

Pocchiari, Martina; Roos, Jason M.T.

Working Paper

Digitization and Community Participation

CESifo Working Paper, No. 10841

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Pocchiari, Martina; Roos, Jason M.T. (2025) : Digitization and Community Participation, CESifo Working Paper, No. 10841, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/282529.2>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

CES ifo

**10841
2025**

Original Version: December 2023
This Version: July 2025

Working Papers

Digitization and Community Participation

Martina Pocchiari, Jason M. T. Roos

CES ifo

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

Digitization and Community Participation

Martina Pocchiari*

Jason M.T. Roos[†]

Abstract

Digitization technologies reshape the way people engage in social and business activities, transitioning events from in-person to virtual formats. However, existing evidence on the effects of digitization on participation remains fragmented and contradictory. The authors address this gap by analyzing data describing 645,013 events organized in 2024 by 12,105 diverse communities comprising 493,143 individuals. Leveraging the structure, granularity, and attributes of the data, which include large amounts of unstructured text, the study finds that digitization has, on average, positive effects on participation. These effects, however, are highly heterogeneous, depending on group interests, event features, and even time. For example, digitization negatively affects participation in highly specialized and technical events, especially those emphasizing skill development in the areas of technology and business. Digitization effects fluctuate significantly over the year for certain types of events and groups, occasionally changing sign throughout 2024. The results reconcile some of the contradictory evidence on digitization effects and highlight potential gaps between the supply of digitized events and participants' preferences. The study also identifies key opportunities for businesses and policy-makers driving digital transformation, emphasizing that delivering specialized knowledge and supporting physical tasks represent the most impactful avenues for the digitization of future events.

Keywords: Digitization, Community engagement, Natural language processing, Heterogeneity.

JEL Classification Codes: M31, O33

*Esade Business School, Universitat Ramon Llull; corresponding author. martina.pocchiari@esade.edu

[†]Rotterdam School of Management, Erasmus University. jroos@rsm.nl

The authors acknowledge the support of all those who contributed to the improvement of this study over time: the faculty and PhD students in the Marketing Management department at the Rotterdam School of Management, and the participants of EMAC 2019-2021, Marketing Science Conference 2020, Italian Marketing Society conference 2020, CESifo area conference on the economics of digitization 2023, Digital Economy Workshop Berlin 2025, and Consumers + Technology Dialogues conference 2025. The authors thank Francesco Capozza, Max Pachali, Andreas Bayerl, and JK Chong for their helpful feedback, as well as the community managers who have shared their expert insights on community digitization over the years: Anastasia "Stacy" Raspopina, Tristan Lombard, Scott Tran, Tali Vasilevsky, Samy Zerrouki, and the hosts of the Communities Show.

1 Introduction

Digitization enables social and business activities to take innovative formats, including live streams, webcasts, online meetings, and virtual classes. Compared to in-person alternatives, these formats are often more affordable, accessible, and efficient to organize and attend. As a result, many events – such as trade shows, product launches and demonstrations, press conferences, work meetings, art exhibits, social support groups, and sports classes – have migrated in recent years from face-to-face to digital formats. The global market size for digitized activities and gatherings is estimated at \$98B, and is projected to grow at a compound annual growth rate of 20% from 2025 to 2030 (Grand View Research, 2025).

Despite the increasing digitization of events, managerial advice on whether to digitize different business and social activities remains limited in scope and often fragmented. There are several reasons for this. First, studies of digitization are typically conducted using highly specific samples and organizations, impeding generalizability. These samples include, for example, students at a for-profit university (Bettinger et al., 2017), workers at an Italian multi-utility company (Angelici & Profeta, 2024), high school students participating in a financial education program (Sconti, 2022), and recent graduates employed by a technology company (Bloom et al., 2024). Second, studies that estimate digitization effects sometimes yield contradictory results, even within the same managerial domain. This hinders the development of comprehensive, practical recommendations. For instance, in the educational domain, Gortazar et al. (2024), Lee et al. (2023), and Sconti (2022) suggest that digitizing courses may improve academic performance, whereas Di Pietro (2023), Gortazar et al. (2024), Kofoed et al. (2024), and Lichand et al. (2022) find negative effects of digitization on learning and socialization. In remote work, Angelici and Profeta (2024) and Bloom et al. (2024) find that virtual meetings for co-working teams can improve business relationships and employee engagement, while Aarons-Mele (2022) and Capossela (2022) suggest that online meetings and remote work make employees more anxious and disengaged. Third, the stability of digitization effects over time is unclear, as short-term effects can either dissipate (Sconti, 2022) or persist (Bettinger et al., 2017; Lee et al., 2023) in the long run. These limitations and discrepancies have not prevented the emergence of “one-size-fits-all” heuristics and rules of thumb regarding the digitization of business and society. Yet, the contradictory evidence about digitization suggests the need for a more comprehensive and nuanced understanding of digitization effects.

In this article, we conduct a large-scale empirical investigation of the effects of digitization on participation in heterogeneous activities organized by thousands of diverse communities and spanning myriad categories and activity types. We compile a novel, extensive, and highly detailed data set collected from Meetup, a global platform for organizing local communities and events. The data describe 12,105 public communities active in 10 U.S. cities. The data include the full population of 645,013 events, of which 215,452 are digitized, organized by the communities in 2024, and contain detailed characterizations of these events in terms of their purpose, format, and logistics. The types of communities and events in the data are highly diverse, ranging from fitness classes to software development workshops, from social support meetings to group meditations. The data also describe 493,143 community members and their 231,227,896 event registration decisions, which comprise the outcome of interest for this study.

The scope and granularity of the data allow us to study the effect of digitizing diverse events on the participation of individual members, as measured by event registrations. However, identifying the causal impact of digitization on event participation is not trivial. The greatest challenge in estimating causal effects across thousands of diverse communities is that randomized field experiments are impractical, or even unethical, for many types of events (e.g., social support groups for people with health concerns, or accessible fitness classes for people with disabilities). Using observational data allows for heterogeneous analyses without ethical or practical issues, but introduces concerns about unobservable factors that may influence both event digitization and individual participation.

To tackle this problem, we closely follow the approach of recent marketing studies, which leverage large data sets and the properties of their empirical settings to perform observational causal inference on granular units of data (Luo & Ranjan, 2025; Shapiro et al., 2021). Specifically, we employ the following strategies. First, we survey the academic and managerial literature in marketing, management, and economics to construct a comprehensive set of observable and unobservable variables that might simultaneously influence (i) a group’s decision to digitize an event, and (ii) a member’s decision to register for a prospective event. Thus, we derive an extensive set of potential confounders of the causal relationship between event digitization and member participation. This set includes variables that represent characteristics of events and organizing groups. This effort yields a comprehensive list of antecedents and confounders of digitization across a broad range of empirical contexts, answering the call by Brynjolfsson et al. (2021) for more research on previously unmeasured factors affecting the digitization of the economy.

