What Industries Cluster Around Hyperscale Data Centers?

Dany Bahar   Greg C. Wright

May 13, 2026

Abstract

The main paper documents that hyperscale data center entry raises information-sector employment by 22 percent in the host county. This note opens the black box. A county-level NAICS subsector decomposition shows that the gain is concentrated in NAICS 519 (other information services), with a smaller positive contribution from NAICS 517 (telecommunications); NAICS 518 (data processing and hosting) responds at the county level but reflects the operator's own on-site staffing, not downstream agglomeration. Switching to sub-county evidence from LinkUp job postings (2007–2026) and a curated company–NAICS crosswalk, we estimate within-site ring difference-in-differences and find a staggered distance gradient: NAICS 518 colocation peers cluster within 5 km of hyperscale builds (+184%), streaming and media firms at 5–10 km (+133%), and web-search and information firms at 10–25 km (+59%). The mechanism is consumption-side agglomeration: hyperscale builds anchor a tiered cluster of internet-economy firms, with the tightest co-location for the firms most dependent on physical interconnect.

1 Motivation

The main paper finds that data center entry raises information-sector employment by 22 percent in the host county at $t = 6$. The gain is concentrated entirely in counties that host a hyperscale operator. Colocation entry, despite delivering a similar capital shock, produces no information-sector gain. We argued there that the operator-type contrast was consistent with hyperscale operators procuring fiber, power, and managed services from a local supplier network that colocation operators do not bring with them. That argument identified a candidate mechanism but did not test it directly.

This note tests the mechanism by decomposing the headline NAICS 51 (information sector) effect into its component three-digit subsectors. The QCEW reports county-level employment at the three- and four-digit NAICS level via the BLS annual singlefile archive going back to 2003. We pull NAICS 511 (publishing), 5112 (software publishers), 512 (motion picture and video), 515 (broadcasting), 517 (telecommunications), 518 (data processing and hosting), and 519 (other information services) for the same county panel as the main paper, then re-estimate the synthetic control specification on each subsector separately.

The decomposition speaks directly to two mechanism stories. First, if the hyperscale info-sector effect were driven by hyperscale operators themselves hiring at scale on the campus, the effect should be concentrated in NAICS 518 (data processing and hosting), the code under which AWS, Google, Microsoft, and Meta classify their data center establishments. Second, if the effect were driven by a localized supplier network of fiber installers, NOC contractors, and managed services firms, the effect should be concentrated in NAICS 517 (telecommunications) and possibly the construction-adjacent subsectors. As we show below, neither of these is the dominant story. The dominant channel is NAICS 519 (other information services), which during 2003–2024 catches internet publishing, web search portals, and internet-based information services more generally. The mechanism is on the consumption side: internet-economy firms cluster around hyperscale data center campuses to be close to the compute and bandwidth they consume.

2 Empirical strategy

We use the same county-year panel as the main paper, restricted to counties that have either no large data center over 2003–2024 (the never-treated donor pool) or whose first large data center opens during 2008–2024 (the treated set). Treatment, hyperscale/colocation classification, and panel construction follow the main paper exactly.

For each NAICS subsector s, we re-estimate the synthetic control design used for the main paper's headline NAICS 51 result. For each treated county, we form a synthetic counterfactual as a non-negative weighted average of never-treated counties, where the weights minimize the squared distance between the treated county and its synthetic counterpart on the pre-treatment outcome trajectory in subsector s. We restrict the donor pool for each treated county to the 300 never-treated counties whose pre-period mean outcome is closest to the treated unit, which screens out clearly incomparable donors and yields well-conditioned matches without changing the estimand. The reported treatment effect is the average across treated counties of $\log(1 + \text{emp}^{\text{treated}}_{s,t=6}) - \log(1 + \text{emp}^{\text{synthetic}}_{s,t=6})$, expressed as a percent change.

We do this on three subsamples: (i) all treated counties pooled; (ii) hyperscale-host treated counties (those whose first DC opening involves at least one hyperscale operator); (iii) colocation-host treated counties (those whose first DC is colocation only). The hyperscale-host versus colocation-host comparison is the same one that drives the headline result in the main paper.

3 Results

3.1 The headline NAICS 51 result replicates

Table 1 reports the SCM estimates by subsector and subsample. The first row replicates the main paper's headline finding. Pooling all treated counties, the NAICS 51 (information sector) effect at $t = 6$ is 28 percent. The hyperscale-host effect is 64 percent. The colocation-host effect is essentially zero (−2.7 percent). The qualitative pattern is the same as in the main paper: hyperscale entry generates a large information-sector ramp; colocation entry does not.

Table 1: Synthetic control estimates of post-DC effect on NAICS 51 and its subsectors, ATT at $t = 6$

	All treated	Hyperscale-host	Colocation-host
NAICS 51 (information, all)	28.3%	64.2%	−2.7%
NAICS 511 (publishing)	−6.1%	−24.4%	27.4%
NAICS 5112 (software publishers)	7.5%	3.6%	18.1%
NAICS 512 (motion picture)	−5.7%	−20.2%	15.4%
NAICS 515 (broadcasting)	−11.3%	−7.4%	−26.0%
NAICS 517 (telecommunications)	40.1%	45.9%	36.3%
NAICS 518 (data processing, hosting)	64.0%	67.8%	95.1%
NAICS 519 (other information)	28.9%	48.7%	−3.9%
Treated counties (N)	90	44	22

Notes: Each cell is the average treatment effect at $t = 6$ from a separate synthetic control regression on the indicated subsector. Treated counties are matched to a non-negative weighted average of never-treated counties on the pre-period outcome. Donor pool screened to the 300 never-treated counties with the closest pre-period mean outcome. ATT expressed as a percent change in employment relative to the synthetic counterfactual; the underlying coefficient is in log points. NAICS 5112 is a four-digit subsector within NAICS 511 (software publishers). NAICS 511, 515, and 5112 cover 2003–2021 only because the 2022 NAICS revision reorganized those codes. The main paper uses the NAICS 51 aggregate from the same QCEW source.

