in practice<sup>63</sup>. To address this, we evaluate the per-capita on sale frequency, which reflects how frequently product groups (product categories, nutritional profiles, or environmental ratings) go on sale relative to the number of customers each store serves. The per-capita on sale frequency is defined as the ratio between validated store-product connections within each community  $c$  and product category  $k$  (or Nutri-score and Environmental Score),  $E_{ck}$ , and the total population  $P_c$  of distinct stores in the community  $c$ . We scale this index by a factor of 100,000, so that it reflects the number of on sale items per 100,000 potential customers. The per-capita on sale frequency quantifies both oversupply conditions (indicating demand forecasting failures and waste risk) and availability patterns that mirror consumer preferences. Thus this measure is able to capture the relationship between supply chain strategies and community characteristics.

Figure 3a shows systematic differences in promotional patterns across the three community types, each exhibiting distinct profiles<sup>37,52</sup>. A Countryside versus Metropolitan Areas comparison reveals systematic differences in the promotional intensity of meat: Countryside shows a total meat frequency of 0.47–2.2 times higher than the Metropolitan Areas’ 0.21. In fact Countryside dominates across all meat categories: Pork is nearly twice as frequent as in Metropolitan Areas, while Chicken is absent in the Metropolitan Area. These patterns suggest that the Countryside has higher promotional intensities for specific meat categories—Pork, Chicken, Lamb, and Sausage—than Metropolitan Areas. Additionally, Butter shows markedly higher promotional intensity in the Countryside, with per-capita on sale frequencies two times greater than those observed in the Metropolitan Areas. Conversely, products such as Tapas, Pasta, and Crackers are more frequently observed in the Capital Region and the Metropolitan Areas. Additionally, the Capital Region demonstrates a pronounced preference for convenience and on-the-go products: Cold Coffee, Dessert Snacks, and Crackers all rank among its top five items.

To better understand the environmental and nutritional impact of these differences, we have mapped each available product to its corresponding Nutri-Score (72 validated products) and Environmental score (66 validated products), which respectively rate nutritional quality (from A for the healthiest to E for the least healthy) and environmental footprint (from A+ for the lowest impact to E for the highest impact). Figure 3b shows how categories B and D are the most prevalent across all community types, indicating that the healthiest products (A) are under-represented in promotions<sup>13</sup>. Critically, poor nutritional scores (D+E combined) reach



**Fig. 3 | Community-specific promotional composition profiles and prevalence of unhealthy products in peripheral danish communities.** **a** Per-capita on sale frequency (per 100,000 potential customers) for product categories (color) across Metropolitan Areas, Capital Region, and Countryside. The Cheese category has been removed, as its values exceed 0.6 across all communities, and values below 0.01 have also been excluded to improve readability. **b** Distribution of Nutri-Score ratings (A: healthiest to E: least healthy) across the three community types, showing

prevalence of categories B and D and under-representation of the healthiest products in all regions. **c** Environmental Score distribution (A+: lowest impact to E: highest impact) reveals mid-grade products (B) dominating across communities, with Countryside showing more environmentally positive options while Metropolitan Areas and Capital Region lean toward less environmentally favorable products.

0.58 in Countryside, 0.37 in Capital Region, and 0.25 in Metropolitan Areas, which presents the most balanced profile with a relatively even distribution across all score categories. In contrast, rural areas exhibit both higher promotional frequencies for high-quality foods and elevated discount rates for low-quality options, suggesting a wider but potentially riskier dietary range<sup>6,52</sup>. On the other hand, Fig. 3c suggests a complementary picture regarding the Environmental score, where mid-grade (B) dominates overall, and the Countryside demonstrates environmentally positive products. However, meat products are not mapped, despite being well known to exert a disproportionately large environmental impact<sup>18</sup>. Notably, the Metropolitan Areas and Capital Region have a distribution shifted towards less environmentally positive products (C-D).

