Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Food co-purchasing networks encode geographic disparities in cardiometabolic health

Antonio Desiderio

desiderio.d.antonio@gmail.com

ISI Foundation https://orcid.org/0000-0002-9971-450X

Leonardo Rodovero

Technical University of Denmark

Sune Lehmann

Technical University of Denmark https://orcid.org/0000-0001-6099-2345

Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11756230/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Food co-purchasing networks encode geographic disparities in cardiometabolic health

Antonio Desiderio[a,b,c,*], Leonardo Rodovero[a], Sune Lehmann[a,d,*]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800, Copenhagen, Denmark.

[b]ISI Foundation, Via Chisola 5, 10126, Turin, Italy.

[c]Centro Ricerche Enrico Fermi, Via Panisperna 89a, 00184, Rome, Italy.

[d]Copenhagen Center for Social Data Science (SODAS), University of Copenhagen, Øster Farimagsgade 5, 1353, Copenhagen, Denmark.

*Corresponding authors: antde@dtu.dk and sljo@dtu.dk.

Abstract

Unhealthy diets are a major driver of cardiometabolic disease, yet the geospatial structure of dietary behavior and its links to population health remain poorly understood. While large-scale retail transaction data offer a unique view of population dietary behavior, conventional approaches reduce purchases to isolated items, not accounting for how foods are actually combined. Here, we analyze 51 million grocery receipts across 95 Danish municipalities using a statistically validated bipartite network framework to isolate co-purchases that product popularity alone cannot explain. We find that, although municipalities buy largely the same products, their co-purchasing networks reorganize along a continuous urban-to-peripheral gradient in how those products are combined. The network backbone can be resolved into modular dietary communities whose local cohesion allows us to predict cardiometabolic prescriptions, explaining a quarter to a half of the variance left after sociodemographic adjustment, beyond what product prevalence alone accounts for. Ultra-processed foods (UPFs) show no spatial gradient in prevalence and are located at the network periphery, where they share baskets with fresh foods but rarely form the combinations that hold the network together. Taken together, our findings show that geographic dietary differences are encoded not only in what people purchase, but in how foods are systematically combined within shopping baskets, offering a scalable network approach for population nutritional epidemiology.

Introduction

Non-communicable diseases, including cardiovascular disease and type 2 diabetes, impose a major global health burden [1, 2], and unhealthy dietary patterns contribute significantly to their incidence and mortality [3–11]. Food-system transformation [12–14] – driven by industrialization, urbanization, and the expansion of retail chains [15–18] – has reconfigured the availability and affordability of food across the past decades [19–21], while the growing market penetration and widespread consumption of ultra-processed foods [20, 22–28] have changed individual dietary profiles, in a society-wide shift often associated with adverse cardiometabolic outcomes [8, 29–36]. These health risks are not distributed uniformly across society; there is geographical variation in dietary patterns that reflects spatial and socioeconomic inequalities [37–40], rooted in underlying structural disparities in local food environments [12, 41–46].

Measuring how diets vary across places, however, is difficult. Traditional nutritional epidemiology relies on dietary surveys and longitudinal cohorts, which provide detailed information on food consumption patterns but are constrained by self-reporting biases, high implementation costs, and limited spatial coverage [2, 47]. Importantly, this tradition has long recognized that foods are eaten in combination, as dietary-pattern analyses capture how individuals combine foods, and these patterns reveal diet–disease relationships that single foods or nutrients miss [3, 48, 49].

Computational methods now offer a way around the constraints of surveys, opening up a data-driven nutritional epidemiology [50]. This field uses large-scale digital records, such as retail transaction data [45, 51–56], to trace the complex system of human interactions with food at a level of detail that was previously out of reach [57]; such records provide unprecedented opportunities to quantify dietary habits across entire populations [58, 59]. Yet these rich transaction datasets are almost always reduced to product- or category-level summaries [45, 51–54, 60], discarding exactly the combination information that dietary-pattern research has shown to matter [61, 62].

Bringing the dietary-pattern perspective to transaction data – understanding how the network organization of food co-purchases varies across geographic contexts, and whether this relational structure encodes population health signals beyond individual product prevalence – is therefore the natural next step in nutritional modeling, and a new way to identify drivers of nutritional disparities [53, 54, 63–65].

Here, we characterize population dietary behavior through food co-purchasing patterns, and ask what food combinations reveal that individual products cannot [61, 63]. By examining how foods are combined rather than considering individual products in isolation, we identify a layer of dietary structure that varies across places and tracks population health. Specifically, we develop a network-based framework to identify non-random co-purchasing relationships and examine how their organization varies across geographic contexts [53, 66–68]. We then assess how these network structures capture population-level cardiometabolic health signals beyond individual product prevalence and demographic composition [59, 65, 69]. We apply our approach to large-scale grocery purchasing data from Denmark, where eating out is relatively rare: about 85% of consumers describe their most recent dinner as home-prepared, leftovers, or based on ready-made components [70], so grocery purchases capture a substantial share of the food the

population eats. We find that co-purchasing networks reorganize along a continuous urban-to-peripheral gradient, and that dietary community cohesion tracks municipal cardiometabolic prescriptions beyond sociodemographics and local product prevalence. Notably, while ultra-processed food sales show no geographic variation, these products consistently sit at the network periphery.

Results

Every receipt from a nationwide supermarket chain. To investigate the spatial organization of dietary habits and its relationship with population-level cardiometabolic health, we draw on retail purchasing records combined with national nutrition databases [71] and official public-health registries in Denmark [72]. Through a collaboration with Salling Group (Denmark's largest retailer), we analyzed 51 million point-of-sale receipts from 553 Netto supermarket locations across 95 of Denmark's 98 municipalities, collected over 79 days in 2025 (see Methods). To isolate dietary purchases, non-food transactions (carrier bags, personal care, household goods, seasonal non-food merchandise, and related categories) were removed, and edible items were harmonized at the brand-and-microcategory level (e.g. "Brand of Milk" – whole milk), yielding 3,405 "products" after exclusion (see Methods and Supplementary Materials, Section 1).

Receipt data linked to nutritional and processing measures. We matched each food product to a reference entry in the Danish national food-composition database (Frida) [71] using a hierarchical multi-stage semantic matching approach (see Methods for the complete pipeline). This mapping establishes both a nutritional and an industrial processing profile for each item, capturing energy-adjusted macronutrients alongside a continuous ultra-processed food (UPF) risk index ($F\text{Score} \in [0,1]$) and categorical NOVA classes [22, 23], estimated via machine learning (FoodProX) [62] and benchmarked against Open Food Facts records (see Methods for further details).

To evaluate population-level outcomes, we aggregated these item-level metrics and aligned the resulting municipal grocery profiles with 2024 national health registries from Statistics Denmark across the 95 municipalities with retail coverage [72]. The cardiometabolic burden for a municipality was measured as the share of residents who redeemed at least one prescription in 2024 for antidiabetic medicines (ATC A10), cardiovascular-system medicines (ATC C), or lipid-modifying agents (ATC C10). We note that these rates are treated-disease proxies rather than clinical diagnoses; for instance, A10 combines insulin and non-insulin therapies, effectively capturing treated diabetes as a whole – roughly 93% of which is type 2 (estimated from regional rather than municipality-level data; see Methods). Finally, we linked municipal grocery and health series to Statistics Denmark's municipality-level data for population, age, and disposable family income (see Methods for further details).

The geography of how foods are co-purchased

Co-occurrences beyond product popularity. Raw co-occurrence frequencies [53, 54, 63, 73] can confound genuine dietary affinities because of the heavy-tailed distribution of product popularity: ubiquitous staples such as milk or bread will tend to co-occur frequently with a wide range

of items simply because their high purchase rate inflates the baseline probability of random overlaps, rather than because they reflect specific dietary complementarity. To filter out these popularity artifacts, we validate basket-level co-purchasing against the Bipartite Configuration Model (BiCM) [66–68], representing transactions as a bipartite network of shopping baskets and food products and preserving observed marginal degrees – namely, basket sizes and product purchase frequencies (see Figure 1**a** and Methods). The resulting validated backbone retains only product pairs that co-occur significantly more often than expected by chance. We extract this backbone for each Danish municipality, allowing regional dietary structures to emerge from the way foods are combined in actual purchases rather than from differences in overall product popularity or purchasing volume.

While products are broadly shared, foods are combined differently across the country. Using our co-purchasing backbone for each municipality, we first examine whether we find a uniform 'national diet' or distinct local patterns of food associations. We compare the overall wiring of the municipal networks using the spectral distance between their normalized Laplacians, which captures differences in network structure independently of geographic distance (see Methods for further details). Denmark's four largest cities – København, Aarhus, Odense and Aalborg – stand out as a distinct cluster, which we call the *urban core* (pairwise spectral distances below 10, against a median of 25 among the remaining municipalities; hierarchical clustering isolates the same four-city block). Clustering on shared edges instead recovers the same partition (Jaccard distance between validated edge sets; adjusted Rand index 0.97). Conversely, municipal product composition is highly uniform across the country (median Jaccard similarity $= 0.65$) and does not isolate the four-city cluster, yielding no clear geographic partitions beyond a few small-catalog outliers.