Second, we explicate the relationships among these variables to derive a sufficient set of assumptions for the causal identification of heterogeneous digitization effects. Our identification strategy depends on observing potential confounders, some of which are only represented in free-text event descriptions. We construct quantitative representations of these variables by combining unstructured text with generative artificial intelligence.

Third, we leverage important sources of identifying variation in the data provided by (i) the same groups organizing events with different formats over time, and (ii) the repeated exposure of the same individuals to both digitized and in-person events. This data structure, in combination with the identifying assumptions, allows us to control for the potentially confounding effects of the aforementioned variables. We thus obtain granular units of observational data that are highly homogeneous apart from the digitization status of the events, and ensure that any remaining variation does not bias estimates of digitization effects.

Fourth, we estimate digitization effects using a doubly robust regression estimator, which provides an additional safeguard against model misspecification (Bang & Robins, 2005; Gordon et al., 2019). Because digitized events can almost always be converted to in-person, but not all in-person events can be digitized, we estimate conditional average treatment effects among the treated (i.e., the effects of digitization on participation among the subset of events that were digitized).

The results yield important managerial insights. Across 215,452 digitized events, the digital format is estimated to be responsible for a net gain of 67,430 event registrations, or an average increase of 0.13 registrations out of an average of 3.55 attendees per digitized event. This positive average digitization effect,

however, masks the high degree of heterogeneity in how digitization impacts different types of groups and events. Positive effects of digitization on event registration are more likely for events organized by groups with community and lifestyle-oriented topics, such as *Spirituality* (+0.21% individual registrations), *Travel* (+0.14%), *Community/Culture* (+0.12%), *Community/Causes* (+0.13%), and *Dancing/Fitness* (+0.13%), as well as events that place less emphasis on *social interactivity* (+0.11%). Negative effects of digitization on event registration are more likely among groups and events with a technical focus, such as the *Tech* (−0.19%) and *Business/Careers* (−0.05%) interest categories and events requiring participants to hold prior *specialized knowledge* (−0.11%). Events emphasizing both *physicality* and *social interactivity* (e.g., an art-making workshop where attendees can meet new friends) show the steepest declines in registrations when digitized, whereas physical activities that do not entail social interaction (e.g., DIY food tastings) often benefit from digitization. A set of sensitivity analyses and falsification tests shows that these results are robust to potential uncontrolled sources of confounding bias.

To provide even more nuanced managerial insights, we also estimate digitization effects for specific combinations of topics and event features. We find that *Business/Careers* events that emphasize both *physicality* and *social interactivity* see a −0.12% decline in individual-level registrations when digitized, whereas business events that focus on *skill development* but do not require *physicality* incur almost no penalty from digitization. Similarly, digitization is positive for events that promote *skill development* that do not require prior *specialized knowledge*, but negative for advanced courses that require *specialized knowledge*.

The results also show that digitization effects fluctuate from month to month in 2024, and to an unexpected extent in a few notable cases. The typically negative digitization effect for *Tech* events turns strongly positive (+1.55%) around September. Similarly, the generally negative effect for *Business/Careers* becomes positive in early summer (+0.05% in July) and again in early autumn (+0.04% in September), and the usually negative effect for *Travel* events changes direction in July (+1%) and November (+0.48%). These results suggest that the success of online formats might not be static and that contextually relevant and time-varying factors, such as exam calendars, travel seasons, and industry release cycles, could also influence participation in digitized activities.

Finally, the results suggest a potential supply-demand mismatch in the digitization of different activities. Events that exhibit positive digitization effects (for instance, events in *Travel* topics that involve no *physicality* or *social interactivity*) appear to be under-supplied in digital form, suggesting organizers are not adopting digitization technologies for events with certain combinations of characteristics, even though those combinations may benefit from digitization. At the same time, events that see lower registrations when digitized, such as highly interactive *Tech* events focused on *skill development*, are frequently offered in digitized formats. This misalignment might point to untapped growth opportunities for both event organizers and developers of digitization technologies. We conjecture that the most impactful opportunities for event organizers and digitization companies might lie in delivering specialized knowledge and remotely performing physical tasks. Collectively, the findings of this study demonstrate that digitization is not a one-size-fits-all intervention. Rather, its benefits hinge on what type of experience is offered, how it is delivered, and when it is scheduled.

The remainder of the paper is organized as follows. We first describe the empirical context, the scope and structure of the data collection, and the data patterns. Then, we outline the identification strategy and empirical models. We then present the results and conclude with a discussion of their implications for businesses and policy-makers, as well as suggestions for future research.

2 Empirical Context

We compiled a comprehensive dataset of groups, events, members, and individual event registrations from Meetup, a leading global platform for organizing shared interest groups and events. In this section, we first give an overview of the Meetup platform and the institutional context in which Meetup groups, events, and members operate. We also describe the factors that influence organizers’ and members’ decisions related to event digitization based on prior literature, including theories on the economics of digitization and industry guidelines. Next, we provide a detailed description of the data set used to estimate the effects of digitization on event participation. We highlight its extensive coverage, granularity, and potential to measure covariates and fixed effects related to the supply and demand for digitization that we identify. We conclude with descriptive evidence showing that the data contain sufficient variation for causal estimation.

2.1 Overview of the Meetup Platform

Meetup is a leading global digital platform for community building, active since 2002. The primary purpose of Meetup is to help shared-interest *groups* organize *events* for their *members*. Members can then use the platform to *register* for events. As of early 2025, Meetup provided a platform for 300,000 groups hosting 100,000 events per week, and 52 million members in 193 countries (Meetup, 2025). The platform also has strong business relevance: many companies use Meetup to develop and support brand communities, and Meetup offers a paid service (“Meetup Pro”) to help companies manage groups professionally.

Meetup *groups* span a wide range of interests and topics, including technology, business, sports, education, and entertainment. Meetup users utilize the platform to discover groups whose interests match their own and to stay informed about events hosted by those groups. When a Meetup user joins a group, they become a *member* of that group. The primary purpose of group membership is to facilitate event planning for organizers and event registration for members.

Group organizers use Meetup’s platform primarily to schedule *events*. These events can vary widely in their purpose and format, including workshops, product previews, coding tutorials, conferences, parties, dance lessons, and book clubs. Group organizers can choose to organize these events digitally or in person. Digitized events can take various forms, including webinars, virtual conferences, discussion panels, online classes, and asynchronous video feeds. When an event is digitized, organizers mark it as “online” on the event interface, and the event is displayed as “online event” to prospective attendees on Meetup.¹ In addition

¹Organizers can also mark their events as “hybrid”, if the event combines both in-person and digital formats. However, in 2024, only 469 events were marked as hybrid (0.072% of the total events.)