**Discussion**

Human interactions with food systems generate system-wide challenges<sup>24</sup>, creating complex feedback loops between dietary behaviors, environmental impacts, and social outcomes that reinforce one another<sup>19,20,22,27,29</sup>. Our analysis reveals regional differences in the near-expiry discounted products across Denmark, demonstrating that spatial context shapes both the composition and intensity of near-expiry product offerings<sup>44</sup>. The bipartite approach identifies three distinct retail communities—Countryside, Metropolitan Areas, and Capital Region—each exhibiting unique profiles in meat, dairy, convenience, nutritional, and environmental product assortments<sup>63,64</sup>. Our findings highlight the need for food waste reduction strategies that are sensitive to local retail dynamics<sup>39</sup>, as neglecting such spatial differences may reinforce supply chain inefficiencies and amplify environmental burdens. While general interventions, such as the phasing out of quantity discounts, have been effective in Denmark, our results suggest that further reductions could arise from regionally differentiated policies—for example, targeting

single-household packaging sizes in capital regions to reduce convenience food waste, while focusing on cold-chain inventory optimization for meat products in rural areas<sup>69</sup>.

Our work comes with limitations. First, our study relies on 153 days of data from a single retail chain in Denmark, which may not capture the full spectrum of retail behaviors across all supermarket formats. Recognizing the temporal limitation, we have performed the same analysis removing days from the dataset to ensure the robustness of our main results. Second, our approach relies on promotional frequency. As discussed above, this is a composite proxy that integrates multiple underlying drivers, including demand forecasting accuracy, inventory management efficiency, supply chain dynamics, and consumer behavioral patterns – rather than direct measures of spoilage or waste quantities, limiting our ability to disentangle the precise contribution of each factor. Third, the mapping of products to Nutri-Score and Environmental scores covers only a subset of items (see Methods), potentially introducing biases in the assessment of nutritional and environmental profiles across regions.

Overall, our results underscore the strong role of geography in shaping human behavior patterns<sup>38,40,44,48</sup>. Food choices are grounded in local social and spatial contexts, influenced by the organization of urban and rural areas and how promotions are tailored to these settings. The regional differences in per-capita on sale frequencies and product composition patterns highlight that supply-demand mismatches vary across locations, emphasizing that effective food waste reduction strategies must be tailored to local retail dynamics rather than applied uniformly<sup>70</sup>. Retail patterns shift across regions and lifestyle clusters, showing how deeply place and daily decisions are linked<sup>46</sup>. Food waste, then, is the sum of these everyday choices – highlighting the need for strategies that tackle both geographic disparities and individual behavior to drive sustainable consumption<sup>50</sup>.

## Methods

### Data descriptions

We have retrieved Data from the Salling Group Anti Food Waste API, encompassing both store and product information (See Supplementary Materials, Section 1 Table S1–4). Store data include geographic coordinates, unique identifiers for each location and store brand. Product data capture items on discount offers, including EAN codes, product categories, stock levels, and temporal information such as discount start and end times. Additionally, we have mapped each product to its corresponding Nutri-Score and environmental score using OpenFoodFacts, an open database of food products from around the world that aggregates ingredient lists, nutritional information, and sustainability metrics. Nutri-Score is a front-of-pack labeling system that grades foods from A (healthiest) to E (least healthy) based on their nutritional composition, while the environmental score quantifies the product’s ecological footprint across factors like greenhouse gas emissions, water usage, and land use<sup>71</sup>. We have collected Data during three predefined periods (2024-08-08 to 2024-10-10, 2025-03-05 to 2025-04-10, and 2025-05-01 to 2025-06-21) and analyses are limited to these observation windows. When multiple discount offers existed for the same product on the same day at a given store, only the record with the lowest price has been retained. This processing approach results in a structured dataset that captures the daily availability of discounted products across the Netto store network throughout 153 days. We have computed the shortest road-network distance between every pair of stores using the Open Source Routing Machine (OSRM) on OpenStreetMap data<sup>72</sup>, and obtained municipality boundaries and resident population figures for Copenhagen’s level-2 administrative units (municipality level) from the Copernicus Global Degree of Urbanisation Classification dataset<sup>73</sup>.