Measuring each municipality's mean spectral distance d to the urban core gives us a continuous urban-to-peripheral gradient, mapped in Figure 1**b** alongside the networks of København and Norddjurs (a non-core municipality with an overlapping catalog), drawn with identical node coordinates to highlight differences in edge wiring. For instance, across ∼1,000 common products, København shows a more heterogeneous co-purchasing topology (degree entropy 0.66 vs. 0.39; random-walk entropy rate 4.30 vs. 2.96), a contrast that persists under equalized sample sizes and matched basket-size distributions. Holding out the urban core, we partition the remaining 91 municipalities at the median distance ($d = 25.1$) into *near* ($n = 47$, median $d = 22$) and *mid–far* ($n = 44$, median $d = 29$) groups. Along this gradient, edge Jaccard similarity with the urban core drops monotonically from 0.55 within the core to 0.25 in the near group and 0.12 in the mid–far group (Kruskal–Wallis test $H = 70.7$, $p < 0.001$; Mann–Whitney test near vs. mid–far, $p < 0.001$). Recomputing spectral distances strictly on pairwise shared products without spectrum padding preserves this urban-to-peripheral ordering ($\rho = 0.78$ pairwise and $\rho = 0.85$ to the urban core), and the geographic patterns remain stable across per-capita sampling ratios and random sampling seeds (see Methods and Supplementary Materials, Section 2). Per-capita sampling retains population scale as part of the ecological description of municipal purchasing activity, although spectral distance remains strongly associated with sample size ($\rho = -0.975$) and validated-catalog size ($\rho = -0.999$). To isolate basket composition from sample size under Bonferroni validation, we first compared København and Norddjurs directly at equal sample sizes, under a size-invariant effect threshold ($z/\sqrt{N}$), and under permutations matched on basket size. Across all three controls, the spectral distance between København and Norddjurs stays

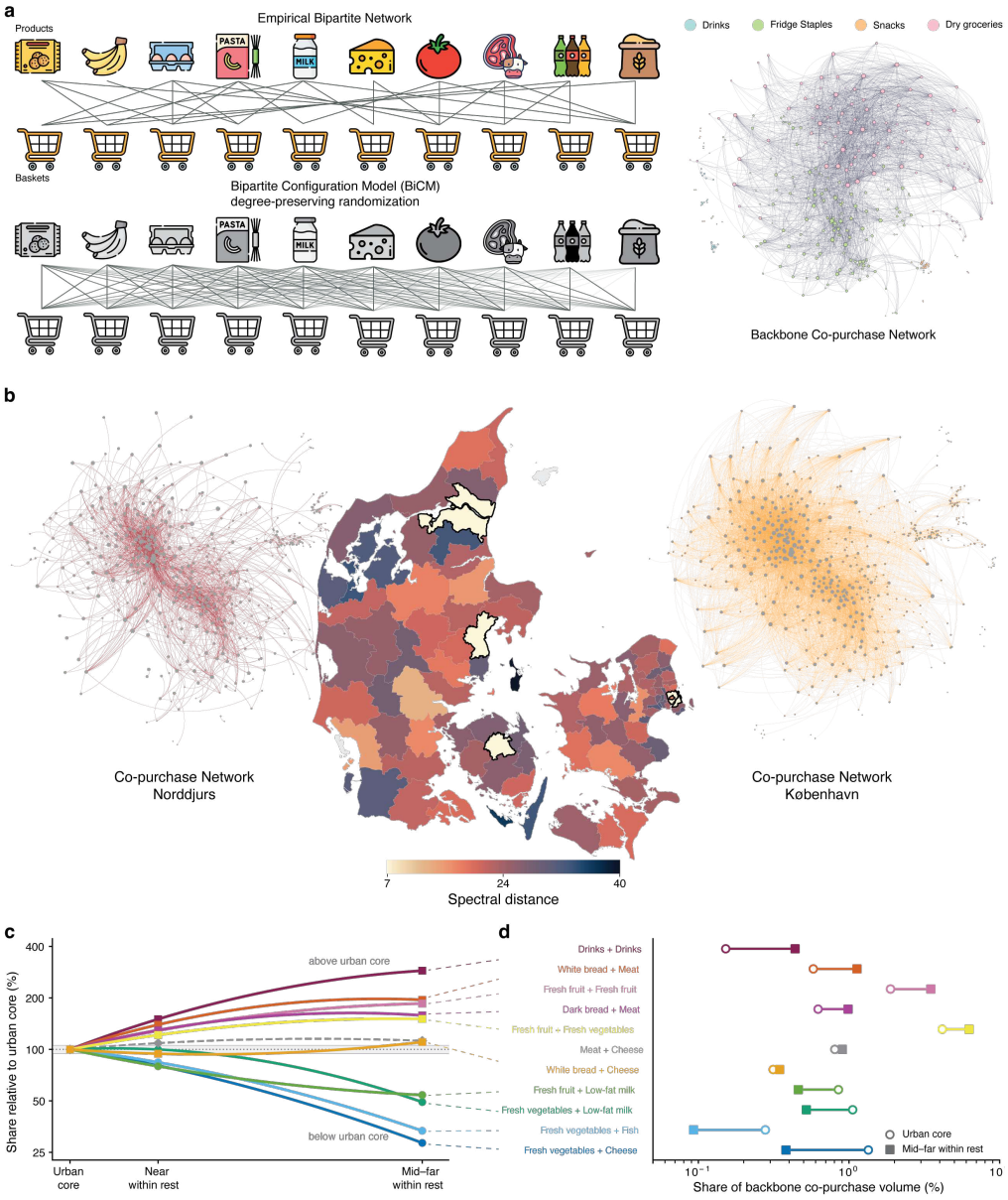


Fig. 1 Statistical validation and spatial reorganization of food co-purchasing networks across Denmark. a) Empirical basket–product records (top left) are compared with a Bipartite Configuration Model (BiCM, bottom left; edges are shaded proportionally to connection probabilities under the null model), which preserves observed basket sizes and product degrees. Statistically validated associations are projected onto a monopartite product–product network – representing the Randers co-purchasing network (right, nodes colored by broad product group) – retaining co-purchasing relationships occurring more often than expected under the BiCM. **b)** Danish municipalities colored by their spectral distance from the four-city urban core (outlined), based on the normalized-Laplacian spectra of their validated networks; municipalities without data are gray. Norddjurs (left) and København (right) are shown with identical node positions, illustrating topological rewiring despite a largely shared product set. On this shared product set, normalized degree entropy is 0.66 in København and 0.39 in Norddjurs, and the random-walk entropy rate is 4.30 and 2.96; this topological contrast persists under equalized sample sizes, basket-size matching, and effect-size filtering (see Supplementary Materials, Section 2). **c)** Mean share of backbone co-purchase volume for selected meta-category pairs, relative to the urban-core mean (100% = parity) across the urban core, near, and mid–far groups; pair colors as in **d**. Warm and cool colors indicate increases and decreases away from the urban core, respectively. Points show group means; curves are $\log_2$-space interpolations. The gray band denotes the largest standard error of the mean of the meat–cheese control pair. **d)** Absolute mean share of backbone co-purchase volume for selected pairs in the urban core (open circles) and mid–far municipalities (filled squares), highlighting a peripheral shift toward staple and within-category pairings alongside a decline in fresh food–dairy/fish combinations.

well above the label-permutation null ($p < 0.001$), with Odense at closer distance and neighboring Frederiksberg indistinguishable from København ($p \geq 0.10$). Extending this equal-sample, basket-size-matched design to all 95 municipalities ($N^* = 53{,}251$ baskets) consolidates 21 of the 27 smaller Capital Region suburbs into the closest tertile around København, while the largest regional cities remain distinct from their rural surroundings. This suburban consolidation compresses edge overlap when evaluated on the unadjusted population-scaled groups (0.47, 0.45, and 0.43), yet municipal rankings outside the Capital Region remain consistent with the per-capita gradient ($\rho = 0.61$, $p < 10^{-7}$) and edge overlap with København declines monotonically across fixed-N^* tertiles (0.44, 0.40, and 0.36; $\rho = -0.85$ nationwide and $\rho = -0.80$ outside the Capital Region; see Supplementary Materials, Section 2).

Shifts in food co-occurrence patterns across the urban–peripheral gradient. Having established that co-purchasing networks are reorganized along a spectral urban-to-peripheral gradient despite largely shared product catalogs, we next ask which food combinations carry this rewiring. For each municipality, we take the backbone's co-purchase volume – the number of baskets containing each validated pair, summed over all pairs – and compute the share falling on each pair of meta-categories. We then track each pair's mean share relative to the urban core along spectral distance (see Figure 1**c** and **d**). Pairings between fresh foods and dairy or fish, which are more prominent in the urban core, decline towards the mid–far municipalities, with vegetables–cheese falling from 1.34% to 0.38% of the backbone (28% of the urban-core share; $\rho = -0.89$, $p < 0.001$), vegetables–fish reaching 33% of the urban-core share ($\rho = -0.50$, $p < 0.001$), and vegetables or fruit with low-fat milk to 49–54% ($\rho = -0.58$, $p < 0.001$, and $\rho = -0.23$, $p = 0.025$). Conversely, fruit–vegetable rises from 4.16% to 6.27% (151% of the urban-core share) and fruit–fruit from 1.89% to 3.49% (185%; $\rho = 0.59$ and 0.76; both $p < 0.001$), while white bread–meat reaches 195% of the urban-core share ($\rho = 0.42$, $p < 0.001$). Drinks–drinks shows the steepest relative increase (290% of the urban-core share; $\rho = 0.61$, $p < 0.001$), while remaining a small share of the backbone. These urban-to-rural shifts in food pairings hold both on unpadded shared-product spectra and when comparing rural municipalities directly to the urban core under fixed-N^* sampling (with rural-to-urban ratios of 0.58 vs. 0.59 for vegetables–cheese, 0.45 vs. 0.30 for vegetables–fish, 1.84 vs. 1.65 for white bread–meat, and 2.13 vs. 2.20 for drinks–drinks; see Supplementary Materials, Section 2). Thus, despite the nationwide standardization of supermarket supply chains, the ways in which consumers combine products in shopping baskets remain strongly localized.

Food co-purchasing structure is associated with cardiometabolic health

Deriving a robust national food co-purchasing backbone. To evaluate whether and how local food purchasing structures correlate with municipal-level differences in cardiometabolic health [54], we need a shared national reference space across municipalities. To establish this framework, we pool population-normalized receipts and apply the bipartite configuration model (BiCM) to construct a national co-purchasing backbone (see Methods). We then filter the backbone using the BiCM-derived z-score at a percolation threshold of $z^* = 28$, retaining 95% of products within a single giant connected component (2,659 nodes and 28,435 edges, resulting in a density of 0.008) [66, 67]. We choose this threshold to preserve network connectivity while reducing the dense catalog of significant co-purchases; importantly, z-score filtering isolates the

structural backbone without biasing the underlying community partition (see Supplementary Materials, Section 3). The resulting sparse network has community structure, with a modularity of $Q = 0.48$. Community detection identifies 12 major coherent dietary communities – including Ready Meals, Cheese & cold cuts, Candy & snacks, Wine & beer, and Pasta & toppings. Crucially, this mesoscale organization is structurally robust. We find strong agreement between two community detection methods built on different principles, Louvain [74] and Infomap [75] (Adjusted Mutual Information ≈ 0.70). A Jaccard-based co-occurrence network [53, 63], however, does not recover this organization, remaining dense (~ 1.2 million edges, density ~ 0.3) with little modular structure at the median threshold, while progressively stricter thresholds reduce connectivity without revealing a comparable community backbone ($Q \lesssim 0.2$).