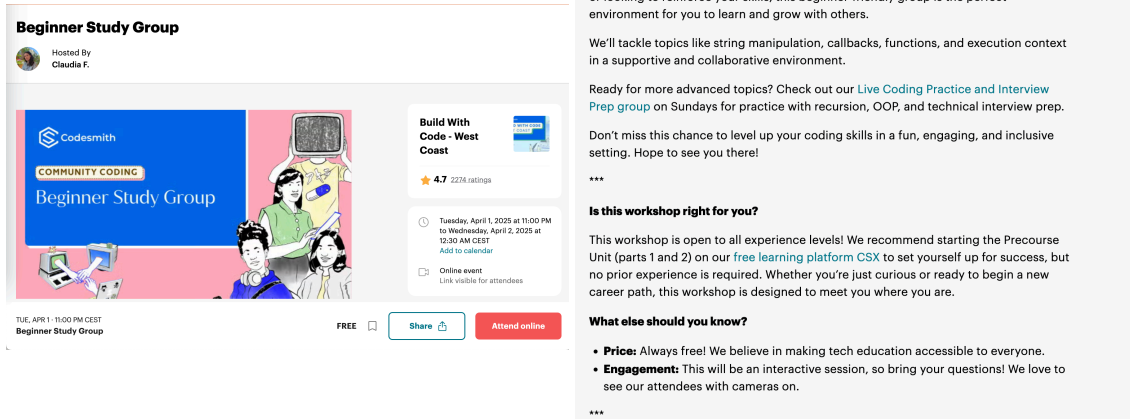


Figure 1: Examples of the Meetup Event and Registration Interface in 2024-2025.

to the structured information provided via the user interface, organizers also provide members with detailed information about the event’s purpose, format, and logistics through a free-text event description.

The *event registration* system allows members to indicate their intention to attend upcoming events. Although group organizers strongly encourage registering for events for effective event management, registration is not mandatory, which often results in fewer registrations than the number of group members. This study focuses on registrations as a form of community engagement and a means of expressing the intention to participate in Meetup events. This is in line with recent work exploring the impact of digitization on intention outcomes (e.g., Lee et al. 2023; Touré-Tillery and Wang 2022). In addition to helping organizers at Meetup, registration rates are considered key performance indicators (KPIs) for measuring the success of events and communities in the industry (e.g., EventUp Planner 2024).²

Figure 1 shows examples of the interface that Meetup presents to prospective attendees who are considering attending an event. The event depicted in Figure 1 shows the event title (“Beginner Study Group”) and a detailed event text description (“Details”), the organizing group and organizers (“Build with Code - West Coast”, “Claudia F.”), the logistics (date and time of the event), the event format (“Online event” in this case), and the registration button (“Attend” or “Attend online” in the case of online events).

2.2 Organization of and Participation in Digitized Events

There is no consensus among researchers or practitioners about why some events are digitized and others are not, or why some individuals participate in digitized events when others do not. Furthermore, discussions about the determinants and consequences of digitization often take place in relatively narrow empirical contexts. Nevertheless, when looking across prior studies, several recurring factors stand out as potentially

²A complementary measure of members’ participation in Meetup communities would be their actual event attendance. However, this variable is not available in the data. Industry sources report that no-show rates in both virtual and in-person events range between 30% and 50%, and do not report significant differences in no-show rates between event formats (Bremer, 2022). We discuss this potential intention–behavior gap in Section 6.

important across a wide range of groups and events: the *interest topics* of a group; whether the event emphasizes *skill development*, *specialized knowledge*, *social interactivity*, and/or *physicality*; and *time*. Some of these factors have received attention only as determinants of event digitization; in those cases, we posit that event organizers digitize events in part because they believe attendees will benefit from digitization. Accordingly, the factors leading to digitization are also expected to be related to members' heterogeneous responses to digitized events. We discuss these factors and potential sources of treatment effect heterogeneity next.

2.2.1 Groups and Interest Topics

Past research suggests that the main *interest topics* of a group should be strongly related to digitization, and that digitization effects should vary across groups catering to different interests. A notable example is the “education” interest topic. Bettinger et al. (2017), Lee et al. (2023), and Sconti (2022) show evidence that it is relatively easy for organizers operating in education (e.g., primary schools and colleges) to digitize their activities, but also that the effect of digitizing educational activities on learners' engagement can be positive or negative. Similarly, Sarabipour (2020) and Zajdela et al. (2025) suggest that digitization has enabled greater participation by researchers in academic conferences, but also that in-person conferences remain more conducive to forming social bonds compared to virtual ones. Other industry reports suggest that digitization is most likely to be adopted by organizers of events catering to specific topics, such as technology, finance, and healthcare (Business Research Insights, 2024; D'Souza, 2025). Although the literature is ambiguous about which additional topics are more or less amenable to digitization, there is a fair degree of consensus that the interest topics of the organizing groups, in general, are important. Furthermore, particular groups could have intrinsic preferences for adopting digitized formats; for example, a group organized by a software developer may be more likely to organize digitized events compared to a group organized by a chef.

2.2.2 Event Characteristics

At the event level, academic and industry literature suggest that the characteristics and goals of an event can influence decisions regarding digitization. One of these characteristics is whether the event is focused on *skill development*. Bonsall (2023) suggests that organizers may find transactional meetings, which focus on task completion and skill development, easier to conduct online. Similarly, people often prefer to participate in digitized activities, such as digital courses, to learn a skill or increase their knowledge of a particular topic (Ingram & Drachen, 2022).

Another important characteristic is the *social interactivity* of an event. Particularly, organizers and attendees may prefer in-person formats for events and business activities focused on socialization and networking (Capossela, 2022; Mills, 2023). In a similar discussion, Bonsall (2023) and Ingram and Drachen (2022) suggest that online formats are less convenient when gatherings have a relational focus and aim to strengthen social connections.

Digitization decisions could also be influenced by the extent to which events require different levels of expertise and *specialized knowledge*. For example, recent academic research has found that the effects of

digitization in the context of education vary depending on the educational level of the students (e.g., Bettinger et al. 2017; Kofoed et al. 2024; Lee et al. 2023 and Sconti 2022).

Finally, events involving activities that hinge on *physicality* and are inherently more “hands-on” could be more difficult to digitize. For instance, a group organizing a barbecue dinner in the park or a run through a hiking trail could be less likely to digitize its activities without altering the logistics of the events, compared to a group organizing a book club or a software demonstration.