### Hierarchical clustering

To identify spatially coherent retail clusters, we have employed Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), a density-based clustering algorithm that extends DBSCAN by converting it into a hierarchical clustering framework<sup>58</sup>. HDBSCAN constructs a hierarchy of clusters by varying the density parameter and extracting stable clusters across multiple density levels. We have used the pre-computed shortest path distance matrix between all store locations and set the minimum cluster size parameter to 10 stores to ensure that identified clusters represent meaningful retail agglomerations rather than isolated store pairs. The clustering procedure identified five distinct spatial clusters across Denmark’s retail landscape, grouping part of Sjælland and all of Hovedstaden into one cluster, combining a mix of Midtjylland and part of Syddanmark in another, capturing a separate part of Syddanmark in a third, forming clusters corresponding mainly to Nordjylland and Sjælland, while noise is scattered primarily across Jutland. To assess clustering quality, we have computed three complementary metrics: the Davies-Bouldin Index (0.44), which measures the average similarity ratio between clusters where lower values indicate better separation; the Silhouette Score (0.65), quantifying how well-separated clusters are with values approaching 1.0 indicating optimal clustering; and Maximum Persistence (0.364), representing the stability of clusters across the hierarchical structure where lower values suggest less robust cluster boundaries.

To quantify promotional composition differences between stores, we have computed Bray-Curtis dissimilarity indices for all store pairs based on their category-specific promotion frequencies. The Bray-Curtis index<sup>60</sup> is defined as

$$\text{Bray-Curtis Similarity}_{ij} = 1 - \frac{\sum_{k=1}^n |x_{ik} - x_{jk}|}{\sum_{k=1}^n (x_{ik} + x_{jk})}, \quad (1)$$

where  $x_{ik}$  and  $x_{jk}$  represent the abundance of product category  $k$  in stores  $i$  and  $j$ , respectively, and  $n$  is the total number of categories (See Supplementary Materials, Section 2, Figure S1 for the dendrogram). Abundance for each product category is calculated as the cumulative number of products on sale multiplied by the number of days they were offered. We have assessed

the relationship between spatial proximity and product similarity using Mantel correlation analysis across multiple distance scales. The Mantel test evaluates whether two distance matrices are significantly correlated by computing Pearson correlations between corresponding elements and testing significance through permutation<sup>61</sup>. We have constructed a cumulative correlogram by calculating Mantel correlations at increasing distance thresholds, enabling detection of scale-dependent spatial autocorrelation patterns in promotional composition similarity. To statistically validate that identified spatial clusters correspond to genuinely distinct product composition strategies, we have performed Permutational Multivariate Analysis of Variance (PERMANOVA) using the Bray-Curtis dissimilarity matrix and the spatial clusters. PERMANOVA is a non-parametric test that evaluates whether groups differ in multivariate composition by partitioning distance matrices and calculating pseudo F-statistics. Group labels are permuted 1000 times to generate a null F-distribution, and significance is assessed by comparing the observed F to this permutation-based distribution, avoiding assumptions about data distribution.

### Bipartite configuration model and network analysis

To identify products driving regional retail differentiation, we have built a bipartite network linking stores  $s$  to products  $p$  based on the cumulative number of days each product was on sale,  $X_{s,p}$ . First, we have applied RCA filtering to identify meaningful store-product associations. RCA evaluates whether a store puts on offer certain products more often than would be expected based on overall market trends<sup>65</sup>. It is calculated as the ratio of a store’s share of a particular product to the product’s share of total offerings across all stores:

$$\text{RCA}_{s,p} = \frac{\frac{X_{s,p}}{\sum_{p'} X_{s,p'}}}{\frac{\sum_{s'} X_{s',p}}{\sum_{s',p'} X_{s',p'}}}, \quad (2)$$

where  $\sum_{p'} X_{s,p'}$  is total number of product-days in store  $s$ ,  $\sum_{s'} X_{s',p}$  is the total number of days product  $p$  was on sale across all stores, and  $\sum_{s',p'} X_{s',p'}$  total number of product-days across all stores and all products. RCA values greater than 1 indicate relative specialization.