3.2 The decomposition: NAICS 519 is the dominant channel

The remaining rows decompose this aggregate. Three patterns stand out.

First, NAICS 519 (other information services) is the largest contributor to the hyperscale-versus-colocation gap. Hyperscale-host counties show a 49 percent increase; colocation-host counties show a 4 percent decline. The 53 percentage-point gap on this single subsector alone explains most of the 67 percentage-point hyperscale-versus-colocation gap on the NAICS 51 aggregate. NAICS 519 in our sample period predominantly captures internet publishing, web search portals, and broadly internet-based information services. Until the 2017 NAICS revision, code 519130 (Internet Publishing and Broadcasting and Web Search Portals) classified the bulk of major internet platforms by their establishment of activity, and that code lived in 519. We return to the implications of the 2017 reclassification below.

Second, NAICS 517 (telecommunications) shows large positive effects in both groups: 46 percent in hyperscale-host counties and 36 percent in colocation-host counties. This is the channel that fits the original supplier-network mechanism story most closely, capturing fiber installers, last-mile telecommunications providers, and dedicated carriers serving the data center campuses. But it is similarly sized in both kinds of treated metros, so it does not contribute to the hyperscale-versus-colocation gap. It does contribute meaningfully to the overall information-sector effect: a 40 percent gain on telecommunications employment is large by historical standards and is consistent with the engineering reality that hyperscale and colocation campuses both require dedicated fiber and edge infrastructure.

Third, NAICS 518 (data processing and hosting), which captures the data center operators themselves, is *larger* in colocation-host counties (95 percent) than in hyperscale-host counties (68 percent). On-site staffing of the data center campus does respond to entry, but the effect is substantially bigger for colocation operators than hyperscale operators. This is consistent with the engineering reality that colocation operators (Equinix, Digital Realty, CyrusOne) staff their facilities more heavily on-site than hyperscale operators, who typically run lights-out facilities with leaner local headcount per facility. The on-site operator response cannot be the source of the hyperscale-versus-colocation gap on the NAICS 51 aggregate, because the gap on this subsector runs in the opposite direction.

The remaining subsectors are smaller and contribute roughly offsetting amounts. NAICS 511 (publishing) is −24 percent in hyperscale-host counties and +27 percent in colocation-host counties. NAICS 512 (motion picture) shows a similar pattern. NAICS 515 (broadcasting) is moderately negative in both groups. NAICS 5112 (software publishers, a four-digit subset of 511) is small and roughly equal across groups. The cross-subsector heterogeneity reflects the distinct industrial composition of the two operator types' host counties: hyperscale-host counties skew toward smaller, tech-adjacent metros where legacy publishing and broadcasting were thin to begin with; colocation-host counties are mostly large established metros where those legacy media subsectors have their own time trends.

3.3 Reclassification check

The headline NAICS 519 finding raises an immediate concern. Internet publishing has moved across NAICS revisions: it was placed in 519130 in the 2007 NAICS, retained there in 2012 and 2017 (with minor adjustments), and then reorganized in the 2022 NAICS revision into a new sector 516 (Broadcasting and Content Providers). If the QCEW is mechanically reclassifying establishments rather than reflecting real entry and exit, the post-DC ramp in NAICS 519 could be a measurement artifact concentrated at the 2017 or 2022 revision boundaries.

Table 2 restricts the SCM sample to four windows that isolate the relevant NAICS regimes. The full-sample result is 49 percent in hyperscale-host counties and −4 percent in colocation-host counties. Restricting to the pre-2017 NAICS regime (2003–2016), when internet publishing was unambiguously in 519130, the hyperscale-host effect grows to 123 percent and the colocation-host effect deepens to −61 percent. The hyperscale-versus-colocation gap is roughly 184 percentage points in this cleanest window, much larger than the 53 percentage-point gap from the full sample. Restricting to the pre-2022 window (2003–2021) gives an intermediate result, consistent with the 2022 reorganization gradually pulling internet-publishing employment out of NAICS 519 over time. The post-2017 window is small (eleven hyperscale and four colocation treated counties with sufficient post-period coverage) and is dominated by a different cohort of treated counties opened in 2017 and later, which complicates inference.

Table 2: NAICS 519 ATT at $t = 6$ across NAICS-revision sample windows

Sample window	Hyperscale-host	Colocation-host	N (h / c)
Full sample (2003–2024)	+48.7%	−3.9%	37 / 20
Pre-2017 revision (2003–2016)	+122.9%	−60.7%	18 / 10
Pre-2022 revision (2003–2021)	+73.0%	−14.3%	31 / 17
Post-2017 (2017–2024, t=4)	−61.4%	+151.9%	11 / 4

Notes: Each cell is an ATT at $t = 6$ from a separate synthetic control restricted to the indicated calendar window. The pre-2017 window covers the regime in which Internet Publishing and Broadcasting and Web Search Portals (NAICS 519130) was unambiguously inside three-digit NAICS 519. The 2017 NAICS revision retained the code but began moving establishments out; the 2022 revision reorganized the bulk of internet-publishing activity into a new sector 516. The post-2017 window contains only treated counties that opened a first DC after 2016 and is dominated by a small later-cohort sample with limited post-period runway (we report ATT at $t = 4$ in that window).

The pattern is informative in two ways. First, the NAICS 519 finding is not a 2017 reclassification artifact: the effect is present in the cleanest pre-2017 window, where the code definition is most stable, and is in fact substantially larger there. Second, the gradual attenuation of the effect as the sample extends through the 2017 and 2022 revisions is what we would expect if real economic activity is being rebadged into other codes (especially the new NAICS 516) rather than disappearing.

3.4 Event-time profiles

Figure 1 plots the average treated-versus-synthetic gap in log employment over event time for the headline NAICS 51 outcome and the two largest contributing subsectors (NAICS 517 and NAICS 518). The pre-period gap is close to zero in each panel, consistent with the SCM matching pre-treatment trajectories well. The post-period ramps are gradual and rise through $t = 6$, with the steepest profile in the hyperscale column for NAICS 51 and the most pronounced operator-type contrast on the same outcome.