Local community cohesion predicts municipal prescription rates. To relate national dietary communities to cardiometabolic burden, we summarize each municipality by the intensity of its locally dense co-purchase communities, measured as the mean intra-community log-lift I_c of the national backbone evaluated on local baskets (see Methods for details). First, we predict age-aggregated 2024 redemption rates (see Figure 2**b**) from the intensity vector I_c alone and jointly with sociodemographic block S (log population, shares aged < 30 and ≥ 60, and affluence). Network intensities alone yield leave-one-out R^2 of 0.20, 0.70, and 0.61 and sociodemographics alone 0.52, 0.84, and 0.54, while the joint model reaches 0.61, 0.94, and 0.75, respectively, corresponding to an additional 18%, 63%, and 43% of the variance remaining after sociodemographic adjustment (see Figure 2**c**). To ensure that age composition does not confound or attenuate the network signal in age-aggregated rates, we repeat the analysis using 2024 redemption rates restricted to residents aged ≥ 60, excluding age composition from S. Here, the network alone yields R^2 of 0.27, 0.58, and 0.50, and the joint model reaches 0.49, 0.61, and 0.46, with the sociodemographic baseline explaining less than 0.07 in all cases. For C10, the joint model (0.46) falls slightly below the network alone (0.50), as adding uninformative features introduces noise that can penalize out-of-sample performance. Thus, local co-purchase cohesion accounts for approximately 49%, 58%, and 48% of the residual variance, respectively, with a similar residual contribution across both diabetes management and cardiovascular disease burden. Recomputing municipal community intensities I_c on the fixed-N^* basket-size-matched samples yields nearly identical out-of-sample predictive performance (R^2_{S+I} of 0.59, 0.93, and 0.73 across all ages and 0.43, 0.57, and 0.47 among residents aged ≥ 60) and preserves the sign of all headline community associations (see Supplementary Materials, Section 2).

Disentangling co-purchase cohesion from baseline product volumes. The lift normalizes co-occurrence by product-level frequencies, but it is still possible for I_c to covary with how strongly a community is purchased across municipalities. To account for this effect, we compare I_c to both local popularity, $M = \text{mean}(\log n_i + \log n_j)$, and co-purchase residualized on local product prevalence, $\widetilde{P}$, both computed on the same BiCM-validated intra-community edges. On age-aggregated rates, with municipal log population, age composition, and affluence as controls, I_c and M yield similar fits at the block level ($S + I$ vs. $S + M$: $R^2 \approx 0.61$ vs. 0.60 for A10; 0.94 vs. 0.95 for C; 0.75 vs. 0.76 for C10). We therefore further examine $\widetilde{P}$, obtained by partialling $\text{mean}(\log P_{ij})$ on $\text{mean}(\log P_i) + \text{mean}(\log P_j)$ across municipalities. This strategy allows us to separate variation in co-purchasing from variation in local product prevalence while retaining the same BiCM-defined network structure. After S, we find that $\widetilde{P}$ still adds $\Delta R^2 \approx 0.12$, 0.07,

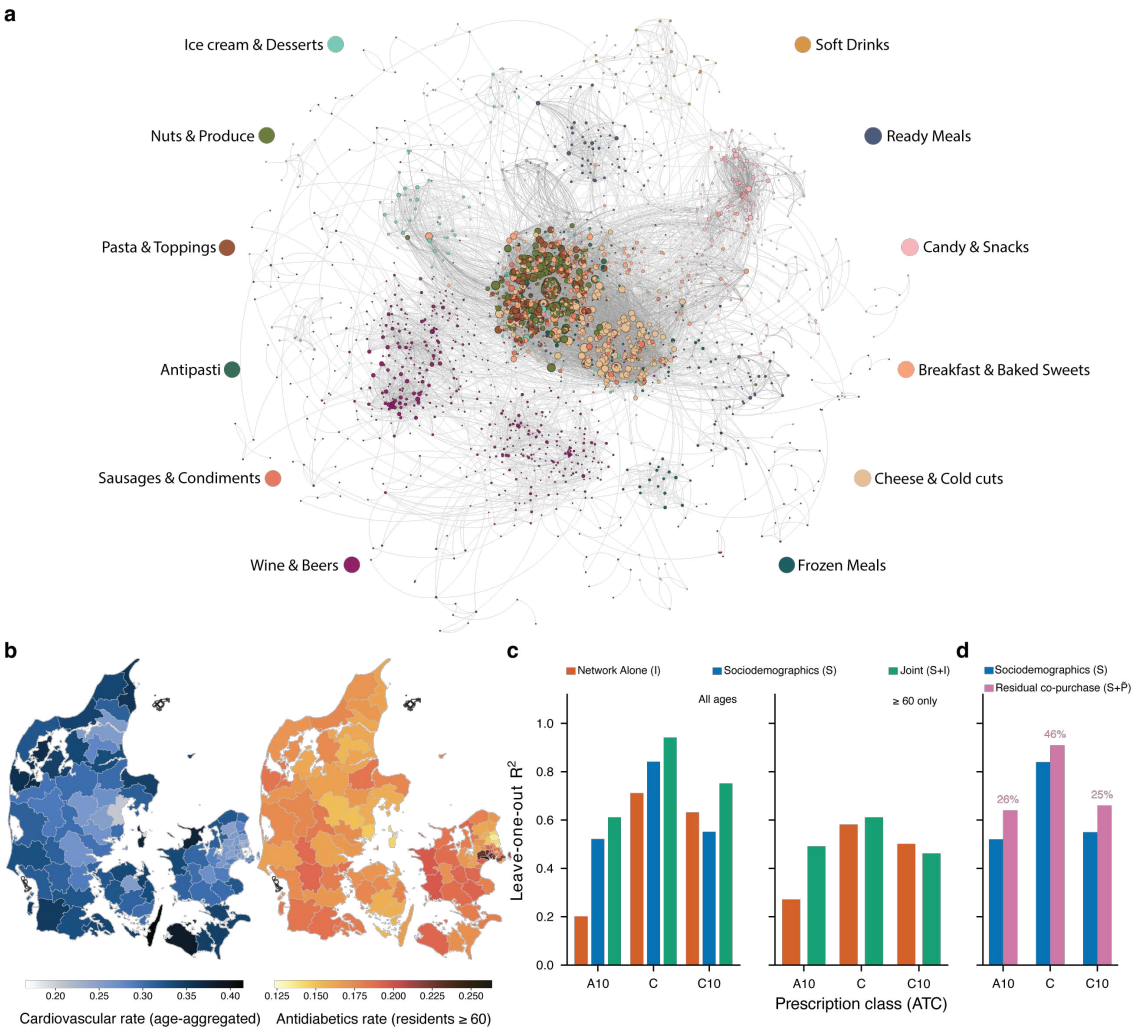


Fig. 2 Food co-purchasing network and its association with cardiometabolic health disparities across Denmark. a) National validated co-purchasing backbone ($z^* = 28$; 2,659 nodes, 28,435 edges), partitioned into 12 major dietary communities (node colors) with strong modular organization ($Q = 0.48$). **b)** Share of residents redeeming at least one prescription in 2024 for cardiovascular medicines (ATC C, all ages; left) and antidiabetic medicines (ATC A10, residents aged $\geq$ 60; right). Fanø and Læsø are white-hatched (no complete coverage; excluded), Bornholm omitted. **c)** Out-of-sample predictive performance (leave-one-out R^2) for antidiabetic (A10), cardiovascular (C), and lipid-modifying (C10) prescription rates, for all ages (left) and residents aged $\geq$ 60 (right). Bars compare local network intensity alone (I, orange), sociodemographics alone (S, blue; omitted for residents aged $\geq$ 60 as $R^2 < 0.07$), and both together ($S + I$, green). **d)** Predictive gain from co-purchasing structure beyond product prevalence (all ages). Adding residualized co-purchase cohesion ($S + \widetilde{P}$, pink) to sociodemographics (S, blue) improves the fit for all three outcomes, closing 25–46% of the variance left unexplained by S (percentages above the bars; see Supplementary Materials, Section 4 for the community-level regression coefficients).

and 0.11 for A10, C, and C10 ($R^2_{S+\widetilde{P}} \approx 0.64$, 0.91, and 0.66), closing about a quarter of the variance left unexplained by sociodemographics for A10 and C10 and nearly half for C (26%, 46%, and 25%; see Figure 2**d**). At the community level, municipalities where Candy & snacks and Cheese & cold cuts are more cohesively co-purchased have a higher share of residents on diabetes medication (A10), for both I_c and $\widetilde{P}$ and after conditioning on sociodemographic factors and local

product prevalences, while Wine & beer cohesion tracks cardiovascular and lipid-lowering medication (C and C10) (see Supplementary Materials, Section 4 for the community-level regression coefficients).

Ultra-processed food prevalence and network position in co-purchasing networks

The share of ultra-processed food shows no urban-to-peripheral gradient. Because ultra-processed foods (UPFs) [22] reshape household procurement and concentrate along spatial retail gradients [20, 26], we test whether the urban-to-peripheral rewiring of co-purchasing networks reflects differences in UPF purchase prevalence or in the topological position of ultra-processed products. To quantify processing at the product level, we use FoodProX scores from the retail-Frida mapping [55, 62], which separates NOVA 1 from NOVA 4 products with 78.3% accuracy in our catalog (see Methods and Supplementary Materials, Section 5). The UPF share of purchases varies little across municipalities (from 0.46 to 0.56) and is unrelated to distance from the urban core ($\rho = -0.02$, $p > 0.05$). This holds whether products are labeled with hard NOVA classes or with FoodProX's continuous scores (class probabilities, *F*Score, high-confidence labels only, or resampled labels), and the choice does not change any of the results below.