2.2.3 Time-Varying Factors

In addition to factors that vary by group and event, idiosyncratic factors might determine both event digitization decisions (e.g., weekly weather, the launch of new digitization tools) and members’ decisions to participate in digitized events (e.g., individual digitization preferences). One particularly important idiosyncratic factor is *time*. The results of recent research on the stability of short-term digitization effects suggest that digitization effects could be time-varying and could either persist (Bettinger et al., 2017; Lee et al., 2023) or dissipate over time (Sconti, 2022).³

In light of the discussion above and our objective to estimate the effect of digitization on event participation, we expect each group to decide whether to digitize its events, and each individual to decide whether to participate in those events. These decisions could be influenced by factors that vary by group, time, event, and individual. *Within a group* organizing events and *within individual members* deciding whether to participate – keeping constant a detailed list of factors that vary by group, event, individual, and time – we expect individual participation to differ between events that are digitized and those that are not. Furthermore, we expect participation in digitized versus in-person events to be heterogeneous across groups with different substantive interest topics, event characteristics (namely, emphasis on *skill development*, *social interactivity*, *specialized knowledge*, and/or *physicality*), and time.

3 Data

3.1 Data Collection and Structure

We collected data via Meetup’s public API in early 2025. The data include all publicly listed Meetup groups active in the 10 largest U.S. cities (US Census Bureau, 2022). For each group that matches these criteria, we compile complete case information on active members and public events organized by these groups between January and December 2024. Web Appendix A provides more details on the data collection process.

³These studies encountered limitations that are mostly related to their empirical settings, such as the inability to follow up with the same experimental participants over periods longer than 6–9 months. Furthermore, most studies that have examined the stability of digitization effects over time refer to the digitization of specific economic sectors, such as education (e.g. Bettinger et al. 2017; Lee et al. 2023), and remote work (e.g. Bloom et al. 2024).

The group-level data contain the group’s name, a free-text description explaining the group’s purpose, and a high-level interest category.⁴ As we explain below, we do not use these category indicators to measure interest topics; instead, we extract latent interest topics from groups’ text descriptions. We index each group in the data by g .

The event-level data include the event name, a detailed text description, the organizing group, and an indicator for whether the event is online or in-person. We index each event in the data by e . When presenting the estimation strategy, we use the notation $g[e]$ to refer to the group g organizing event e , and denote the set of events organized by group g in 2024 by \mathcal{E}_g . The event-level data also contain variables related to event logistics, such as the date and time of the event; the venue; indicators for outdoor, recurring, and “hybrid” events; and COVID-related prerequisites, such as mask and vaccine requirements. We index the week in 2024 when the event took place by t (or $t[e]$), and create an indicator for whether the event was organized on a weekend. We also extract additional event descriptors from their text descriptions, as explained below.

The user-level data include unique identifiers for each user of the Meetup platform, which we index by i . Although membership always exists in the context of a particular group g , users carry their individual identifier i across all the groups to which they belong. Group g ’s membership changes over time as new users join and old users leave, but we observe a snapshot of group g ’s membership with every event. This set, which we denote by \mathcal{J}_e , contains all *potential* attendees to event e .

We aim to estimate various treatment effects of digitizing events. The treatment variable of interest for this study is thus the digitization status of event e , which we denote by D_e . We set $D_e = 1$ if event e is marked as online, and $D_e = 0$ otherwise. The outcome of interest in this study is whether group member i registers for a particular event e . We observe these decisions in the data and denote them by Y_{ie} . We set $Y_{ie} = 1$ if member i registers for event e , and $Y_{ie} = 0$ otherwise.

The complete data set comprises 12,105 groups, 645,013 events (averaging 53.3 events per group), 493,143 users, and 231,227,896 registration statuses recorded at the individual and event levels over 53 weeks. The data reflect variation in (i) the extent to which events are digitized every week and over time, (ii) the extent to which the same groups organize events of different formats, and (iii) the extent to which the same individuals are exposed to different event formats. In reference to point (i), 215,452 events (32.7%) were digitized in 2024, and Figure 2 shows that the share of digitized events per week varied between 31% and 38% over the course of the year.

In reference to points (ii) and (iii), the data show that individual users belonged to a median of 2 groups and had the opportunity to attend a median of 124 events throughout the year. 276,261 (56.02%) users were members of groups that organized both in-person and digitized events. 48,958 (9.93%) users were members of groups that organized only digitized events, and 167,924 (34.05%) were members of groups that organized only in-person events. 272,852 (55.33%) users were members of two or more groups, and 204,671 of them

⁴The group-level raw data also contain a set of more fine-grained interest topics. The group organizer can choose these from a predefined list comprising more than 36,000 options, such as *art exhibit*, *angel investors*, *mindfulness meditation*, and *house parties*. However, the API queries described in Web Appendix A, which require the input of a location and a keyword, return the same list of topics for all groups corresponding to the same input location-keyword pair. Therefore, due to a lack of variation in topics across groups, the raw topics data cannot be used to estimate heterogeneous treatment effects.

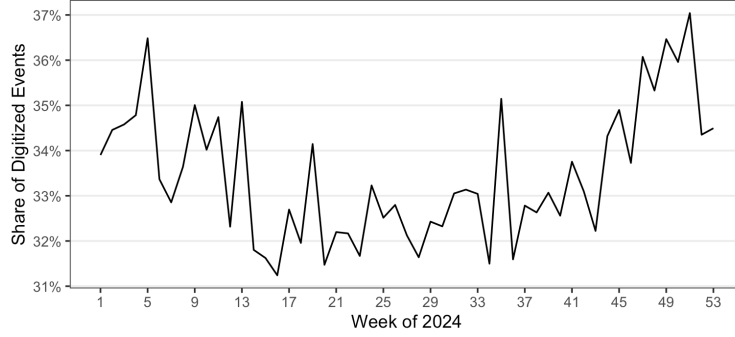


Figure 2: Share of Digitized Events Per Week in 2024.

(75.01%) were exposed to both digitized and in-person events. Among the 220,291 members belonging to only one group, 71,590 (32.49%) were exposed to both digitized and in-person formats.

The aforementioned variation in event digitization within and across groups and individual members is crucial for estimating the causal effect of digitization on event registrations. In the presence of such variation, we can leverage the repeated observation of the same individual member exposed to different digitization formats, while keeping everything else as constant as possible, to estimate causal effects.

3.2 Feature Engineering: Group Interest Categories and Topics

As discussed above, groups that organize events relevant to different interests may differ in their likelihood of offering digitized events, and their members may vary in their average preference for digitized events. We therefore seek to estimate the effects of digitization on participation that potentially vary by interest topic. As previously mentioned, the data contain information about each group’s high-level interest category. Each group can choose only one category from a set of 21 options, including *career and business*, *health and wellbeing*, and *social activities*.

Although the interest categories contain information about groups, there is a conceptual concern about what this variable represents, as each group can choose only one interest category, excluding all others. A group that, for instance, promotes better mental health by organizing museum visits might struggle to choose between the *arts and culture* and *health and wellbeing* categories. In contrast, the group-level text descriptions typically contain all of the relevant information needed to understand a group’s purpose.