3.5 Cross-checking with TWFE event study

Figure 2 repeats the decomposition using a two-way fixed-effects event study with metro-clustered standard errors, as a robustness check. The pattern is broadly consistent with the SCM. Pre-trends are mostly small and not jointly significant. NAICS 519 has a non-trivial pre-trend in both subsamples under TWFE; the SCM design absorbs this by matching on pre-period levels, which is why the SCM estimates the underlying treatment effect more cleanly here.

4 Sub-county evidence from job postings

The county-level SCM identifies the subsector that absorbs the headline employment response, but it cannot say whether the responding firms are physically next door to the data center campus or thirty miles away in the same county. For the agglomeration interpretation to be sharp, we need finer spatial resolution. We bring that resolution by switching to job-postings microdata.

4.1 Data

We use the LinkUp archive of US online job postings for 2007–2026, which provides company name, ZIP code, posting date, and removal date for every posting it indexes. The archive contains 226 million USA postings during our window, of which 2.0 million are

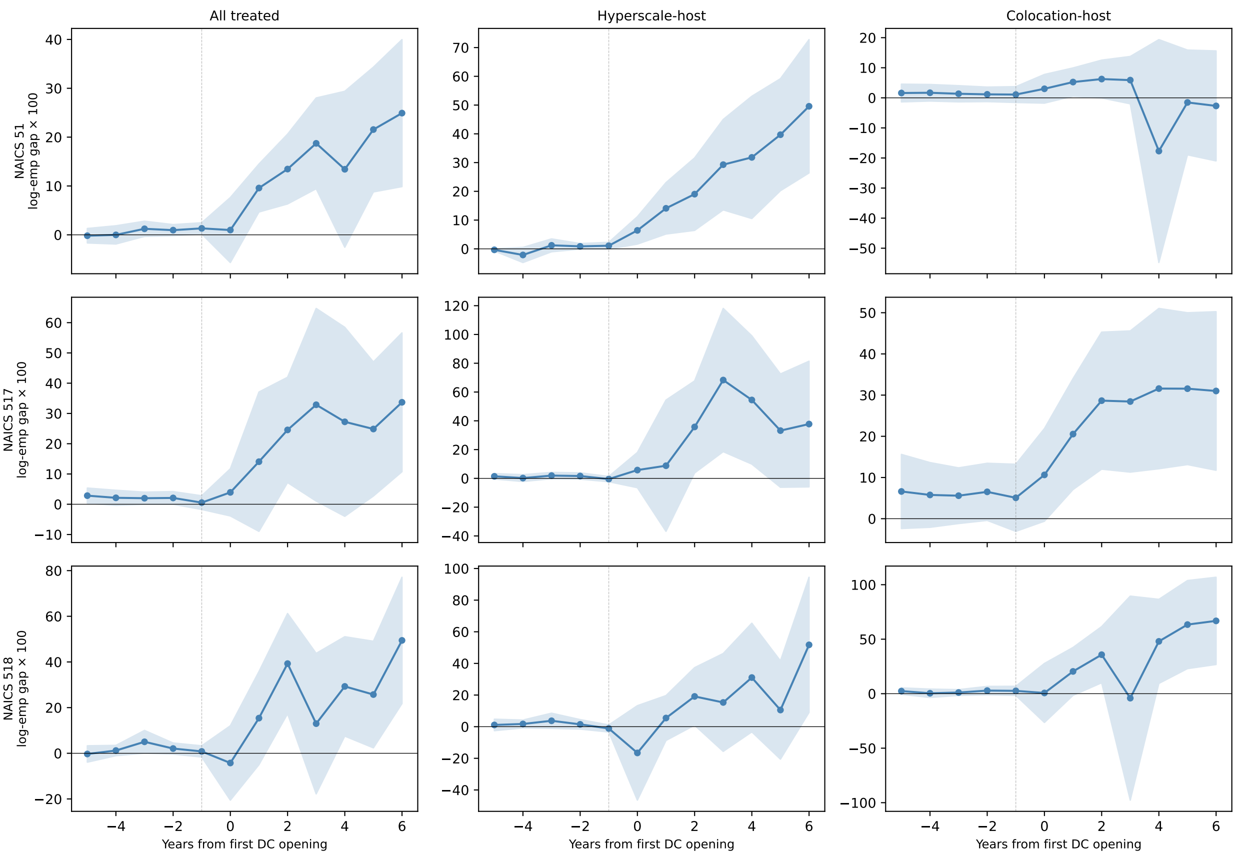

Figure 1: Event-time profiles of the SCM treated-minus-synthetic gap, by NAICS subsector and subsample

Notes: Each panel plots the average treated-minus-synthetic gap across treated counties for a single outcome–subsample combination. Bands are 95 percent confidence intervals based on cross-county standard errors. $t = 0$ is the calendar year of the first DC opening. Vertical reference at $t = -1$.

from the six hyperscale parent corporations directly. We aggregate to a ZIP $\times$ month panel with two outcome counts: postings created in the month (new) and postings alive at any point in the month (active). For each posting we tag the issuing company against a curated COMPANY_NAME-to-NAICS crosswalk we built by classifying the top 500 US LinkUp employers (51.7 percent of all postings) by hand. The classification distinguishes NAICS 5112 (software publishers), 512/516 (motion picture, sound recording, and streaming providers), 515 (broadcasting), 517 (telecommunications), 518 (computing infrastructure, data processing, web hosting), and 519 (web search portals and other information services), plus a non-NAICS-51 control group of 5415 (computer-systems-design consulting) and 5417 (pharmaceutical R&D). Each ZIP is assigned to its nearest hyperscale data center via 2020 ZCTA Gazetteer centroids; that distance then maps the ZIP into a distance ring around the DC.