Ultra-processed foods are located at the periphery of the co-purchasing network. Receipts containing both a NOVA 1 and a NOVA 4 item hold 76.6% of mapped food purchase volume (see Supplementary Materials, Section 6). Despite this within-basket co-occurrence, UPFs and minimally processed foods do not occupy the same structural role in local co-purchasing network topologies (see Figure 3**a** for København). When products are ranked into degree deciles, UPFs are under-represented among the most connected products in all 95 municipalities (see Figure 3**b**). The probability that a randomly chosen UPF has more co-purchase partners than a randomly chosen minimally processed food (the NOVA 4 versus NOVA 1 degree AUC) is below 0.5 in every municipality (mean AUC = 0.43), indicating that UPFs consistently occupy peripheral rather than hub-like positions in local food networks (see Figure 3**b**). This peripheral role does not imply purchasing segregation across separate shopping events. Evaluating the baskets underlying each co-purchasing tie reveals that the minimally processed (NOVA 1–1) core of the network is supported by baskets where ultra-processed items already account for approximately 35% of purchase volume. By contrast, NOVA 4–4 ties occur in baskets with nearly 60% UPFs, indicating that the UPF–UPF pole concentrates within carts that are already UPF-dense rather than defining an exclusive UPF basket type (see Supplementary Materials, Section 6). In other words, shoppers do not make separate *healthy* and *junk* trips, as UPFs travel in the same baskets as fresh food, but they are rarely part of the specific combinations that hold the network together, and they pair up with each other mainly in the most UPF-heavy baskets.

Along the geographic gradient, network structure reorganizes but ultra-processed food positioning remains invariant. We next consider whether this shared UPF peripheral position is associated with geographic differences in how processed products connect and in how strongly UPF–UPF relationships are expressed. Neither the co-purchase mass of NOVA 4–4 pairs (C_{44}) nor their excess co-prevalence above product-frequency expectations (ΔC_{44}) shows a significant geographic gradient ($\rho = 0.06$ and $\rho \approx 0$, respectively; $p > 0.05$), indicating that aggregate

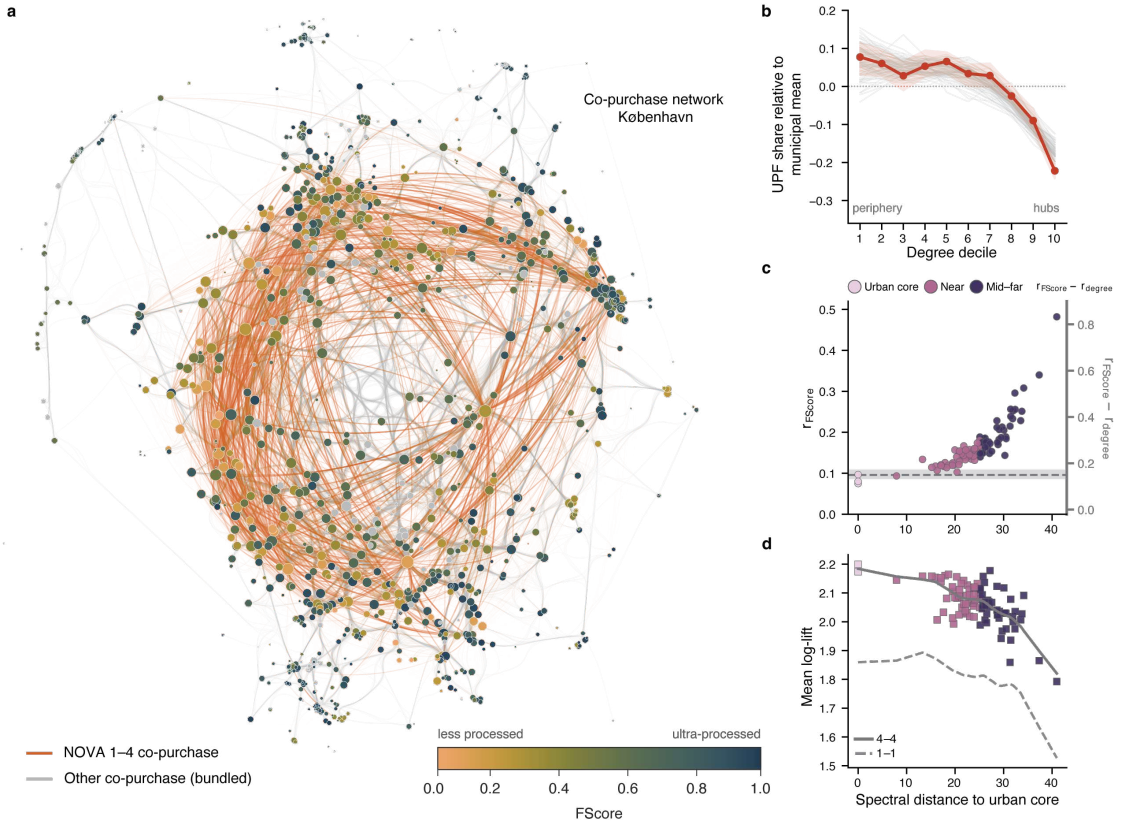


Fig. 3 Ultra-processed foods are located at the periphery of co-purchasing networks across the country. a) Co-purchasing network for København. Node color shows the processing score *F*Score (gray where no score is available), and node size scales with the logarithm of degree. Orange edges are co-purchases between minimally processed (NOVA 1) and ultra-processed (NOVA 4) products (6,012 links, 27% of all edges), with opacity increasing with local co-purchase strength; all other links are shown in gray and edge-bundled. Minimally processed products form the hubs of the network, while UPFs are linked to them throughout rather than forming a separate cluster. **b)** UPF share within each degree decile, from the periphery (1) to the hubs (10), relative to the municipality's overall UPF share (gray lines, 95 municipalities; red, mean). UPFs are under-represented among the most connected products; the NOVA 4 versus NOVA 1 degree AUC is below 0.5 in every municipality (mean 0.43). **c)** Processing assortativity r_{FScore}, the tendency of products with similar processing scores to be co-purchased, against spectral distance from the urban core. It rises along the urban-to-peripheral gradient ($\rho = +0.89$, $p < 0.001$) and exceeds a null model that shuffles *F*Score across products in all 95 municipalities ($z > 2$). Its excess over degree assortativity, $r_{\text{FScore}} - r_{\text{degree}}$ (dashed line and band, right axis), stays positive but shows no gradient. **d)** Local strength (mean log-lift) of the nationally validated NOVA 4–4 pairs (points, solid line) and NOVA 1–1 pairs (dashed line). Both decline with spectral distance ($\rho = -0.56$, $p < 0.001$ for 4–4; $\rho = -0.27$, $p = 0.007$ for 1–1), while the 4–4 over 1–1 gap remains positive in all 95 municipalities and shows no significant gradient ($\rho = -0.16$, $p = 0.13$).

UPF–UPF co-purchasing does not systematically rise or fall along the spectral axis. In contrast, assortativity in processing scores, r_{FScore}, increases strongly with spectral distance from the urban core ($\rho = +0.89$, $p < 0.001$; see Figure 3**c**), alongside rising degree assortativity. Relative to a node-label shuffle of *F*Score on the same graphs, r_{FScore} exceeds the null in all 95 municipalities ($z > 2$). The excess assortativity beyond degree, $r_{\text{FScore}} - r_{\text{degree}}$, remains positive across municipalities but shows no geographic gradient, indicating that the processing-specific component of assortative mixing is broadly shared nationwide (see Figure 3**c**). Finally, holding fixed the nationally validated NOVA 4–4 edges and reweighting them by local mean log-lift, association

strength decreases with spectral distance ($\rho = -0.56$, $p < 0.001$), while remaining higher than that of NOVA 1–1 pairs in all 95 municipalities (see Figure 3**d**). NOVA 1–1 pairs weaken as well ($\rho = -0.27$, $p = 0.007$), and the gap between the two shows no gradient ($\rho = -0.16$, $p = 0.13$).

Thus, across the country, UPF purchasing remains invariant in peripheral network position (AUC < 0.5 in all 95 municipalities), aggregate UPF–UPF volume, and the higher log-lift of NOVA 4–4 over NOVA 1–1 pairs across all 95 municipalities. Under fixed-N^* sampling, processing assortativity r_{FScore} continues to increase with unpadded shared-product spectral distance from København ($\rho = +0.210$, $p = 0.041$; $\rho = +0.38$, $p < 0.001$ on padded spectra), while degree assortativity decreases or flattens ($\rho = -0.312$, $p = 0.002$ unpadded; $\rho = -0.11$, $p = 0.31$ padded). Consequently, excess processing assortativity $r_{\text{FScore}} - r_{\text{degree}}$ increases with distance ($\rho = +0.419$, $p < 0.001$; see Supplementary Materials, Section 2). Repeating the analyses with Frida IDs as nodes in place of retail products recovers the same patterns, with UPF purchase share remaining uncorrelated with spectral distance across the 95 municipalities ($\rho = -0.04$, $p > 0.05$); on local networks for 30 municipalities spanning the same gradient, UPFs remain peripheral (mean AUC $= 0.34$, all < 0.5) while r_{FScore} still increases with spectral distance ($\rho = +0.89$).

Conclusion

As research efforts increasingly focus on epidemiological and policy frameworks for ultra-processed foods, the field remains polarized between defenders of the NOVA classification [28, 31] and critics pointing to taxonomic ambiguities and reporting artifacts in dietary recall [36, 76–78]. Here, we address both taxonomic and observational limitations by pairing continuous processing scores [62] with passive, population-scale transaction data, leveraging 51 million grocery receipts across 95 Danish municipalities to link real-world purchasing behavior to population health outcomes [51, 52]. Rather than treating products in isolation, we use a Bipartite Configuration Model to isolate non-random co-purchasing patterns from popularity-driven co-occurrence [66, 67]. We show that while municipal UPF purchase share spans a narrow band (0.46–0.56) uncorrelated with the urban-to-peripheral spectral axis, co-purchasing networks are structurally reorganized along that gradient, and UPFs sit at the periphery of these networks everywhere. Importantly, the structural cohesion of these local networks predicts municipal cardiometabolic prescription rates, accounting for residual variance beyond standard sociodemographic factors and product prevalence [54]. Together, our results indicate that geographic cardiometabolic variation reflects not only *what* is purchased locally, but also *how* products are combined in shopping baskets [28, 61].