For these reasons, we estimate a latent Dirichlet allocation (LDA) topic model using information contained in group names and text descriptions. This procedure yields a rich characterization of the various interests catered to by each group, while keeping the number of latent interests tractable. Using quantitative criteria from Arun et al. (2010), Cao et al. (2009), and Deveaud et al. (2014), we set the number of latent topics to eight. The LDA model assumes that each group reflects a mixture of these eight topics, with mixture weights given by a vector of topic probabilities denoted by C_g . These topic probabilities reflect the extent to which each group caters to each of the eight latent topics. In the remainder, we refer to these eight LDA topics as “topics”, “latent topics” or “interest topics”. Table 1 characterizes the eight interest topics based

Topic	Label	Most Frequent Words
1	Tech	Technology, gaming, learning, development, data, AI
2	Community/Causes	Club, community, local, vegan, volunteering, animal
3	Travel	Travel, people, social, culture, adventure, Asian, trip, photo
4	Dancing/Fitness	Dancing, fit, sport, class, player, level, trainer
5	Business/Careers	Business, networking, career, professional, market, entrepreneur
6	Community/Culture	Community, women, book, art, explore
7	Socialization	Activity, social, people, fun, friend
8	Spirituality	Spirituality, lifestyle, support, health, meditation, healing, practice

Table 1: Latent Interest Topics Recovered via LDA.

on the words that emerge most frequently from the group titles and descriptions. Further details about the LDA procedure are provided in Web Appendix B.

The latent topics, C_g , correspond well with the topic category variable provided by Meetup. To compare the two, we calculate the average topic weight for all groups with the same Meetup interest category. Table W1 in Web Appendix B shows that groups in the same Meetup interest category may cater to relatively different latent topics. For instance, groups in the *Identity & language* category load high not only on *Community/Culture*, but also on *Dancing/Fitness* and *Socialization*. An example of a group catering to multiple interests is “La Mesa Espanola - Houston,” whose organizers describe it as a “social / study group for learning the Spanish language and culture [...] and participat[ing] in other Spanish related activities such as going to a movie or dinner.” Another is “The New Language Exchange”, which provides “weekly gathering on Sundays to practice languages” as well as “walking tours, picnics, hiking, going to the beach, and more.” Although there are only 8 latent topics compared to Meetup’s 21 predefined categories, allowing groups to be described by a mixture of latent topics leads to a richer representation of their focal topics, which at the same time is more amenable to estimation.

Table W2 in Web Appendix B also shows that the LDA topic modeling produces a more balanced distribution of digitization across latent topics compared to Meetup’s one-group-one-category variable. Although not the objective of extracting topics loadings from text descriptions, the resulting balance in the treatment variable is favorable for estimation.

3.3 Feature Engineering: Unstructured Event Data and Event Characteristics

In the earlier discussion of theory-relevant factors that explain the variation in the supply and demand of digitized and non-digitized events, we identified that *social interactivity*, *specialized knowledge*, *skill development*, and *physicality* are important event characteristics. Accordingly, we construct four variables to represent these event-level attributes.

The variables are created using the rich and detailed, but unstructured, information provided in the text descriptions for each event. Following recent studies that have successfully used large language models

(LLMs) for similar classification tasks (e.g. Krugmann and Hartmann 2024; Le Mens and Gallego 2025), we use `ChatGPT-4o-mini` to classify event descriptions along the theory-relevant constructs. In Web Appendix C, we demonstrate that the classification results obtained using `ChatGPT-4o-mini` achieve high inter-rater reliability across multiple classification rounds and that the results are robust to the use of a different large language model (`deepseek-reasoner`). Web Appendices C.1 and C.2 report more details on the prompts, performance, stability, and reliability of the classifications.

The resulting binary variables, each of which measures whether an event emphasizes one of the four theoretically relevant characteristics, are denoted by the event-level vector X_e (Web Appendix C.3 shows the prevalence of the four variables across events). These variables play crucial roles both as part of the identification strategy, allowing for unbiased estimation of causal effects, and as important dimensions of heterogeneity in those causal effects. Due to their importance as treatment effect moderators, they are notated separately from all other event-level variables, which we denote W_e . The variables in W_e (i.e., whether the event is organized over the weekend, is part of a recurring series, has a limit on the number of attendees, takes place outdoors, is hybrid, has COVID-related prerequisites, and the character length of the event description) serve as adjustment variables for the purpose of obtaining unconfounded causal estimates, but are not expected to be meaningful sources of treatment effect heterogeneity.

3.4 Data Summary

To conclude the overview of the data, Table 2 provides summary statistics for all relevant variables in the estimation data set, suggesting significant variation in event- and group-level variables.

4 Estimating Heterogeneous Effects of Event Digitization on Event Participation

We aim to estimate the causal effects of digitizing events, D_e , on members’ registrations to those events, Y_{ie} , conditional on particular group and event features. We focus on treatment effects among digitized events due to an asymmetry in how digitization relates to other event features. Whereas a digitized event can almost always be converted to an in-person format without substantially altering its core aspects, the reverse is not always possible. For example, converting a webinar to in-person might entail assembling attendees in one location and having the speaker present to a live audience, without any other changes to the format or content. By contrast, a hike through the woods cannot be digitized without fundamentally altering the nature of the activity and member participation. Hence, the causal effects we seek to estimate in the main analysis are conditional average treatment effects among the treated (CATT).

Nonparametric identification of these CATTs follows from a core set of causal assumptions. We first describe these core assumptions, then present the empirical strategy for estimating these treatment effects using the extensive and granular data collected from Meetup.

	Mean	SD	Min.	Median	Max.
<i>Events, Groups, and Members</i>					
Events per group	53	150	1	12	4 598
Members per group	157	272	1	68	6 243
Events per member	469	1 233	1	124	93 430
Groups per member	3.84	6.45	1	2	602
<i>Event Registrations</i>					
N. registrations	0.01	0.10	0	0	1
N. registrations per group	206	578	0	56	25 019
N. registrations per event	3.86	8.49	0	1	1 785
<i>Event Features</i>					
Online events	0.33	0.47	0	0	1
Social interactivity (Yes = 1, No = 0)	0.59	0.49	0	1	1
Specialized knowledge (Yes = 1, No = 0)	0.09	0.28	0	0	1
Skill development (Yes = 1, No = 0)	0.45	0.50	0	0	1
Physicality (Yes = 1, No = 0)	0.34	0.48	0	0	1
Mask required (Yes = 1, No = 0)	0.02	0.16	0	0	1
Vaccination required (Yes = 1, No = 0)	0.03	0.16	0	0	1
Outdoor venue (Yes = 1, No = 0)	0.01	0.10	0	0	1
Hybrid event (Yes = 1, No = 0)	0.001	0.03	0	0	1
Weekend event (Yes = 1, No = 0)	0.34	0.47	0	0	1
Event series (Yes = 1, No = 0)	0.28	0.45	0	0	1
Positive registrations limit (Present = 1, Absent = 0)	0.22	0.42	0	0	1
Description length (N. characters)	1 462.32	1 153.16	0	1 135	43 215
<i>Group Topics</i>					
Business/Careers	0.16	0.29	0	0.002	1
Community/Causes	0.08	0.16	0	0.001	1
Community/Culture	0.19	0.26	0	0.057	1
Dancing/Fitness	0.11	0.21	0	0.001	1
Socialization	0.15	0.24	0	0.003	1
Spirituality	0.14	0.27	0	0.001	1
Tech	0.13	0.26	0	0.001	1
Travel	0.04	0.11	0	0.001	1

Table 2: Summary Statistics. Notes: N. events = 645,013, N. groups = 12, 105, N. members = 493, 143.