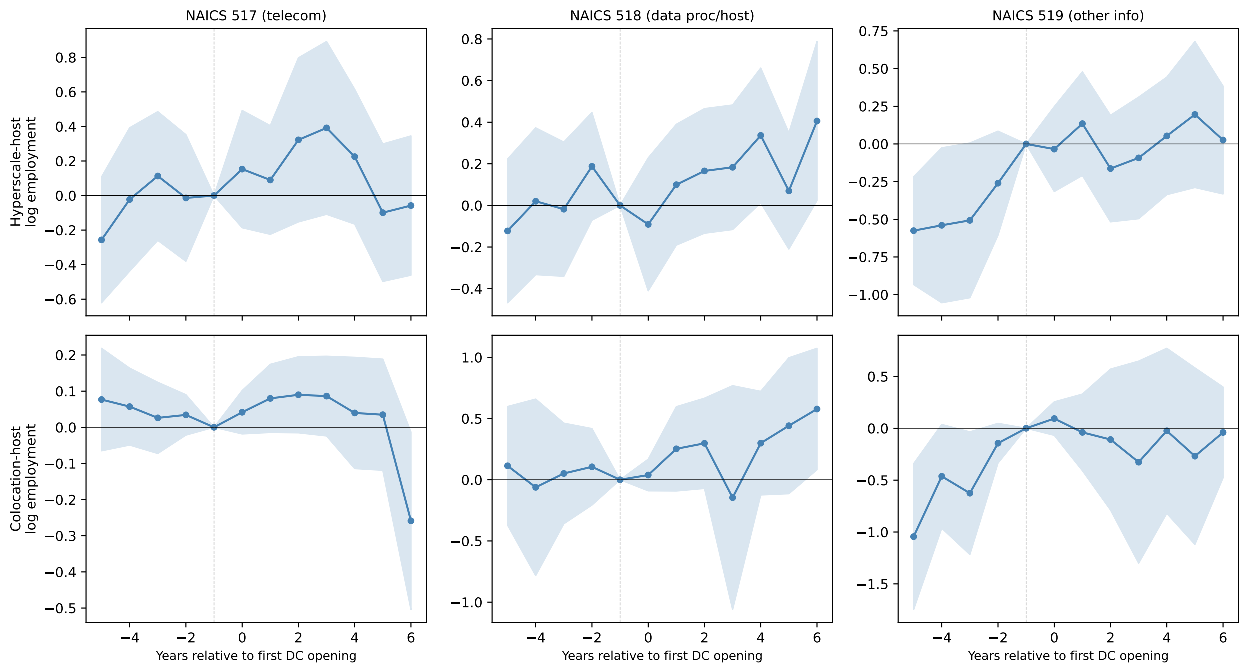

Figure 2: TWFE event-study estimates by NAICS subsector, hyperscale-host vs colocation-host

Notes: Each panel is a separate TWFE regression of $\log(1 + \text{emp}_s)$ on event-time dummies anchored at $t = -1$, with county and year fixed effects and metro-clustered standard errors. Bands are 95 percent confidence intervals.

4.2 Spec

We restrict the sample to 99 hyperscale clean-confidence data centers opening 2014–2024. For each inner ring $r \in \{0\text{–}5 \text{ km}, 5\text{–}10 \text{ km}, 10\text{–}25 \text{ km}\}$ separately, we estimate

$$\log(1 + \text{postings}_{i,r',t}) = \beta_r(\text{post}_{i,t} \cdot \mathbb{1}[r' = r]) + \gamma\, \text{post}_{i,t} + \alpha_{i,r'} + \lambda_t + \varepsilon_{i,r',t}$$

where i indexes the hyperscale data center, $r' \in \{r,\ 25\text{–}50 \text{ km outer}\}$, t is calendar month, and $\text{post}_{i,t}$ is an indicator for the months on or after data center i's opening. We absorb DC-by-ring fixed effects $\alpha_{i,r'}$ and calendar-month fixed effects λ_t, and cluster standard errors at the data-center level. The event window is ± 48 months. We run the specification separately for each NAICS subgroup, restricting to non-hyperscale postings so that hyperscale operators' own employment does not contaminate the estimated co-location effect.

4.3 Results: a staggered downstream-customer gradient

Figure 3 shows the cross-sectional density gradient. Panel (a) plots mean active postings per ZIP-month against the ZIP's distance to its nearest hyperscale DC. Panel (b) normalizes each subgroup to its own 50–100 km baseline, so the gradient is visible regardless of how large the subgroup is in absolute terms. NAICS 518 firms are 3.9 times more dense at 2–5 km than at 50–100 km. NAICS 5112 software publishers are 2.9 times denser, and NAICS 519 web-search firms are 2.0 times denser, both at 2–5 km. NAICS 5415 tech consulting is 3.3 times denser inside 1 km of the DC, consistent with on-site contractor and integrator presence. NAICS 5417 pharma R&D, which has no functional relationship to data-center workload, is flat across the gradient, as expected for a placebo.

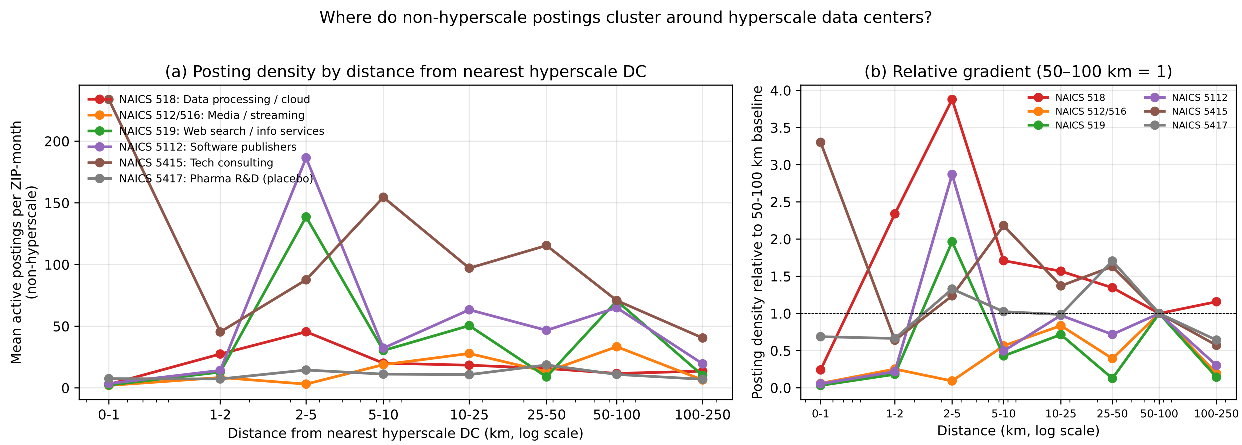

Figure 3: Where non-hyperscale postings cluster around hyperscale data centers, cross-section over 2007–2026