While our passive-data approach offers a scale that surveys cannot match, it entails inherent methodological limitations. First, receipts record what a household buys, not what each person eats: they miss food waste, sharing within the household, and food bought elsewhere – in other supermarket chains, restaurants or canteens [52, 54]. Second, our design is cross-sectional and correlational, so it does not establish a causal link between co-purchasing network structure and cardiometabolic outcomes. Third, although Danish population registries enable individual-level life-course analyses [79], we have not linked purchases and prescriptions at the individual level: because we aggregate by municipality, the households generating local baskets need not

be the same residents redeeming cardiometabolic prescriptions. Unmeasured confounders, such as physical activity, are also absent from transaction records. Nevertheless, the reported associations remain robust to spatial cross-validation, receipt downsampling, age-restricted endpoints, Frida-ID node definitions, and filtering against maximum-entropy bipartite null models [66] (see Supplementary Materials, Sections 2–5).

Moving beyond product-centric nutrition policy requires a complex-systems view of dietary behavior and food environments [57, 61]. Current instruments, such as sugar taxes or front-of-pack warning labels, judge one product at a time [28, 61]. Our results suggest that health tracks how foods are combined, not only which foods are bought – and a rule that looks at single products cannot see those combinations. Retailers could complement item-level rules by acting on the combinations themselves, for example through which products are placed or promoted together, or through price promotions that bundle healthier ingredients [20, 55, 80]. Future work can close the gaps listed above by following co-purchasing networks over time in longitudinal retail panels [52] and by linking household purchases to individual health trajectories in Danish registry settings [54, 79] – testing whether intervening on food combinations, not only on single products, shifts cardiometabolic risk at scale [28, 57, 69].

Methods

Data Descriptions

Transaction records were provided by Salling Group for Netto discount stores in Denmark from 1 February to 20 April 2025 (see Supplementary Materials, Section 1). Each line records an encrypted basket identifier, store code, calendar day, and product metadata. We discarded non-food lines and excluded the Easter weekend (17–20 April 2025), leaving 75 days of data. Edible items were aggregated to brand $\times$ level-3 microcategory nodes (lowest level in the hierarchy), yielding 3,405 products after exclusion. Store coordinates were assigned to municipal boundaries; analyses are restricted to the 95 of Denmark's 98 municipalities with complete coverage (Fanø, Læsø, and Bornholm excluded). To retain population scale while standardizing sampling intensity, we downsample receipts in proportion to municipal population, anchoring the baskets-per-resident ratio at the 25th percentile of coverage; municipalities below this target retain all receipts. We have performed sensitivity analyses varying this coverage ratio (10th–50th percentiles) and the sampling seed, as well as enforcing a fixed absolute sample size ($N^* = 53{,}251$) and an identical basket-size distribution across all municipalities (see Supplementary Materials, Section 2).

Cardiometabolic and sociodemographic indicators for 2024 were retrieved from Statistics Denmark [72]. We use municipal prescription-redemption rates as treated-disease proxies [54]: the share of residents redeeming at least one medicine in ATC classes A10 (antidiabetics; treated diabetes), C (cardiovascular system), and C10 (lipid-lowering agents, a subclass of C). ATC A10 pools insulin (A10A) and non-insulin antidiabetics (A10B), but Statistics Denmark publishes only the pooled class at municipality level. In the same prescription register, at national and regional resolution, 92.8% of people who redeemed an A10 medicine in 2024 also redeemed an A10B agent (range 91.5–93.7% across the five regions; primary sector, number of users) [81]. We

therefore read the municipal A10 rate as treated diabetes that is predominantly type 2. The main analysis uses age-aggregated rates – the number of residents who redeemed at least one relevant prescription, divided by the municipal population – and adjusts for a sociodemographic block S: log total population, the population shares aged < 30 and ≥ 60, and an affluence index contrasting high-income ($\geq$ 700,000 DKK) versus low-income ($<$ 300,000 DKK) households. As a robustness check, we repeat the analysis on rates restricted to residents aged ≥ 60 (redeemers aged ≥ 60 divided by the population aged ≥ 60; Ærø and Langeland lack cardiovascular rates in this stratum and are dropped); because this outcome is already age-stratified, S then drops the age shares and keeps log total population and affluence. Retail purchases cover February–April 2025; health and sociodemographic registries are the most recent complete calendar year (2024), immediately preceding the sales window.

Retail food harmonization and validation

Each retail product was assigned a single reference entry in the DTU Frida food-composition database [71] using exclusively its product name and store category. We first established an automated string-matching baseline that identifies the Frida entry with highest token similarity to the retail title, prioritizing food groups compatible with the store category and searching the unconstrained catalog only when no category rule applied. We then refined assignments through a hierarchical cascade: (i) adopting the product's EFSA FoodEx2 code [82] whenever it directly matched a Frida vocabulary entry; (ii) otherwise retrieving the nearest candidate using a pretrained multilingual text embedding model (multilingual-e5-base) restricted to category-compatible Frida food groups; and (iii) retaining the initial string match only when its token Jaccard similarity exceeded the validation median of 0.15. For embedding-based retrieval, candidate matches were accepted only when cosine similarity met or exceeded the median threshold of 0.825. Residual ambiguities related to preservation or preparation state (e.g., canned tuna in oil versus water, or salted versus raw nuts) were resolved via a compact set of deterministic, name-based domain rules (see Supplementary Materials, Section 5). We validated the mapping against Open Food Facts NOVA classifications for a subset of products across food catalog families (n = 2,944), labeling Frida foods via FoodProX Random Forest estimates [55, 62], under the NOVA framework [22]. The complete cascade – FoodEx2 linkage, category-gated embedding retrieval, filtered string matching, and process-state rules – achieved 68.9% overall accuracy, an ultra-processed (NOVA 4) F_1 score of 0.822, and 78.3% accuracy on the binary unprocessed-versus-ultra-processed contrast (NOVA 1 vs. 4). As an independent benchmark, FoodProX [62] applied directly to product-level Open Food Facts macronutrient panels (without mapping to Frida) recovered NOVA classifications with 82.7% accuracy and a UPF F_1 of 0.908 on items with complete protein–fat–carbohydrate records (n = 2,607). On items with local retailer nutrition panels, FoodProX scores derived from on-pack labels and those derived from mapped Frida profiles agreed on the categorical NOVA class for 76.8% of paired items (UPF F_1 = 0.864; median absolute difference in continuous FScore = 0.08). Route-specific accuracies, error decomposition, and the resulting catalog composition are detailed in the Supplementary Materials, Section 5.

Bipartite Configuration Model and Network Analysis

The Bipartite Configuration Model constructs a maximum-entropy null ensemble conditioned on the observed degree sequences, i.e., the number of unique products purchased in each basket and, conversely, the number of baskets in which each product occurs [66–68]. Within this ensemble, each basket–product pair is assigned a connection probability $p_{sc} = \frac{e^{-\lambda_s-\theta_c}}{1+e^{-\lambda_s-\theta_c}}$, where λ_s and θ_c are the Lagrange multipliers enforcing the observed degree constraints on baskets and products, respectively. For a pair of products a and b, the probability that both occur in basket s is $q_s = p_{sa}p_{sb}$, giving an expected co-occurrence $\mu_{ab} = \langle V_{ab}\rangle = \sum_s q_s$ and variance $\sigma^2_{ab} = \sum_s q_s(1-q_s)$. The observed co-occurrence is standardized relative to this null expectation as $z_{ab} = \frac{V_{ab}-\mu_{ab}}{\sqrt{\sigma^2_{ab}}}$.

We then assessed the statistical significance of each observed co-occurrence using the upper tail of its Poisson–Binomial null distribution [66]. For the exact formulation, the p-value was computed as $p^{\text{exact}}_{ab} = 1 - \sum_{x=0}^{V_{ab}-1} \pi(x \mid a, b)$, where $\pi(x \mid a, b)$ is the Poisson–Binomial probability of observing x common baskets under the BiCM. We additionally computed the Refined Normal Approximation (RNA), which corrects the Gaussian tail probability using the standardized third cumulant of the Poisson–Binomial distribution [83, 84], $\gamma_{ab} = \frac{\sum_s q_s(1-q_s)(1-2q_s)}{[\sum_s q_s(1-q_s)]^{3/2}}$, yielding $p^{\text{RNA}}_{ab} = \bar{\Phi}(z_{ab}) + \frac{\gamma_{ab}}{6}(z^2_{ab}-1)\phi(z_{ab})$, where $\bar{\Phi}$ and ϕ denote the standard normal survival function and density, respectively. Municipal backbones are validated with exact Poisson–Binomial p-values, while the national backbone uses the RNA approximation. Both exact and RNA p-values were subjected to Bonferroni correction for multiple hypothesis testing ($\alpha = 0.001$), retaining only the links whose corrected p-values passed this threshold to indicate co-occurrences significantly stronger than expected from the marginal degree structure alone.

We characterized municipal network structure using the spectrum of the symmetric normalized Laplacian, which captures the organization of connectivity across the network. Municipalities were compared by the Euclidean (L^2) distance between their sorted eigenvalue spectra. When graphs differed in size, the shorter spectrum was padded with 2, the theoretical upper bound of the normalized Laplacian. As a robustness check, we recomputed pairwise spectral distances on the induced subgraphs over the shared product set $V_i \cap V_j$ (median $|V_i \cap V_j| = 780$), with equal Laplacian order and no padding, and recovered the same urban-to-peripheral axis (Spearman $\rho = 0.78$). 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 [75], and Louvain, which greedily maximizes modularity [74] – a measure of the density of edges inside communities versus edges between communities. The two partitions were compared using the Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI), both chance-corrected. ARI measures pairwise co-assignment from the contingency matrix of community labels, whereas AMI measures shared information between the labelings; statistical significance was assessed by permutation of labels. Because Infomap typically yields a finer partition nested within Louvain modules, AMI remains high even when ARI is only moderate, indicating agreement on mesoscale organization. On each municipal backbone, we quantify the tendency of products with similar characteristics to connect by measuring Newman assortativity for the continuous FoodProX *F*Score and for degree, capturing respectively the tendency of foods with similar processing levels and similar connectivity to co-occur in the

network. We further assess whether network connectivity ranks processing extremes using the degree AUC for NOVA 4 versus NOVA 1 products, where an AUC above (below) 0.5 indicates that NOVA 4 products systematically occupy higher-degree (lower-degree) positions than NOVA 1 products. Finally, we assess the significance of the observed *F*Score assortativity against a node-label shuffle null model, obtaining $z > 2$ in all 95 municipalities.