The RCA filtering yields an unweighted bipartite network that can be validated via the Bipartite Configuration Model<sup>62–64</sup>. BiCM defines a maximum-entropy null ensemble constrained by the observed degree sequences – that is, the number of unique products connected to each store and vice versa. The model assigns connection probabilities

$$p_{sc} = \frac{e^{-\lambda_s - \theta_c}}{1 + e^{-\lambda_s - \theta_c}}, \quad (3)$$

where  $\{\lambda_s\}$  and  $\{\theta_c\}$  are the multipliers enforcing the degree constraints for stores and products, respectively. From these probabilities, the expected number of common stores between two products  $a$  and  $b$  is just  $\langle V_{ab} \rangle = \sum_s p_{sa} p_{sb}$ , and the observed co-occurrence  $V_{ab}$  is tested against its Poisson-Binomial distribution via  $p[V_{ab}] = 1 - \sum_{x=0}^{V_{ab}-1} \pi(x|a, b)$ . Finally, we apply False Discovery Rate (FDR) to control for multiple hypothesis testing when validating individual links (significance level  $\alpha = 0.05$ ), ensuring that retained connections represent statistically significant deviations from random expectation rather than spurious correlations.

We perform community detection on the resulting projections using Infomap, which identifies modules by modeling information flow on the network as a random walk and optimizing a map equation<sup>68</sup>, and Louvain, which greedily maximizes modularity<sup>67</sup>—a measure of the density of edges inside communities versus edges between communities. The two resulting partitions were then compared using the Adjusted Rand Index (ARI) to assess their similarity based on the contingency matrix and statistical significance via permutation testing.

Percolation on the weighted (store-store) network is performed by using travel distance as the link weight. At a given distance threshold, only edges shorter than that threshold are retained (equivalently, distances can be normalized to bond-open probabilities), allowing connected stores to form clusters that merge as the threshold increases<sup>40</sup>.

### Data availability

The Salling Group API provides open access to food waste data through their developer portal <https://developer.sallinggroup.com/api-reference>. OSRM provides publicly accessible routing services through their API endpoints at <http://router.project-osrm.org/>. The Nutri-Score and environmental score mappings were obtained by querying the Open Food Facts database via its REST API <https://openfoodfacts.github.io/openfoodfacts-server/api/>. Municipality boundaries and resident population were obtained from the Copernicus Global Degree of Urbanisation Classification dataset [https://human-settlement.emergency.copernicus.eu/ghs\\_duc2023.php](https://human-settlement.emergency.copernicus.eu/ghs_duc2023.php). The data and code to reproduce the analysis is released on [GitHub](#). For inquiries, please contact A.D. antde@dtu.dk.

### Code availability

The data and code to reproduce the analysis is released on [GitHub](#). For inquiries, please contact A.D. antde@dtu.dk.

Received: 25 November 2025; Accepted: 10 February 2026;