Notes: Panel (a) plots mean active postings per ZIP-month over 2007–2026, by NAICS subgroup, against the ZIP's distance to its nearest hyperscale data center. Panel (b) normalizes each subgroup to its 50–100 km baseline. Hyperscale-operator postings are excluded so that the figure reflects downstream firms only. Distance bins on log scale.

Table 3 reports the within-site ring DiD estimates by inner ring and NAICS subgroup. The estimates align with the cross-section but layer on a clean before-versus-after interpretation. Three findings stand out.

First, NAICS 518 colocation peers (Equinix, Digital Realty, CoreSite, CyrusOne, and other non-hyperscale data-processing firms) rise by 184 percent in the 0–5 km ring after a hyperscale opens ($p = 0.005$, $n = 17$ DCs), and by 62 percent in the 5–10 km ring ($p = 0.069$). This is the cleanest evidence of downstream-industry colocation: the firms that physically locate immediately adjacent to hyperscale builds are themselves data-center and cloud-infrastructure firms. The mechanism is cross-connection and shared utility infrastructure; the result is consistent with the engineering reality that colocation operators want to be one fiber-pull away from a hyperscale region.

Second, NAICS 512/516 streaming and media firms (Netflix, Disney, Spotify, Roku, and similar content providers) rise by 133 percent in the 5–10 km ring ($p = 0.02$, $n = 11$ DCs).

NAICS 519 web-search and information firms (Google search-side activity, Meta Platforms, Snap, Pinterest, Yelp, and the various information-services and online-marketplaces classified into 519 in the post-2017 NAICS regime) rise by 59 percent in the 10–25 km ring (p = 0.01, $n = 20$ DCs). These two findings give the spatial gradient the staggered structure we did not see in the county-level SCM: the closer the firm's value proposition is to physical proximity (interconnect for 518, then CDN for 512/516, then web latency for 519), the tighter its co-location with the hyperscale anchor.

Table 3: Within-site ring DiD: log active postings by NAICS subgroup, non-hyperscale only

NAICS subgroup	0–5 km	5–10 km	10–25 km	n DCs (max)
518 data processing / cloud	+183.9%***	+62.3%*	+41.9%	23
512/516 media + streaming	−27.4%**	+133.4%**	−18.9%	14
519 web search / info services	−45.7%	−41.0%**	+58.8%**	20
5112 software publishers	−56.5%	−13.5%	−69.7%	26
5415 tech consulting	−18.9%	−6.9%	−28.6%	43
5417 pharma R&D (placebo)	−0.5%	+9.7%	−6.1%	43
All non-NAICS-51 ("other")	+11.8%	+35.0%***	−0.5%	66

Notes: Each cell is the static within-site ring DiD coefficient on $\text{post}_{i,t} \times \mathbb{1}[r' = r]$, expressed as a percent change in active postings, from a separate regression that pools the indicated inner ring against the same DC's 25–50 km outer ring. DC × ring fixed effects and calendar-month fixed effects are absorbed; standard errors are clustered at the data-center level. Sample: 99 hyperscale clean-confidence data centers opening 2014–2024; event window ±48 months. Hyperscale-operator postings excluded so the estimates reflect downstream firms only. Stars: $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$. "All non-NAICS-51" is the catch-all retail/services/health/manufacturing universe and is reported as a benchmark for the broad local-multiplier effect that complements the tech-specific channels.

Third, NAICS 5112 software publishers and NAICS 5415 tech consulting do *not* show within-site rings: their geography is anchored to historic talent hubs (San Francisco, Seattle, Austin, Boston) that predate the hyperscale data-center build-out, and a new hyperscale opening in Loudoun County or Council Bluffs does not relocate them. The cross-section in Figure 3 shows these firms are concentrated near hyperscale anchors that happen to be in their existing metros (the 2–5 km peak in panel b), but the ring DiD shows the post-opening change in their counts is null. The takeaway is that the agglomeration we identify is genuinely a response to data-center entry, not the time-invariant fact that DCs and software firms both like to be in tech metros.

The bottom row of Table 3 reports the catch-all "other" universe: a +35 percent ring DiD at 5–10 km (p = 0.001, $n = 63$ DCs). The other category is dominated by retail, healthcare, food service, construction, transportation, and the various non-tech sectors. The

+35 percent on this universe is much larger in absolute terms (roughly 800 added active postings per DC-ring-month at the sample mean) than the 184 percent on NAICS 518 (roughly 92 added postings per DC-ring-month). The two channels are complementary: the broad local-multiplier effect lifts the overall labor demand; the NAICS-51 tech-agglomeration effect explains where the information-sector employment specifically goes.

4.4 Reconciling with the county-level NAICS 518 result

Table 1 reports that the county-level NAICS 518 SCM effect is *larger* in colocation-host counties (95 percent) than in hyperscale-host counties (68 percent). Table 3 reports that the sub-county ring DiD on non-hyperscale NAICS 518 is +184 percent inside 5 km of a hyperscale opening. These two findings are not contradictory once we separate two distinct channels.

The county-level NAICS 518 estimate captures the operator's own on-site staffing response: colocation operators run their own facilities at higher headcount per facility than hyperscale operators (who run lights-out facilities with leaner local staffing). This shows up at the county level because the data-center operator's payroll is itself a NAICS 518 establishment in the QCEW data.