4.1 Nonparametric Identification of Causal Effects

The core identifying assumptions comprise several assumed relationships among the observed and unobserved variables related to event digitization and member registration discussed in the previous section. Figure 3 depicts these assumed causal relationships as a directed acyclic graph (DAG; Pearl 2009).

From the perspective of an individual choosing to register for an event (whose registration decision is represented by the node labeled Y), the nodes labeled D , T , and X in Figure 3 represent the digitization status of the event (D_e), its timing ($t[e]$), and all other event features (X_e and W_e).⁵ All of these variables are observed by group members choosing whether to register for an event. The node labeled G represents

⁵We use X to denote both X_e and W_e in the DAG, as their distinction is not relevant for identifiability.

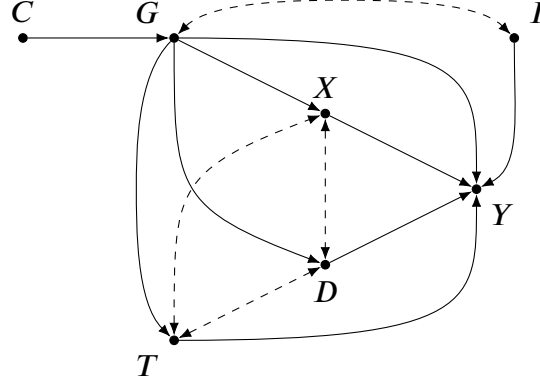


Figure 3: Directed Acyclic Graph Encoding Assumptions Leading to Unconfounded Non-Parametric Estimation of Digitization Effects..

the organizing group (g), C the group’s interest topics (C_g), and I the individual group member (i), who chooses whether to register for the event.

Arrows shown with solid arcs in Figure 3 represent the possibility of a direct causal effect between observed variables, and point in the direction of causality. Double-ended arrows with dashed arcs represent the possible presence of an unobserved common cause that has an influence on more than one observed variable. Unobserved variables that only impact a single observed variable are not shown, as their presence or absence is not consequential for identifiability. Importantly, the absence of an arrow pointing from one node to another represents an *assumed absence* of a direct causal effect. The DAG is a fully nonparametric depiction of causal assumptions. As such, all variables with arrows pointing to the same dependent variable can potentially interact in how they jointly affect it (i.e., interactions have no implied structure and can be of arbitrary complexity).

The arrow pointing from the node labeled D to the node labeled Y in Figure 3 represents the potential for an event’s digitization status, D_e , to have a direct causal effect on a user’s registration status for that event, Y_{ie} . Similarly, the arrow from X to Y allows for the possibility that other event characteristics, namely X_e and W_e , can also have direct causal effects on registrations. Other factors that could have direct causal effects on registrations are depicted with arrows pointing directly to the node labeled Y . These are (i) the group organizing the event, G , as some groups might organize events that are more or less attractive to members; (ii) the timing of events, T , which might cause some events to be more popular than others; and (iii) the individual member, I , as some people might be more likely to attend events (on average) compared to others.

The digitization status (D), timing (T), and other characteristics of an event (X) are all affected by the group organizing the event (G). Moreover, the dashed, double-ended arrows connecting D , T , and X indicate the possibility that one or more factors not observed in the data may have direct, causal effects on event digitization, timing, and other features. The arrow pointing from C to G , and the absence of an arrow from C to D or X , represents the assumptions that (i) the existence of topic interests drives the formation of shared interest communities, and (ii) groups fully mediate the relationship between interest topics and

events organized around shared interests. Finally, the dashed, double-ended arrow pointing to both G and I represents the unobserved processes that cause particular interest groups and individual users to use the Meetup platform to organize and learn about events, and for certain individuals to join some groups and not others.

An important characteristic of this setting, which might not be immediately apparent from the DAG in Figure 3, is that the focal treatment—digitization—is observed at the level of an event (e); whereas the focal outcome—a registration—is observed at the level of an event *and* group member, (i, e). Consequently, for every treated event, multiple instances of the outcome are observed, each corresponding to a different group member. Importantly, individual group members’ registrations to the same event can be correlated due to social interaction without negative consequences for identifiability. The key assumption that allows identifiability in the face of social interactions is that the choice to digitize any *particular event* be unaffected by the anticipated registration(s) from any *particular group member*, as reflected by the absence of a common unobserved factor affecting both D and Y .

Given these assumptions, conditional average treatment effects (CATEs) – conditional on group g , time t , and event features x – are nonparametrically identified (Pearl, 2009).⁶ These CATEs, which we denote by $\tau(x, g, t)$, represent an expected change in an individual group member’s registration from a hypothetical change in digitization while holding all other event and group features constant. Because these CATEs are nonparametrically identified, it follows that the conditional-on-topic and conditional-on-event feature CATTs of digitization are also nonparametrically identified (as these CATTs are just weighted averages of the CATEs, and the weights are also nonparametrically identified).

The conditional-on-topic CATT for each interest topic $c = 1, \dots, 8$ is

$$\text{CATT}(c) = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \tau(X_e, g, t[e]) \cdot n_e \cdot D_e \cdot C_{g,c}}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e \cdot C_{g,c}}, \quad (1)$$

where $C_{g,c}$ is group g ’s weighting for interest topic c . These CATTs represent weighted averages of expected changes in registrations among the members of each group g . Each group g ’s contribution to the overall CATT for interest topic c depends on the set of digitized events it organizes, $\{e \in \mathcal{E}_g | D_e = 1\}$; the size of its membership at the time of each event, $n_e = |J_e|$; and how strongly the group is related to topic c , $C_{g,c}$. The conditional-on-feature j CATTs are

$$\text{CATT}_j(x) = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \tau(X_e, g, t[e]) \cdot n_e \cdot D_e \cdot 1[X_{e,j} = x]}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e \cdot 1[X_{e,j} = x]}, \quad (2)$$

where $1[X_{e,j} = x]$ is 1 if the j feature of event e has a value of x , and 0 otherwise. We also report an overall

⁶This follows because the DAG implies $Y \perp\!\!\!\perp D | G, T, X$, as the set $\{G, T, X\}$ blocks all otherwise open backdoor paths connecting D with Y , and none of $\{G, T, X\}$ is a causal descendant of D along a causal chain starting at D and ending at Y or one of Y ’s causal descendants (Pearl, 2009).