Published online: 20 February 2026

### References

- Tufano, M. et al. Data-driven nutritional assessment of urban food landscapes: insights from Boston, London, and Dubai. *Sci. Rep.* **15**, 24453 (2025).
- Tilman, D. & Clark, M. Global diets link environmental sustainability and human health. *Nature* **515**, 518–522 (2014).
- Ispirova, G., Sebek, M. & Menichetti, G. Informatics for food processing. <https://arxiv.org/abs/2505.17087> Version Number: 1 (2025).
- Kuhl, E. AI for food: accelerating and democratizing discovery and innovation. *npj Sci. Food* **9**, 82 (2025).
- Menichetti, G., Ravandi, B., Mozaffarian, D. & Barabási, A.-L. Machine learning prediction of the degree of food processing. *Nat. Commun.* **14**, 2312 (2023).
- Ravandi, B. et al. Prevalence of processed foods in major US grocery stores. *Nat. Food* **6**, 296–308 (2025).
- Barabási, A.-L., Menichetti, G. & Loscalzo, J. The unmapped chemical complexity of our diet. *Nat. Food* **1**, 33–37 (2019).
- Horn, A. L. et al. Population mobility data provides meaningful indicators of fast food intake and diet-related diseases in diverse populations. *npj Digit. Med.* **6**, 208 (2023).
- Xu, R. et al. Integrating human activity into food environments can better predict cardiometabolic diseases in the United States. *Nat. Commun.* **14**, 7326 (2023).
- García Bulle Bueno, B. et al. Effect of mobile food environments on fast food visits. *Nat. Commun.* **15**, 2291 (2024).
- Wingrove, K. et al. Outcomes and impacts of community food hubs: a rapid review. *Proc. Nutr. Soc.* **84**, E50 (2025).
- Lane, M. M. et al. Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses. *BMJ* e077310. <https://doi.org/10.1136/bmj-2023-077310> (2024).
- Muncke, J., Touvier, M., Trasande, L. & Scheringer, M. Health impacts of exposure to synthetic chemicals in food. *Nat. Med.* **31**, 1431–1443 (2025).
- Fanzo, J. et al. Sustainable food systems and nutrition in the 21st century: a report from the 22nd annual Harvard nutrition obesity symposium. *Am. J. Clin. Nutr.* **115**, 18–33 (2022).
- Minor, K., Bjerre-Nielsen, A., Jonasdottir, S. S., Lehmann, S. & Obradovich, N. Rising temperatures erode human sleep globally. *One Earth* **5**, 534–549 (2022).
- Romanello, M. et al. The 2024 report of The Lancet Countdown on health and climate change: facing record-breaking threats from delayed action. *Lancet* **404**, 1847–1896 (2024).
- Nijdam, D., Rood, T. & Westhoek, H. The price of protein: review of land use and carbon footprints from life cycle assessments of animal food products and their substitutes. *Food Policy* **37**, 760–770 (2012).
- Springmann, M., Godfray, H. C. J., Rayner, M. & Scarborough, P. Analysis and valuation of the health and climate change cobenefits of dietary change. *Proc. Natl. Acad. Sci.* **113**, 4146–4151 (2016).
- Wang, Y. et al. Food waste used as a resource can reduce climate and resource burdens in agrifood systems. *Nat. Food* **6**, 478–490 (2025).