The sub-county NAICS 518 estimate, by contrast, drops the hyperscale parent corporations and asks whether *other* NAICS 518 firms (colocation peers, edge providers, managed hosting) physically locate near the hyperscale build. The +184 percent effect at 0–5 km is a within-county pattern: in counties that contain a hyperscale build, the colocation-peer firms cluster in the same ZIPs as the hyperscale campus. In counties that contain only a colocation build, there is no comparably-sized hyperscale magnet to attract this colocation-peer cluster, so the within-county spatial signature is weaker (which the county-level NAICS 518 estimate does not see, because the county aggregate folds in the operator's own headcount).

The two findings together describe a layered geography: at the county level, colocation operators contribute more on-site headcount, but at the sub-county level, only hyperscale anchors generate the dense industry-peer agglomeration.

4.5 Mapping the agglomeration

Figure 4 shows the spatial pattern at four named hyperscale clusters with the same 25 km zoom. Northern Virginia (Loudoun/Ashburn) is the world's densest hyperscale cluster (127 hyperscale facilities operated by AWS, Google, Microsoft, and Oracle within 25 km), and is also the cluster with the densest surrounding presence of NAICS 5112, 518, and 519 firms in the LinkUp data. Council Bluffs (19 Google facilities) shows the canonical

NAICS 518 colocation-peer pattern: a small number of identifiable colocation firms physically adjacent to the Google campus. Prineville, Oregon (12 Apple and Meta facilities) and Quincy, Washington (12 Microsoft facilities) anchor hyperscale clusters in much smaller metros; the cluster maps for these rural sites are essentially empty of surrounding compute-using firms. The contrast is informative: hyperscale entry generates the downstream cluster only when it occurs in a metro with sufficient density of compatible compute-using activity to begin with. Hyperscale in Loudoun pulls in a dense surrounding ecosystem; hyperscale in Quincy or Prineville does not, at least not within the time window we observe.

5 Case study evidence

The mechanism implied by these results is that hyperscale data center campuses anchor a local cluster of internet-economy firms (web services, content delivery, AI inference, internet publishing) that benefit from low-latency access to compute and bandwidth. We document this pattern qualitatively in four established hyperscale clusters and one industry-wide trend.

Loudoun County, Virginia. Loudoun County is the world's densest data center cluster, hosting roughly 30 million square feet of data-center capacity. Approximately 70 percent of global internet traffic passes through Data Center Alley each day. The cluster originated with America Online's relocation to Loudoun in 1996 and MCI WorldCom's co-location around the same time, both predating the bulk of the data-center build-out. AWS opened its first Loudoun facility in 2006; Equinix's MAE-East peering site opened in 1998; QTS followed. The historical sequence is consistent with internet-economy firms (AOL, NAICS 519; MCI, NAICS 517) anchoring the cluster and pulling in subsequent data-center capacity. A Harvard Business School working paper on tech clusters [Kerr and Robert-Nicoud, 2020] documents that data centers tend to cluster in places with concentrated tech and finance employment, partly because those workers value low-latency access to compute. Vice Chair Mike Turner of the county board of supervisors observed that "all the trade unions are familiar with data centers and what they require," indicating that the cluster has produced a localized labor market for the supporting trades [Bertaccini, 2026].

Prineville and the Hillsboro tenant cluster, Oregon. Meta's Prineville campus (more than $2 billion in cumulative investment) anchors the Oregon hyperscale hub at Prineville, with adjacent QTS, Apple, and Meta facilities accounting for the twelve hyperscale data centers within 25 km of the Prineville centroid in our master registry. Roughly 130 miles to the northwest, the Hillsboro tenant cluster hosts data centers for TikTok, X (formerly Twitter),