For each municipality with N baskets, we measure pairwise product co-occurrence using log-lift $\log \mathrm{lift}_{ij} = \log\left(\frac{N_{ij}N}{n_i n_j}\right)$, where N_{ij} is the joint basket count and n_i, n_j are marginal counts. Projecting these local co-occurrences onto the national BiCM backbone, we summarize each municipality across the Louvain communities by characterizing the local intra-community edges along three complementary dimensions. We quantify network intensity through the mean $\log \mathrm{lift}_{ij}$ of the observed edges, while product popularity is captured by the mean $\log n_i + \log n_j$, reflecting the baseline marginal frequency of the products involved. Finally, to isolate co-purchasing patterns beyond product popularity, we define residual co-purchase as the municipal residual from regressing the mean $\log P_{ij}$ across community edges on the corresponding mean $\log P_i$ and $\log P_j$ across all municipalities, where $P_{ij} = N_{ij}/N$ and $P_i = n_i/N$. To evaluate how these features predict municipal prescription rates, we fit Ridge regressions on standardized predictors (zero mean, unit variance), selecting the penalty α via generalized cross-validation across a 30-point log-spaced grid (10^{-2} to 10^{3}). We measure predictive performance by leave-one-out cross-validated R^2 for sociodemographic controls alone (S), network intensity alone (I), and the combined model ($S + I$), repeating this for $S + M$ and $S + \widetilde{P}$. The incremental explanatory power of the network beyond sociodemographics is quantified by the share of residual variance after S is $\frac{R^2_{S+I} - R^2_S}{1 - R^2_S}$.

Declarations

- Data Availability: Raw data of purchase are not publicly available due to privacy considerations, but are available to researchers who meet the criteria for access to confidential data, sign a confidentiality agreement and agree to work under supervision in Copenhagen. Municipal health and sociodemographic indicators were retrieved from Statistics Denmark's StatBank (https://www.statbank.dk/) via the open REST API (https://api.statbank.dk/). Prescription-redemption counts come from table MEDI1 (https://www.statbank.dk/MEDI1; year 2024; unique persons redeeming at least one prescription in ATC classes A10, C and C10). The regional split of antidiabetic users into insulin (A10A) and non-insulin (A10B) agents comes from Medstat (https://medstat.dk/; primary sector, number of users, 2024). Resident population and age structure come from table BY2 (https://www.statbank.dk/BY2; 1 January 2024), and household disposable-income brackets from table INDKF132 (https://www.statbank.dk/INDKF132; year 2024). The Danish food-composition database Frida v5.5 is publicly available from Technical University of Denmark (https://frida.fooddata.dk/). Open Food Facts was accessed via its REST API https://openfoodfacts.github.io/openfoodfacts-server/api/. Continuous food-processing scores *F*Score and NOVA class probabilities were computed with FoodProX [62] (https://doi.org/10.1038/s41467-023-37457-1; code at https://github.com/menicgiulia/MLFoodProcessing), retrained on a 52-nutrient USDA panel aligned to Frida v5.5 and applied to each mapped Frida food; a seven-macronutrient FoodProX model was used for product-level Open Food Facts and on-pack information.

- Code Availability: The code to reproduce the analysis is released on FoodHealth. For inquiries, please contact A.D. antde@dtu.dk.
- Acknowledgements A.D. acknowledges helpful discussions and insightful feedback from Riccardo Di Clemente, Sándor Juhász, Anna Mancini, and Fabio Saracco. All authors acknowledge helpful feedback from Ulf Aslak at an early stage of this work. Finally, all authors thank the BI & Data Analytics team (Machine Learning & AI) at Salling Group, with special thanks to Philip Lund Thomsen for initiating and driving this collaboration.
- Author Contributions: L.R. performed preliminary analysis. A.D. 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.
- Funding Declaration: S.L. acknowledges support from the Villum Foundation through the Nation-Scale Social Networks project (grant number 34288). A.D. acknowledges support by COLINE DUT European Partnership project (F-DUT-2023-0037) and the Lagrange Project of the ISI Foundation, funded by Fondazione CRT.

References

[1] Zhou, B., Lu, Y., Hajifathalian, K., Bentham, J., Di Cesare, M., Danaei, G., Bixby, H., Cowan, M., Ali, M., Taddei, C., *et al.*: Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. The Lancet **387**(10027), 1513–1530 (2016) https://doi.org/10.1016/S0140-6736(16)00618-8 . Accessed 2026-08-03

[2] Afshin, A., Sur, P.J., Fay, K.A., Cornaby, L., Ferrara, G., *et al.*: Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet **393**(10184), 1958–1972 (2019) https://doi.org/10.1016/S0140-6736(19)30041-8 . Accessed 2026-08-04

[3] Jayedi, A., Soltani, S., Abdolshahi, A., Shab-Bidar, S.: Healthy and unhealthy dietary patterns and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. British Journal of Nutrition **124**(11), 1133–1144 (2020) https://doi.org/10.1017/S0007114520002330 . Accessed 2026-08-04

[4] O'Hearn, M., Lara-Castor, L., Cudhea, F., Miller, V., Reedy, J., Mozaffarian, D., Global Dietary Database, *et al.*: Incident type 2 diabetes attributable to suboptimal diet in 184 countries. Nature Medicine **29**(4), 982–995 (2023) https://doi.org/10.1038/s41591-023-02278-8 . Accessed 2026-08-03

[5] Micha, R., Wallace, S.K., Mozaffarian, D.: Red and Processed Meat Consumption and Risk of Incident Coronary Heart Disease, Stroke, and Diabetes Mellitus: A Systematic Review and Meta-Analysis. Circulation **121**(21), 2271–2283 (2010) https://doi.org/10.1161/CIRCULATIONAHA.109.924977 . Accessed 2026-08-03

[6] Wang, P., Song, M., Eliassen, A.H., Wang, M., Fung, T.T., Clinton, S.K., Rimm, E.B., Hu, F.B., Willett, W.C., Tabung, F.K., Giovannucci, E.L.: Optimal dietary patterns for prevention of chronic disease. Nature Medicine **29**(3), 719–728 (2023) https://doi.org/10.1038/

s41591-023-02235-5 . Accessed 2026-08-03

[7] Huang, Y., Chen, Z., Chen, B., Li, J., Yuan, X., *et al.*: Dietary sugar consumption and health: umbrella review. BMJ **381**, 071609 (2023) https://doi.org/10.1136/bmj-2022-071609 . Accessed 2026-09-15

[8] Lane, M.M., Gamage, E., Du, S., Ashtree, D.N., McGuinness, A.J., et al.: Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses. BMJ, 077310 (2024) https://doi.org/10.1136/bmj-2023-077310 . Accessed 2026-08-03

[9] Springmann, M., Godfray, H.C.J., Rayner, M., Scarborough, P.: Analysis and valuation of the health and climate change cobenefits of dietary change. Proceedings of the National Academy of Sciences **113**(15), 4146–4151 (2016) https://doi.org/10.1073/pnas.1523119113 . Accessed 2026-08-03

[10] De Souza, R.J., Mente, A., Maroleanu, A., Cozma, A.I., Ha, V., Kishibe, T., Uleryk, E., Budylowski, P., Schünemann, H., Beyene, J., Anand, S.S.: Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. BMJ, 3978 (2015) https://doi.org/10.1136/bmj.h3978 . Accessed 2026-09-15

[11] Tilman, D., Clark, M.: Global diets link environmental sustainability and human health. Nature **515**(7528), 518–522 (2014) https://doi.org/10.1038/nature13959 . Accessed 2026-08-03

[12] Popkin, B.M.: The Nutrition Transition and Obesity in the Developing World. The Journal of Nutrition **131**(3), 871–873 (2001) https://doi.org/10.1093/jn/131.3.871S . Accessed 2026-08-04

[13] Bilal, U., Jones-Smith, J., Diez, J., Lawrence, R.S., Celentano, D.D., Franco, M.: Neighborhood social and economic change and retail food environment change in Madrid (Spain): The heart healthy hoods study. Health & Place **51**, 107–117 (2018) https://doi.org/10.1016/j.healthplace.2018.03.004 . Accessed 2026-09-15

[14] Muncke, J., Touvier, M., Trasande, L., Scheringer, M.: Health impacts of exposure to synthetic chemicals in food. Nature Medicine **31**(5), 1431–1443 (2025) https://doi.org/10.1038/s41591-025-03697-5 . Accessed 2026-08-03

[15] Reardon, T., Timmer, C.P.: The Economics of the Food System Revolution. Annual Review of Resource Economics **4**(1), 225–264 (2012) https://doi.org/10.1146/annurev.resource.050708.144147 . Accessed 2026-08-03

[16] Seto, K.C., Ramankutty, N.: Hidden linkages between urbanization and food systems. Science **352**(6288), 943–945 (2016) https://doi.org/10.1126/science.aaf7439 . Accessed 2026-08-03

[17] Ambikapathi, R., Schneider, K.R., Davis, B., Herrero, M., Winters, P., Fanzo, J.C.: Global food systems transitions have enabled affordable diets but had less favourable outcomes

for nutrition, environmental health, inclusion and equity. Nature Food **3**(9), 764–779 (2022) https://doi.org/10.1038/s43016-022-00588-7 . Accessed 2026-08-03

[18] Di Clemente, R., Strano, E., Batty, M.: Urbanization and economic complexity. Scientific Reports **11**(1), 3952 (2021) https://doi.org/10.1038/s41598-021-83238-5 . Accessed 2026-08-03

[19] Schneider, K.R., Fanzo, J., Haddad, L., Herrero, M., Moncayo, J.R., *et al.*: The state of food systems worldwide in the countdown to 2030. Nature Food **4**(12), 1090–1110 (2023) https://doi.org/10.1038/s43016-023-00885-9 . Accessed 2026-08-03