- Anshassi, M. & Townsend, T. G. Improving waste systems in the global south to tackle international environmental impacts. *Nat. Sustain.* <https://www.nature.com/articles/s41893-025-01607-8> (2025).
- Shepon, A., Eshel, G., Noor, E. & Milo, R. The opportunity cost of animal based diets exceeds all food losses. *Proc. Natl. Acad. Sci.* **115**, 3804–3809 (2018).
- Macdiarmid, J. I. The food system and climate change: are plant-based diets becoming unhealthy and less environmentally sustainable? *Proc. Nutr. Soc.* **81**, 162–167 (2022).
- Ambikapathi, R. et al. Global food systems transitions have enabled affordable diets but had less favourable outcomes for nutrition, environmental health, inclusion and equity. *Nat. Food* **3**, 764–779 (2022).
- Schneider, K. R. et al. The state of food systems worldwide in the countdown to 2030. *Nat. Food* **4**, 1090–1110 (2023).
- Crippa, M. et al. Food systems are responsible for a third of global anthropogenic GHG emissions. *Nat. Food* **2**, 198–209 (2021).
- Nerín, C., Aznar, M. & Carrizo, D. Food contamination during food process. *Trends Food Sci. Technol.* **48**, 63–68 (2016).
- Huang, I. Y. et al. Food waste management: a review of retailers' business practices and their implications for sustainable value. *J. Clean. Prod.* **285**, 125484 (2021).
- Campi, M., Dueñas, M. & Fagiolo, G. Specialization in food production affects global food security and food systems sustainability. *World Dev.* **141**, 105411 (2021).
- Filimonau, V. & Gherbin, A. An exploratory study of food waste management practices in the UK grocery retail sector. *J. Clean. Prod.* **167**, 1184–1194 (2017).
- Principato, L. *Food Waste at Consumer Level*. SpringerBriefs in Environmental Science. <https://doi.org/10.1007/978-3-319-78887-6> (Springer International Publishing, 2018).
- Hübner, A. H., Kuhn, H. & Sternbeck, M. G. Demand and supply chain planning in grocery retail: an operations planning framework. *Int. J. Retail Distrib. Manag.* **41**, 512–530 (2013).
- Tian, Y., Lautz, S., Wallis, A. O. G. & Lambiotte, R. Extracting complements and substitutes from sales data: a network perspective. *EPJ Data Sci.* **10**, 45 (2021).
- Fox, E. J., Montgomery, A. L. & Lodish, L. M. Consumer shopping and spending across retail formats. *J. Bus.* **77**, S25–S60 (2004).
- Pennacchioli, D., Coscia, M., Rinzivillo, S., Giannotti, F. & Pedreschi, D. The retail market as a complex system. *EPJ Data Sci.* **3**, 33 (2014).
- Moragues-Faus, A. & Battersby, J. Urban food policies for a sustainable and just future: Concepts and tools for a renewed agenda. *Food Policy* **103**, 102124 (2021).
- Aiello, L. M., Schifanella, R., Quercia, D. & Del Prete, L. Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci.* **8**, 14 (2019).
- Lee, M., Kim, H., Jun, B. & Park, J. We are what we buy: extracting urban lifestyles using large-scale delivery records. <https://arxiv.org/abs/2504.15618> Version Number: 1.(2025).