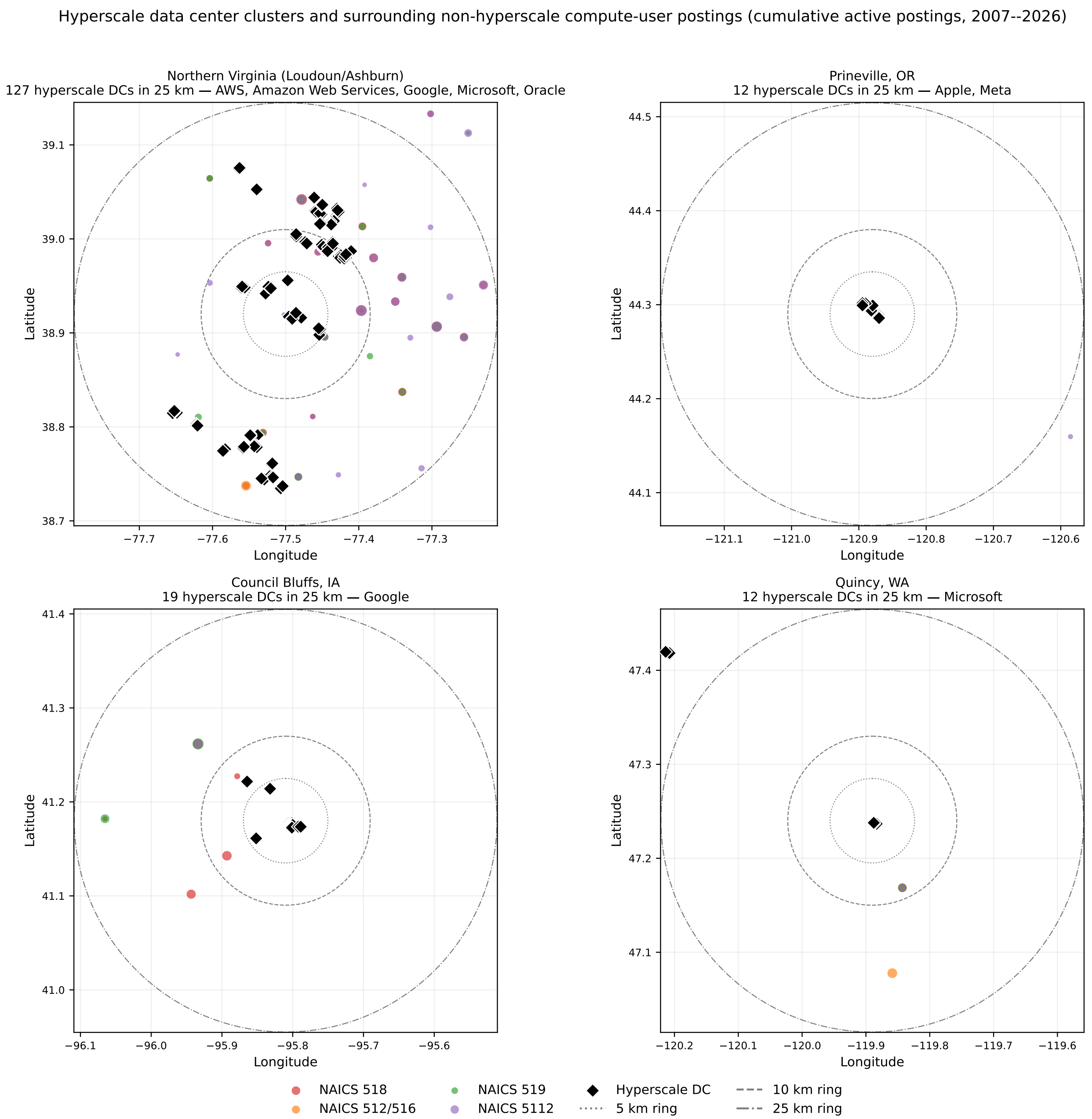

Figure 4: Hyperscale data center clusters and surrounding non-hyperscale compute-user firm postings

Notes: Each panel shows a 25 km radius around the named hyperscale cluster centroid. Black diamonds are individual hyperscale-operator data centers from the master registry; colored circles are ZIP-level cumulative active postings (2007–2026) from non-hyperscale firms in the indicated NAICS subgroup, sized by log posting count. Dashed circles are geodesic at 5, 10, and 25 km. The contrast between Loudoun (dense ecosystem) and Prineville/Quincy (rural, isolated hyperscale anchors) illustrates that downstream agglomeration is conditional on metro thickness.

LinkedIn, and Adobe. Three of those four are NAICS 519 internet-services firms; the fourth (Adobe) is a software publisher. The composition of the Hillsboro tenant cluster is a direct illustration of the mechanism we identify: internet-economy firms locate at the Oregon hyperscale facilities (typically in colocation arrangements at Hillsboro-area tenant space rather than at the Prineville campus itself) rather than at their corporate headquarters in San Francisco or Los Angeles, presumably because the latency and bandwidth advantages of being at the data-center hub outweigh the headquarters-metro agglomeration benefits [KLCC, 2026]. The cluster maps in Figure 4 make the surrounding spatial pattern at Prineville explicit: the hyperscale anchor is dense, but the immediate compute-user firm presence within 25 km is thin compared with Loudoun, consistent with the metro-thickness condition we discuss in the interpretation.

Council Bluffs, Iowa. Google has invested over $5.5 billion in its Council Bluffs data center since 2007, with another $1 billion announced. Google's presence anchored a broader Iowa hyperscale cluster: Meta now operates its largest cloud campus in Iowa, and Microsoft built a major campus in West Des Moines. The chained-investment pattern (one hyperscaler attracts another) is the supply-side analogue of the demand-side internet-economy clustering we identify, and is consistent with the local fiber and power infrastructure becoming a fixed cost that subsequent operators can amortize against [Google, 2026].

Quincy, Washington. Microsoft has been building data centers in Quincy for roughly twenty years, and Microsoft's own materials describe Quincy as "the model for our nationwide data center strategy." Sabey's Quincy campus alone employs 250 people, with entry-level technicians earning approximately $60,000 per year without a college degree [Banse, 2026]. The cluster has spawned a new high school, library, and city hall, plus secondary businesses serving construction workers and operations staff. ProPublica's investigation found that the cumulative tax exemptions (more than $300 million through 2023) delivered fewer direct jobs than promised, but the indirect ecosystem effects on supplier and adjacent firms were larger [Mickle, 2024].

The 2025–2026 pivot to AI inference zones. The industry-wide shift toward "inference zones," small latency-optimized facilities deployed near hyperscale regions to serve real-time AI workloads, makes the mechanism we identify mechanically explicit [CoreSite, 2026]. Inference-focused firms locate adjacent to where cloud regions and large language models are deployed; the value proposition is that "shaving milliseconds off inference times directly impacts conversion rates, customer retention, and operational safety." This is the explicit statement of why an AI startup, a CDN, or a real-time-services firm wants to be co-located with a hyperscale data center, and it is exactly the firm type that would classify as NAICS

519 in our QCEW data.

6 Interpretation

The hyperscale information-sector effect documented in the main paper is real, replicates here under the same SCM design, and decomposes cleanly along two dimensions. At the county level, NAICS 519 (other information services) is the dominant channel, NAICS 517 (telecommunications) is a secondary channel that contributes in both operator types, and NAICS 518 (data processing and hosting) shows up larger in colocation-host counties through the operator's own on-site staffing. At the sub-county level, the LinkUp postings data add a staggered downstream-customer gradient: NAICS 518 colocation peers cluster within 5 km of hyperscale builds (+184 percent), NAICS 512/516 streaming and media firms at 5–10 km (+133 percent), and NAICS 519 web-search and information firms at 10–25 km (+59 percent). The two layers describe distinct but complementary phenomena: county-level NAICS 518 reflects who staffs the operator's own facility, while sub-county NAICS 518 reflects who locates next door to the operator. The first is bigger for colocation; the second is bigger for hyperscale.