ATT, which is a weighted average of the CATEs across all digitized events,

$$\text{ATT} = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \tau(X_e, g, t[e]) \cdot n_e \cdot D_e}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e}. \quad (3)$$

All of these treatment effects are expressed as the expected changes in an individual member’s registration resulting from digitizing an event. When reporting these effects, we multiply them by 100 and report them as percentages, making them easier to interpret in proportion to the size of a group’s membership. As an alternative, we also characterize ATTs in terms of the change in the number of attendees to an event that can be attributed to digitization; for any event e , this quantity is equal to $n_e \cdot \tau(X_e, g[e], t[e])$.

Although the identifiability of the CATEs (and thus the CATTs) follows directly from the assumptions encoded in the DAG, the identifying assumptions can also be expressed in terms of potential outcomes, which may be more familiar to some readers. Letting $Y_{ie}(d)$ denote the potential outcome for individual i ’s registration to event e with digitization status $D_e = d$, the CATEs can be expressed as expected differences in potential outcomes:

$$\tau(x, g, t) = E_{i,e}[Y_{ie}(1) - Y_{ie}(0)|x, g, t], \quad (4)$$

where the conditional expectation is taken with respect to group g ’s events (e) and members (i). In the potential outcomes framework, the main identifying assumption allowing consistent, nonparametric estimation of CATEs, is *conditional ignorability*:

$$Y(1), Y(0) \perp\!\!\!\perp D | X, G, T \quad (5)$$

This assumption says that conditional on an event’s timing, organizing group, and all other observed features, any remaining unobserved factors that affect digitization are independent of unobserved factors influencing how an individual would respond to the event, whether digitized ($Y_{ie}(1)$) or not ($Y_{ie}(0)$).

Finally, we note that the assumptions encoded in the DAG are also consistent with an alternative version of the conditional ignorability assumption. This alternative version conditions not only on X , G , and T , but also on the identity of each Meetup user I , and thus on each user’s average tendency to attend digitized or non-digitized events. Given the assumed absence of an unobserved confounder affecting both event digitization and an individual’s response, conditioning on each individual is not necessary for identification. However, conditioning on individual users can improve statistical efficiency and provide a degree of protection against model misspecification. For that reason, we include individual group members in the set of conditioning variables (i.e., controls) that form the basis of the estimation procedure.

4.2 Empirical Strategy

Next, we describe the approach to treatment effect estimation. First, we estimate CATEs using a two-stage, doubly-robust regression estimator (Bang & Robins, 2005; Gordon et al., 2019), and then take weighted averages of the estimated CATEs to obtain estimated CATTs.

4.2.1 First Stage Propensity Score Model

The first stage in the two-stage estimator uses logistic regression to estimate the propensity of event digitization (D_e) conditional on other event features (X_e and W_e) and fixed effects representing the week of the event (t) and the organizing group (g).

$$\hat{\pi}_e \equiv E[D_e | \text{---}] = \text{logit}^{-1} (X_e' \theta^X + W_e' \theta^W + \theta_t^T + \theta_g^G) \quad (6)$$

The estimated propensity scores, $\hat{\pi}_e$, are then used to construct inverse propensity weights (IPW) for CATEs.

$$\phi_e \equiv D_e / \hat{\pi}_e + (1 - D_e) / (1 - \hat{\pi}_e) \quad (7)$$

Web Appendix D shows that both treated and control counts exist in all propensity score bins. We next assess whether the distributions of covariates, conditional on the fixed effects in Equation (6), differ among treated and control units. Any discrepancies would point to potential issues with the estimated CATTs.

A common recommendation for assessing balance is to compute the difference in the standardized means (SMDs) of each covariate between treatment and control units over different propensity score strata. This measures the dissimilarity between the treated and control events in terms of their covariates, among subsets of events with the same propensity for digitization. The literature recommends setting a threshold between 0.1 and 0.25 to determine whether the absolute value of SMDs is too high, and to compute absolute SMDs for both covariates and their simple interactions (Austin, 2009).

Web Appendix D shows that the absolute SMDs for most covariates and covariate interactions score below the 0.1 threshold across propensity score bins, and all absolute SMDs are below the 0.25 threshold, providing evidence for covariate balance across treatment and control units. The only term equal to the recommended threshold is the interaction term *Physicality* \times *Socialization* ($|\text{SMD}_{\text{Bucket } 6}| = 0.25$), which we expect to correlate strongly with digitization status. Although we cannot rule out the influence of additional omitted variables at this stage, these diagnostics lend plausibility to our proposed identification strategy, which leverages highly granular data, adjusts for covariates, interactions, and fixed effects, and relies on a doubly robust estimator.

4.2.2 Doubly-Robust Second Stage

The second stage regression in the two-stage estimator incorporates the estimated IPWs not as observation weights, as in a Horvitz-Thompson estimator, but rather as an additional covariate for doubly-robust adjustment.⁷ We specify the second stage regression as a linear probability model, with effects of digitization on registrations that are heterogeneous depending on a subset of theory-relevant event features, X_e , and groups' interest topics, C_g , as well as several control variables, including the estimated IPWs, ϕ_e .

$$E[Y_{ie} | \text{---}] = D_e (X_e' \beta^{DX} + C_g' \beta^{DC}) + X_e' \Gamma^{XC} C_g + W_e' \gamma^W + \phi_e \delta + \eta_g + \eta_t + \eta_i \quad (8)$$

⁷This is called a “clever covariate” in Bang and Robins (2005) and “Inverse-Probability-Weighted Regression Adjustment” in Gordon et al. (2019)

The coefficients β^{DX} and β^{DC} in Equation (8) allow for heterogeneity in treatment effects across events that differ in their features X_e and/or the organizing groups' interest topics C_g . The remaining terms control for factors correlated with levels of either digitization or registrations. These control variables include the additional event features W_e , the group identifier g , the week of the event t , the individual member i , the first-stage IP weights ϕ_e , and linear interactions between event features and interest topic weights. The group-level fixed effects η_g control for stable, group-specific characteristics, such as the group organizers' preferences and the stable, underlying demographic characteristics of the group's membership. The weekly fixed effects η_t account for market-level temporal shocks (e.g., technology trends). The individual fixed effects η_i capture individual-level idiosyncrasies, in particular, each member's baseline propensity to participate in Meetup events.

The regression equation models treatment effects that potentially vary across events with different features and across groups catering to different interest topics, and the contributions to treatment effects from these separate sources of heterogeneity are assumed to be additively separable. Both of these modeling choices place restrictions on higher-order treatment effect heterogeneity. Both are motivated by (i) having a finite sample, and (ii) a desire not to overfit the data.