[20] Scapin, T., Romaniuk, H., Feeley, A., Corrêa, K.P., Kupka, R., Gomez-Donoso, C., Orellana, L., Gupta, A., Sacks, G., Cameron, A.J.: Global food retail environments are increasingly dominated by large chains and linked to the rising prevalence of obesity. Nature Food **6**(3), 283–295 (2025) https://doi.org/10.1038/s43016-025-01134-x . Accessed 2026-08-03

[21] Popkin, B.M., Barquera, S., Corvalan, C., Hofman, K.J., Monteiro, C., Ng, S.W., Swart, E.C., Taillie, L.S.: Towards unified and impactful policies to reduce ultra-processed food consumption and promote healthier eating. The Lancet Diabetes & Endocrinology **9**(7), 462–470 (2021) https://doi.org/10.1016/S2213-8587(21)00078-4 . Accessed 2026-09-15

[22] Monteiro, C.A., Cannon, G., Levy, R.B., Moubarac, J.-C., Louzada, M.L., Rauber, F., Khandpur, N., Cediel, G., Neri, D., Martinez-Steele, E., Baraldi, L.G., Jaime, P.C.: Ultra-processed foods: what they are and how to identify them. Public Health Nutrition **22**(5), 936–941 (2019) https://doi.org/10.1017/S1368980018003762 . Accessed 2026-08-03

[23] Monteiro, C.A., Cannon, G., Moubarac, J.-C., Levy, R.B., Louzada, M.L.C., Jaime, P.C.: The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Public Health Nutrition **21**(1), 5–17 (2018) https://doi.org/10.1017/S1368980017000234 . Accessed 2026-09-15

[24] Scrinis, G., Monteiro, C.: From ultra-processed foods to ultra-processed dietary patterns. Nature Food **3**(9), 671–673 (2022) https://doi.org/10.1038/s43016-022-00599-4 . Accessed 2026-09-14

[25] Vandevijvere, S., Jaacks, L.M., Monteiro, C.A., Moubarac, J., Girling-Butcher, M., Lee, A.C., Pan, A., Bentham, J., Swinburn, B.: Global trends in ultraprocessed food and drink product sales and their association with adult body mass index trajectories. Obesity Reviews **20**(S2), 10–19 (2019) https://doi.org/10.1111/obr.12860 . Accessed 2026-08-03

[26] Coyle, D.H., Huang, L., Shahid, M., Gaines, A., Di Tanna, G.L., Louie, J.C.Y., Pan, X., Marklund, M., Neal, B., Wu, J.H.Y.: Socio-economic difference in purchases of ultra-processed foods in Australia: an analysis of a nationally representative household grocery purchasing panel. International Journal of Behavioral Nutrition and Physical Activity **19**(1), 148 (2022) https://doi.org/10.1186/s12966-022-01389-8 . Accessed 2026-08-04

[27] Martínez Steele, E., Baraldi, L.G., Louzada, M.L.D.C., Moubarac, J.-C., Mozaffarian, D., Monteiro, C.A.: Ultra-processed foods and added sugars in the US diet: evidence from a

nationally representative cross-sectional study. BMJ Open **6**(3), 009892 (2016) https://doi.org/10.1136/bmjopen-2015-009892 . Accessed 2026-08-03

[28] The Lancet: Ultra-processed foods: time to put health before profit. The Lancet **406**(10520), 2601 (2025) https://doi.org/10.1016/S0140-6736(25)02322-0 . Accessed 2026-08-04

[29] Elizabeth, L., Machado, P., Zinöcker, M., Baker, P., Lawrence, M.: Ultra-Processed Foods and Health Outcomes: A Narrative Review. Nutrients **12**(7), 1955 (2020) https://doi.org/10.3390/nu12071955 . Accessed 2026-08-04

[30] Hall, K.D., Ayuketah, A., Brychta, R., Cai, H., Cassimatis, T., *et al.*: Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake. Cell Metabolism **30**(1), 226 (2019) https://doi.org/10.1016/j.cmet.2019.05.020 . Accessed 2026-08-03

[31] Monteiro, C.A., Louzada, M.L., Steele-Martinez, E., Cannon, G., Andrade, G.C., *et al.*: Ultra-processed foods and human health: the main thesis and the evidence. The Lancet **406**(10520), 2667–2684 (2025) https://doi.org/10.1016/S0140-6736(25)01565-X . Accessed 2026-09-14

[32] Bonaccio, M., Di Castelnuovo, A., Costanzo, S., De Curtis, A., Persichillo, M., Sofi, F., Cerletti, C., Donati, M.B., De Gaetano, G., Iacoviello, L.: Ultra-processed food consumption is associated with increased risk of all-cause and cardiovascular mortality in the Moli-sani Study. The American Journal of Clinical Nutrition **113**(2), 446–455 (2021) https://doi.org/10.1093/ajcn/nqaa299 . Accessed 2026-08-03

[33] Cordova, R., Viallon, V., Fontvieille, E., Peruchet-Noray, L., Jansana, A., *et al.*: Consumption of ultra-processed foods and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. The Lancet Regional Health - Europe **35**, 100771 (2023) https://doi.org/10.1016/j.lanepe.2023.100771 . Accessed 2026-09-14

[34] Srour, B., Kordahi, M.C., Bonazzi, E., Deschasaux-Tanguy, M., Touvier, M., Chassaing, B.: Ultra-processed foods and human health: from epidemiological evidence to mechanistic insights. The Lancet Gastroenterology & Hepatology **7**(12), 1128–1140 (2022) https://doi.org/10.1016/S2468-1253(22)00169-8 . Accessed 2026-09-15

[35] Juul, F., Martinez-Steele, E., Parekh, N., Monteiro, C.A.: The role of ultra-processed food in obesity. Nature Reviews Endocrinology **21**(11), 672–685 (2025) https://doi.org/10.1038/s41574-025-01143-7 . Accessed 2026-09-14

[36] O'Neill, H.M.: Ultra-processed foods and cardiometabolic risk: from evidence to policy. npj Metabolic Health and Disease **4**(1), 21 (2026) https://doi.org/10.1038/s44324-026-00116-2 . Accessed 2026-08-04

[37] NCD Risk Factor Collaboration (NCD-RisC), *et al.*: Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature **569**(7755), 260–264 (2019) https://doi.org/10.1038/s41586-019-1171-x . Accessed 2026-08-03

[38] He, P., Liu, Z., Baiocchi, G., Guan, D., Bai, Y., Hubacek, K.: Health–environment efficiency

of diets shows nonlinear trends over 1990–2011. Nature Food **5**(2), 116–124 (2024) https://doi.org/10.1038/s43016-024-00924-z . Accessed 2026-08-04

[39] Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., Schnell, M.: Food Deserts and the Causes of Nutritional Inequality. Technical Report w24094, National Bureau of Economic Research, Cambridge, MA (December 2017). https://doi.org/10.3386/w24094 . http://www.nber.org/papers/w24094.pdf Accessed 2026-08-03

[40] Zheng, L., Adalibieke, W., Zhou, F., He, P., Chen, Y., *et al.*: Health burden from food systems is highly unequal across income groups. Nature Food **5**(3), 251–261 (2024) https://doi.org/10.1038/s43016-024-00946-7 . Accessed 2026-08-09

[41] Cummins, S., Macintyre, S.: Food environments and obesity—neighbourhood or nation? International Journal of Epidemiology **35**(1), 100–104 (2006) https://doi.org/10.1093/ije/dyi276 . Accessed 2026-09-15

[42] Franco, M., Diez-Roux, A.V., Nettleton, J.A., Lazo, M., Brancati, F., Caballero, B., Glass, T., Moore, L.V.: Availability of healthy foods and dietary patterns: the Multi-Ethnic Study of Atherosclerosis. The American Journal of Clinical Nutrition **89**(3), 897–904 (2009) https://doi.org/10.3945/ajcn.2008.26434 . Accessed 2026-09-15

[43] Moore, L.V., Diez Roux, A.V., Nettleton, J.A., Jacobs, D.R., Franco, M.: Fast-Food Consumption, Diet Quality, and Neighborhood Exposure to Fast Food: The Multi-Ethnic Study of Atherosclerosis. American Journal of Epidemiology **170**(1), 29–36 (2009) https://doi.org/10.1093/aje/kwp090 . Accessed 2026-09-15

[44] Xu, R., Huang, X., Zhang, K., Lyu, W., Ghosh, D., Li, Z., Chen, X.: Integrating human activity into food environments can better predict cardiometabolic diseases in the United States. Nature Communications **14**(1), 7326 (2023) https://doi.org/10.1038/s41467-023-42667-8 . Accessed 2026-08-03

[45] Althoff, T., Nilforoshan, H., Hua, J., Leskovec, J.: Large-scale diet tracking data reveal disparate associations between food environment and diet. Nature Communications **13**(1), 267 (2022) https://doi.org/10.1038/s41467-021-27522-y . Accessed 2026-08-03

[46] Desiderio, A., Galdeman, A., Bäuerlein, F., Lehmann, S.: Mapping regional disparities in discounted grocery products. npj Science of Food **10**(1), 112 (2026) https://doi.org/10.1038/s41538-026-00764-0 . Accessed 2026-09-30

[47] Fiolet, T., Srour, B., Sellem, L., Kesse-Guyot, E., Allès, B., Méjean, C., Deschasaux, M., Fassier, P., Latino-Martel, P., Beslay, M., Hercberg, S., Lavalette, C., Monteiro, C.A., Julia, C., Touvier, M.: Consumption of ultra-processed foods and cancer risk: results from NutriNet-Santé prospective cohort. BMJ **360**, 322 (2018) https://doi.org/10.1136/bmj.k322 . Accessed 2026-08-03

[48] Hu, F.B.: Dietary pattern analysis: a new direction in nutritional epidemiology:. Current Opinion in Lipidology **13**(1), 3–9 (2002) https://doi.org/10.1097/00041433-200202000-00002 . Accessed 2026-10-07

[49] Iqbal, K., Buijsse, B., Wirth, J., Schulze, M.B., Floegel, A., Boeing, H.: Gaussian Graphical Models Identify Networks of Dietary Intake in a German Adult Population. The Journal of Nutrition **146**(3), 646–652 (2016) https://doi.org/10.3945/jn.115.221135 . Accessed 2026-10-07