The mechanism that fits the combined evidence is consumption-side rather than supplier-side. Hyperscale builds anchor a local cluster of firms that consume large amounts of compute and bandwidth and that benefit from being physically close to the data center. The distance-graded structure of the sub-county evidence reinforces this interpretation: the firms with the strongest physical-proximity value proposition (interconnect-dependent colocation peers) are tightest to the anchor, those whose value proposition is more bandwidth-and-CDN sensitive (streaming, content delivery) sit a ring out, and those that are sensitive to web-latency aggregates (search, social, marketplaces) sit a further ring out. Colocation builds, by contrast, do not anchor the same kind of consumption-side cluster, because colocation tenants are typically remote enterprises whose workforces stay in the tenant's headquarters metro. The colocation operator's on-site staffing response is real but modest in absolute terms and is overwhelmed in the NAICS 51 aggregate by the absence of a localized internet-economy cluster.

Two important qualifications follow from Figure 4. First, the downstream agglomeration is conditional on metro thickness. Loudoun's hyperscale cluster sits inside the Washington DC metro and inherits the federal IT contractor ecosystem, the NoVA tech-talent base, and the proximity to peering exchanges; its surrounding compute-user-firm density is correspondingly high. Prineville and Quincy host hyperscale facilities in much smaller and more isolated metros, and the surrounding compute-user-firm density remains thin even after a decade of hyperscale presence. The county-level SCM aggregates over these geographies; the postings microdata makes the dispersion explicit. Second, the broad local-multiplier channel is larger in absolute terms than the tech-specific channel. The +35 percent on the catch-all non-NAICS-51 universe adds roughly 800 active postings per DC-ring-month at the sample mean, against roughly 92 added by the +184 percent NAICS 518 effect. The information-sector story is the most distinctive piece of the post-DC labor-market response, but it is not the largest piece by employment count.

The combined evidence refines, rather than overturns, the operator-type mechanism in the main paper. The hyperscale-versus-colocation distinction does matter, and it does shape the local labor market response. But the firms that follow the hyperscale operator into the host county are the operator's downstream *customers* (internet-economy firms that consume the compute and bandwidth) rather than the operator's upstream *suppliers* (fiber installers, NOC contractors). This has direct implications for incentive design: targeting incentives toward operator entry does not automatically generate the supplier ecosystem one might expect from a manufacturing-style supply chain; what it generates is a customer ecosystem that depends both on operator type and on the existing thickness of the host metro's compute-user firm base. Hyperscale operators serve latency-sensitive workloads at scale, and the agglomeration response materializes when the host metro already has compute-using firms that can move next door; colocation operators do neither.

7 Caveats

Two caveats temper the interpretation. First, the NAICS 519 effect is largest in the pre-2017 NAICS regime, where the code most cleanly captured internet publishing. Recent NAICS revisions (2017 and 2022) move pieces of internet-economy activity into adjacent codes, including a new sector 516 created in 2022. For paper revisions or extension to recent years, the relevant aggregation may shift toward NAICS 516 plus 519 plus subsectors of 518; the specific numbers reported here will need updating once the QCEW publishes consistent NAICS 2022 data through the post-treatment period.

Second, the SCM-based decomposition implicitly assumes that within-subsector pre-period matching delivers a credible counterfactual at the subsector level. The pre-period RMSPE values reported in our tables are mostly small (under 0.17 in log points), and pre-trend tests in the TWFE event-study cross-check are mostly insignificant, but the colocation-host subsample is particularly thin (twenty-two treated counties) and a few subsectors have pre-RMSPE values close to zero, suggesting the SCM is fitting the pre-period almost exactly because the small-county subsector employment level is itself near zero pre-treatment. We flag this as a measurement caution rather than a substantive concern; the qualitative pattern is robust across subsamples and across windows.

Third, the sub-county postings spec relies on a manually curated COMPANY_NAME-to-NAICS crosswalk of the top 500 US LinkUp employers, which covers 51.7 percent of all USA postings. The remaining 48 percent is rolled into "other," which captures the broad local-multiplier effect. A more complete crosswalk would tighten the NAICS-subgroup estimates but is unlikely to change the qualitative ring-by-ring pattern. The ring-DiD estimates themselves rest on relatively small DC counts for the NAICS-518 (17–23 DCs), 512/516 (11–14 DCs), and 519 (20 DCs) subgroups; the headline percent-lift point estimates should be interpreted as directional rather than precisely quantified. We show the heatmap-style report in Table 3 rather than highlighting any single coefficient.

Replication. The county-level decomposition uses the BLS QCEW singlefile archive at the three- and four-digit NAICS level. The puller, the SCM decomposition, the NAICS 519 reclassification check, and the table-builder are in the project repository under `code/`. The sub-county postings analysis (Section 4) uses the Dewey LinkUp archive (250 parquet files, 61.7 GB, 2007–2026) staged on the UC Merced cluster. The cluster-side aggregator with the curated COMPANY_NAME-to-NAICS classification, the local panel build, the ring DiD estimator, and the figure builder are all in the project repository under `code/cluster/` and `code/` respectively.

References

Tom Banse. A small town in Central Washington is Microsoft's answer to the data center backlash. KUOW Public Radio, 2026. Available at https://www.kuow.org/stories/a-small-town-in-central-washington-is-microsoft-s-answer-to-the-data-center-backlash.

Maurizio Bertaccini. Loudoun county, virginia: The heart of the data-center boom. City Journal, 2026. Available at https://www.city-journal.org/article/loudoun-county-virginia-data-centers-construction.

CoreSite. Inference zones: How data centers support real-time AI. CoreSite, 2026. Available at https://www.coresite.com/blog/inference-zones-how-data-centers-support-real-time-ai.

Google. Council bluffs, iowa — Google data center location. https://datacenters.google/locations/iowa/, 2026.

William R. Kerr and Frederic Robert-Nicoud. Tech clusters. *HBS Working Paper 20-063*, 2020.

KLCC. Oregon's data center explosion: Who benefits and who bears the cost? KLCC, Oregon Public Broadcasting, 2026. Available at https://www.klcc.org/podcast/oregon-on-the-record/2026-04-02/oregons-data-center-explosion-who-benefits-and-who-bears-the-cost.

Tripp Mickle. A tax break for Washington data centers promised jobs. Is it paying off? ProPublica, 2024. Available at https://www.propublica.org/article/washington-data-centers-tech-jobs-tax-break.