38. Dong, X. et al. Social bridges in urban purchase behavior. *ACM Trans. Intell. Syst. Technol.* **9**, 1–29 (2018).
39. Seto, K. C. & Ramankutty, N. Hidden linkages between urbanization and food systems. *Science* **352**, 943–945 (2016).
40. Arcaute, E. et al. Cities and regions in Britain through hierarchical percolation. *R. Soc. Open Sci.* **3**, 150691 (2016).
41. Lenormand, M. et al. Comparing and modelling land use organization in cities. *R. Soc. Open Sci.* **2**, 150449 (2015).
42. Bassolas, A. et al. Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 4817 (2019).
43. Di Clemente, R., Strano, E. & Batty, M. Urbanization and economic complexity. *Sci. Rep.* **11**, 3952 (2021).
44. Lengyel, B., Bokányi, E., Di Clemente, R., Kertész, J. & González, M. C. The role of geography in the complex diffusion of innovations. *Sci. Rep.* **10**, 15065 (2020).
45. Yang, Y., Pentland, A. & Moro, E. Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Sci.* **12**, 15 (2023).
46. Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961 (2013).
47. Bettencourt, L. M. A., Lobo, J., Strumsky, D. & West, G. B. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5**, e13541 (2010).
48. Sapienza, A., Litlál, M., Lehmann, S. & Alessandretti, L. Exposure to urban and rural contexts shapes smartphone usage behavior. *PNAS Nexus* **2**, pgad357 (2023).
49. PACE Circular. Denmark against food waste (n.d.). <https://pacecircular.org/denmark-against-food-waste>.
50. Otto, I. M. et al. Social tipping dynamics for stabilizing earth's climate by 2050. *Proc. Natl. Acad. Sci.* **117**, 2354–2365 (2020).
51. Committee on a Systems Approach to Reducing Consumer Food Waste et al. *A National Strategy to Reduce Food Waste at the Consumer Level*, 25876. <https://www.nap.edu/catalog/25876> (National Academies Press, 2020).
52. Johansen, D., Friis, K., Skovenborg, E. & Grønbæk, M. Food buying habits of people who buy wine or beer: cross sectional study. *BMJ* **332**, 519–522 (2006).
53. Buisman, M., Haijema, R. & Bloemhof-Ruwaard, J. Discounting and dynamic shelf life to reduce fresh food waste at retailers. *Int. J. Prod. Econ.* **209**, 274–284 (2019).
54. Kayikci, Y., Demir, S., Mangla, S. K., Subramanian, N. & Koc, B. Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers. *J. Clean. Prod.* **344**, 131068 (2022).
55. Atan, Z., Honhon, D. & Pan, X. A. Displaying and discounting perishables: impact on retail profits and waste. *Management Science, Articles in Advance*, pp. 1–18 (2025).
56. Anderson, E. T. & Fox, E. J. How price promotions work: A review of practice and theory. In *Handbook of the Economics of Marketing*, Vol. 1, 497–552 <https://linkinghub.elsevier.com/retrieve/pii/S2452261919300061> (Elsevier, 2019).
57. Salling Group. Food waste (n.d.). <https://sallinggroup.com/en/sustainability/climate/food-waste>.
58. Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. in *Advances in Knowledge Discovery and Data Mining*, vol. 7819, 160–172 (eds Pei, J., Tseng, V. S., Cao, L., Motoda, H. & Xu, G.) [https://doi.org/10.1007/978-3-642-37456-2\\_14](https://doi.org/10.1007/978-3-642-37456-2_14) (Springer Berlin Heidelberg, 2013).
59. Marin, V., Molinero, C. & Arcaute, E. The scalar mismatch of regional governance: a comparative analysis of hierarchical structures. *Environ. Plan. B Urban Anal. City Sci.* **51**, 2126–2145 (2024).
60. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
61. Diniz-Filho, J. A. F. et al. Mantel test in population genetics. *Genet. Mol. Biol.* **36**, 475–485 (2013).
62. Saracco, F., Di Clemente, R., Gabrielli, A. & Squartini, T. Randomizing bipartite networks: the case of the world trade web. *Sci. Rep.* **5**, 10595 (2015).
63. Saracco, F. et al. Inferring monopartite projections of bipartite networks: an entropy-based approach. *N. J. Phys.* **19**, 053022 (2017).
64. Cimini, G. et al. The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71 (2019).
65. Balassa, B. & Noland, M. "Revealed" comparative advantage in Japan and the United States. *J. Econ. Integr.* **4**, 8–15 (1989).
66. Knowles, R. D. Transit oriented development in Copenhagen, Denmark: from the finger plan to Ørestad. *J. Transp. Geogr.* **22**, 251–261 (2012).
67. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
68. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
69. Halloran, A., Clement, J., Kornum, N., Bucatariu, C. & Magid, J. Addressing food waste reduction in Denmark. *Food Policy* **49**, 294–301 (2014).
70. Doğan, N. B., Mejia, A. & Gomez, M. Cities can benefit from complex supply chains. *npj Urban Sustain.* **3**, 20 (2023).
71. Andreani, G., Sogari, G., Wongprawmas, R., Menozzi, D. & Mora, C. Nutri-score and eco-score labeling: A systematic review of their impact on consumer understanding, attitudes, and behaviors. *Food Rev. Int.* **41**, 1981–2005 (2025).
72. Huber, S. & Rust, C. osrmtime: Calculate travel time and distance with OpenStreetMap data using the open source routing machine (OSRM). *SSRN Electron. J.* <https://www.ssrn.com/abstract=2691551> (2016).
73. European Commission. Joint Research Centre. *GHSL data package 2023*. <https://doi.org/10.2760/098587> (Publications Office of the European Union, 2023).

### Acknowledgements

A.D. acknowledges the helpful discussions and insightful feedback with Riccardo Di Clemente, Yuan Liao, Giovanni Mauro, Fabio Saracco and Michele Tizzani.

### Author contributions

A.D. and F.B. gathered the data. F.B. performed preliminary analysis. A.D. and A.G. performed the analysis and made the figures. A.D. and S.L. designed the analysis. S.L. supervised the project. All authors discussed the results and contributed to the final 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/s41538-026-00764-0>.

**Correspondence** and requests for materials should be addressed to Antonio Desiderio or Sune Lehmann.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026