[50] Barabási, A.-L., Menichetti, G., Loscalzo, J.: The unmapped chemical complexity of our diet. Nature Food **1**(1), 33–37 (2019) https://doi.org/10.1038/s43016-019-0005-1 . Accessed 2026-08-03

[51] Bandy, L., Adhikari, V., Jebb, S., Rayner, M.: The use of commercial food purchase data for public health nutrition research: A systematic review. PLOS ONE **14**(1), 0210192 (2019) https://doi.org/10.1371/journal.pone.0210192 . Accessed 2026-08-03

[52] Nevalainen, J., Erkkola, M., Saarijärvi, H., Näppilä, T., Fogelholm, M.: Large-scale loyalty card data in health research. DIGITAL HEALTH **4**, 2055207618816898 (2018) https://doi.org/10.1177/2055207618816898 . Accessed 2026-08-03

[53] Pennacchioli, D., Coscia, M., Rinzivillo, S., Giannotti, F., Pedreschi, D.: The retail market as a complex system. EPJ Data Science **3**(1), 33 (2014) https://doi.org/10.1140/epjds/s13688-014-0033-x . Accessed 2026-08-03

[54] Aiello, L.M., Schifanella, R., Quercia, D., Del Prete, L.: Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Science **8**(1), 14 (2019) https://doi.org/10.1140/epjds/s13688-019-0191-y . Accessed 2026-08-03

[55] Ravandi, B., Ispirova, G., Sebek, M., Mehler, P., Barabási, A.-L., Menichetti, G.: Prevalence of processed foods in major US grocery stores. Nature Food **6**(3), 296–308 (2025) https://doi.org/10.1038/s43016-024-01095-7 . Accessed 2026-08-03

[56] Fox, E.J., Montgomery, A.L., Lodish, L.M.: Consumer Shopping and Spending across Retail Formats. The Journal of Business **77**(S2), 25–60 (2004) https://doi.org/10.1086/381518 . Accessed 2026-08-03

[57] Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, D.T., Greaves, F., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuess, E., Shiell, A., Thomas, J., White, M.: The need for a complex systems model of evidence for public health. The Lancet **390**(10112), 2602–2604 (2017) https://doi.org/10.1016/S0140-6736(17)31267-9 . Accessed 2026-09-15

[58] Ni Mhurchu, C., Vandevijvere, S., Waterlander, W., Thornton, L.E., Kelly, B., Cameron, A.J., Snowdon, W., Swinburn, B., INFORMAS: Monitoring the availability of healthy and unhealthy foods and non-alcoholic beverages in community and consumer retail food environments globally. Obesity Reviews **14**(S1), 108–119 (2013) https://doi.org/10.1111/obr.12080 . Accessed 2026-08-03

[59] Kuhl, E.: AI for food: accelerating and democratizing discovery and innovation. npj Science of Food **9**(1), 82 (2025) https://doi.org/10.1038/s41538-025-00441-8 . Accessed 2026-08-03

[60] Martin, S.L., Howell, T., Duan, Y., Walters, M.: The feasibility and utility of grocery receipt

analyses for dietary assessment. Nutrition Journal **5**(1), 10 (2006) https://doi.org/10.1186/1475-2891-5-10 . Accessed 2026-08-03

[61] Mozaffarian, D., Rosenberg, I., Uauy, R.: History of modern nutrition science—implications for current research, dietary guidelines, and food policy. BMJ, 2392 (2018) https://doi.org/10.1136/bmj.k2392 . Accessed 2026-08-03

[62] Menichetti, G., Ravandi, B., Mozaffarian, D., Barabási, A.-L.: Machine learning prediction of the degree of food processing. Nature Communications **14**(1), 2312 (2023) https://doi.org/10.1038/s41467-023-37457-1 . Accessed 2026-08-03

[63] Johansen, D., Friis, K., Skovenborg, E., Grønbæk, M.: Food buying habits of people who buy wine or beer: cross sectional study. BMJ **332**(7540), 519–522 (2006) https://doi.org/10.1136/bmj.38694.568981.80 . Accessed 2026-08-03

[64] Dong, X., Suhara, Y., Bozkaya, B., Singh, V.K., Lepri, B., Pentland, A.: Social Bridges in Urban Purchase Behavior. ACM Transactions on Intelligent Systems and Technology **9**(3), 1–29 (2018) https://doi.org/10.1145/3149409 . Accessed 2026-08-03

[65] Dong, L., Ratti, C., Zheng, S.: Predicting neighborhoods' socioeconomic attributes using restaurant data. Proceedings of the National Academy of Sciences **116**(31), 15447–15452 (2019) https://doi.org/10.1073/pnas.1903064116 . Accessed 2026-08-03

[66] Saracco, F., Straka, M.J., Clemente, R.D., Gabrielli, A., Caldarelli, G., Squartini, T.: Inferring monopartite projections of bipartite networks: an entropy-based approach. New Journal of Physics **19**(5), 053022 (2017) https://doi.org/10.1088/1367-2630/aa6b38 . Accessed 2026-08-03

[67] Saracco, F., Di Clemente, R., Gabrielli, A., Squartini, T.: Randomizing bipartite networks: the case of the World Trade Web. Scientific Reports **5**(1), 10595 (2015) https://doi.org/10.1038/srep10595 . Accessed 2026-08-03

[68] Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., Caldarelli, G.: The statistical physics of real-world networks. Nature Reviews Physics **1**(1), 58–71 (2019) https://doi.org/10.1038/s42254-018-0002-6 . Accessed 2026-08-03

[69] Hunter, R.F., De La Haye, K., Badham, J., Valente, T., Clarke, M., Kee, F.: Social network interventions for health behaviour change: a systematic review. The Lancet **390**, 47 (2017) https://doi.org/10.1016/S0140-6736(17)32982-3 . Accessed 2026-08-03

[70] Møller, L.K.: Danskernes aftensmad i 2024 beskrives som velsmagende, hjemmelavet og mættende. Markedsanalyse / ernæring & forbrugeranalyse, Landbrug & Fødevarer (February 2025). Survey of Danish evening-meal habits in 2024 (Voxmeter; four waves). Citekey year suffix refers to the survey year. https://lf.dk/media/xjqmowlp/markedsanalyse_danskernes-aftensmad-i-2024_5022025.pdf Accessed 2026-09-15

[71] Poulsen, A., Langwagen, M., Jacobsen, J., Christensen, T.: Dataset from The Danish Food Composition Database, Frida version 5.5. Technical University of Denmark. Artwork Size: 28769892 Bytes Pages: 28769892 Bytes (2026). https://doi.org/10.11583/DTU.29500682 . https://data.dtu.dk/articles/dataset/Dataset_from_The_Danish_Food_Composition_Database_Frida_version_4_1/29500682 Accessed 2026-08-29

[72] Statistics Denmark: StatBank Denmark. Statistics Denmark (2026). https://www.statbank.dk/

[73] Tufano, M., Duarte, F., Mazzarello, M., Eshtiyagh, J., Ratti, C., Camps, G.: Data-driven nutritional assessment of urban food landscapes: insights from Boston, London, and Dubai. Scientific Reports **15**(1), 24453 (2025) https://doi.org/10.1038/s41598-025-08098-9 . Accessed 2026-08-03

[74] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), 10008 (2008) https://doi.org/10.1088/1742-5468/2008/10/P10008 . Accessed 2026-08-03

[75] Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences **105**(4), 1118–1123 (2008) https://doi.org/10.1073/pnas.0706851105 . Accessed 2026-08-03

[76] Lichtenstein, A.H., Khera, A., Anderson, C.A.M., Appel, L.J., DeSilva, D.M., Gardner, C., Hu, F.B., Jones, D.W., Petersen, K.S., on behalf of the American Heart Association: 2026 Dietary Guidance to Improve Cardiovascular Health: A Scientific Statement From the American Heart Association. Circulation **153**(18) (2026) https://doi.org/10.1161/CIR.0000000000001435 . Accessed 2026-09-15

[77] Khandpur, N., Geleijnse, J.M., Kamphuis, C.B.M., Velema, E., Poelman, M.P., Eshuis, R., Djojosoeparto, S.K., Verweij, M.F., Moran, A.J.: The validity and utility of Nova and the concept of ultra-processed foods for science and policy. Nature Food **7**(6), 513–520 (2026) https://doi.org/10.1038/s43016-026-01371-8 . Accessed 2026-08-04

[78] Louie, J.C.Y.: Ultraprocessed food research and scientific discourse. Nature Food **7**(6), 521–526 (2026) https://doi.org/10.1038/s43016-026-01372-7 . Accessed 2026-08-04

[79] Savcisens, G., Eliassi-Rad, T., Hansen, L.K., Mortensen, L.H., Lilleholt, L., Rogers, A., Zettler, I., Lehmann, S.: Using sequences of life-events to predict human lives. Nature Computational Science **4**(1), 43–56 (2023) https://doi.org/10.1038/s43588-023-00573-5 . Accessed 2026-09-14

[80] Taillie, L.S., Ng, S.W., Xue, Y., Harding, M.: Deal or no deal? The prevalence and nutritional quality of price promotions among U.S. food and beverage purchases. Appetite **117**, 365–372 (2017) https://doi.org/10.1016/j.appet.2017.07.006 . Accessed 2026-08-03

[81] Danish Health Data Authority: Medstat.dk. Danish Health Data Authority. Register of Medicinal Product Statistics. Primary sector; number of users; ATC A10, A10A and A10B; 2024; Denmark and the five regions (2026). https://medstat.dk/

[82] European Food Safety Authority (EFSA): The food classification and description system FoodEx 2 (revision 2). EFSA Supporting Publications **12**(5) (2015) https://doi.org/10.2903/sp.efsa.2015.EN-804 . Accessed 2026-08-31

[83] Lugannani, R., Rice, S.: Saddle point approximation for the distribution of the sum of independent random variables. Advances in Applied Probability **12**(2), 475–490 (1980) https://doi.org/10.2307/1426607 . Accessed 2026-09-05

[84] Hong, Y.: On computing the distribution function for the Poisson binomial distribution. Computational Statistics & Data Analysis **59**, 41–51 (2013) https://doi.org/10.1016/j.csda.2012.10.006 . Accessed 2026-09-05

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterials.pdf