The first and second stage regressions are estimated using PySpark's `MLlib` regression modules, with g , t , and i specified as high-dimensional fixed effects. Conditional on feature (x) and group (g) CATE estimates, $\hat{\tau}(x, g)$, are then obtained from the coefficient estimates. This is accomplished by predicting registrations to events with and without digitization and taking their difference:

$$\hat{\tau}(x, g) \equiv \hat{E}[Y_{ie}(1) - Y_{ie}(0)|x, g] = x' \hat{\beta}^{DX} + C_g' \hat{\beta}^{DC} \quad (9)$$

The CATT for the interest topic c is then obtained as a weighted average over the empirical distribution of groups and events,

$$\widehat{\text{CATT}}(c) = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \hat{\tau}(X_e, g) \cdot n_e \cdot D_e \cdot C_{g,c}}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e \cdot C_{g,c}}. \quad (10)$$

Similarly, the CATT for events sharing the same value x for event feature j is⁸

$$\widehat{\text{CATT}}_j(x) = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \hat{\tau}(X_e, g) \cdot n_e \cdot D_e \cdot 1[X_{e,j} = x]}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e \cdot 1[X_{e,j} = x]}. \quad (11)$$

Last, the overall ATT is

$$\widehat{\text{ATT}} = \frac{\sum_g \sum_{e \in \mathcal{E}_g} \hat{\tau}(X_e, g) \cdot n_e \cdot D_e}{\sum_g \sum_{e \in \mathcal{E}_g} n_e \cdot D_e}. \quad (12)$$

⁸We also augment the main analysis by calculating CATTs for subsets of events defined by more than one event feature and/or interest topic.

4.2.3 Monthly Second-Stage Regressions and CATTs

The regression defined by Equation (8) does not include a time-varying term related to digitization. This amounts to an assumption—consistent with much of the prior literature—that although time can jointly affect *levels* of registrations and event digitization, it does not contribute to variation in *treatment effects*. Said differently, the main specification assumes that $E[Y_{ie}(1) - Y_{ie}(0)]$ does not vary systematically over time. Here we describe a relaxation of this assumption that allows digitization effects to potentially vary by month.

In addition to the main specification, we estimate separate models from subsets of the data corresponding to each month of 2024.⁹ The monthly models are structurally identical to those described previously, maintaining the same doubly-robust estimation approach, control variables, fixed effects structure, and heterogeneity dimensions. Because each monthly model replicates the main estimation procedure exactly, the identifying assumptions, if valid in the overall data, should also be valid within each monthly subset. This approach enables us to investigate potential temporal heterogeneity in digitization effects and assess the stability or variation of these effects throughout the year. For each month-specific model, we also calculate CATTs using identical procedures and weighting schemes to those already described, facilitating a consistent interpretation of digitization effects over time.

In Web Appendix E.3, we obtain standard errors for a relevant subset of monthly treatment effects by (i) simulating 30 coefficient estimates from a normal distribution centered on the point estimate, with standard errors obtained from the monthly regression models; (ii) calculating the relevant treatment effects conditional on the simulated coefficients; and (iii) quantifying uncertainty in the simulated treatment effect predictions. Owing to the large sample sizes, the resulting uncertainty bounds for treatment effects are very narrow compared to differences across values of covariates. Hence, we focus on point estimates when reporting all estimated treatment effects.

4.3 Robustness Checks

We assess the robustness of the model specification in Equation (8) in two ways. First, we perform an analysis of the model’s sensitivity to the inclusion of unobserved confounders, as described in Cinelli and Hazlett 2020. This robustness analysis calculates the minimum strength of association that unobserved confounders would need to have in order to alter the qualitative conclusions of the main analysis. Second, we conduct a “falsification test” in which the digitization treatment is randomly assigned to individuals (Haw et al., 2023), to rule out the possibility that the effects are driven by idiosyncratic factors and that any significant results may be spurious. Online Appendices G.1 and G.2 provide more details on these additional analyses and an overview of their conclusions, which support our proposed identification arguments.

⁹We use the `fixest` R package (Bergé, 2018) to estimate the second-stage regressions, which is faster than `PyS-park` when working with smaller data sets.

5 Results

This section presents estimates of the heterogeneous effects of digitization on participation in digitized events. We first highlight the importance of the eight interest topics and four theory-relevant event characteristics as determinants of treatment effect heterogeneity. We then provide insights into the stability of these effects throughout 2024. We conclude the section with additional exploratory analyses that connect our findings with opportunities for firms and policy-makers seeking to further enable digital transformation.

5.1 Digitization Effects

We use the estimated regression coefficients to derive the estimated conditional average treatment effects (CATE; Equation (9)). Web Appendix E reports the coefficient estimates obtained from Equation (8)). From these, we then calculate the various treatment effects described in Equations (10)–(12). These conditional treatment effects represent the incremental likelihood of individual i participating in particular types of digitized events due to digitization. The unconditional average effect of digitization among all digitized events in 2024 (ATT; Equation (12)) is a 0.05% increase in registered attendees. This estimate of 0.05% represents an absolute increase in registrations in proportion to the number of members in a typical group (as opposed to a proportional increase in the average attendance rate).

Although the magnitude of this ATT appears small, there are two important considerations to bear in mind. The first is that a typical digitized event in 2024 was organized by a group with 286 members and had 3.55 people register to attend. Among these average 3.55 positive registrations, 0.13 can be attributed to the digitization of the event, approximately 3.7% of those attending. Due to the size of the platform, these increases in attendance scale such that, among the 215,452 digitized events organized in 2024, there was a net gain of 67,430 attendees resulting from digitization. The second consideration is that this net change in registrations reflects a mix of increases and decreases in registrations across all events. In particular, the net gain of 67,430 attendees is composed of an increase of 141,441 attendees among digitized events that benefited from digitization and a decrease of 74,011 among events that were harmed (in terms of registrations) by digitization. Next, we elaborate on the drivers of these gains and losses.

5.1.1 Group topics

The topics covered by the organizing group stand out as a significant source of heterogeneity in digitization effects. Figure 4a depicts the conditional average treatment effects among digitized events weighted by the organizing group’s interest topic loadings (CATT; Equation (10)). Figure 4a highlights that there are large subsets of events for which digitization leads to lower attendance. In particular, digitized events organized by groups catering to the interest topics *Tech* and *Business/Careers* can attribute decreases in attendance of -0.19% and -0.05% , respectively, to digitization. In contrast, the effects of digitization are positive for events organized by groups catering to the interest topics *Spirituality* (0.21%), *Travel* (0.14%), *Community/Culture* (0.12%), *Community/Causes* (0.13%), and *Dancing/Fitness* (0.13%).

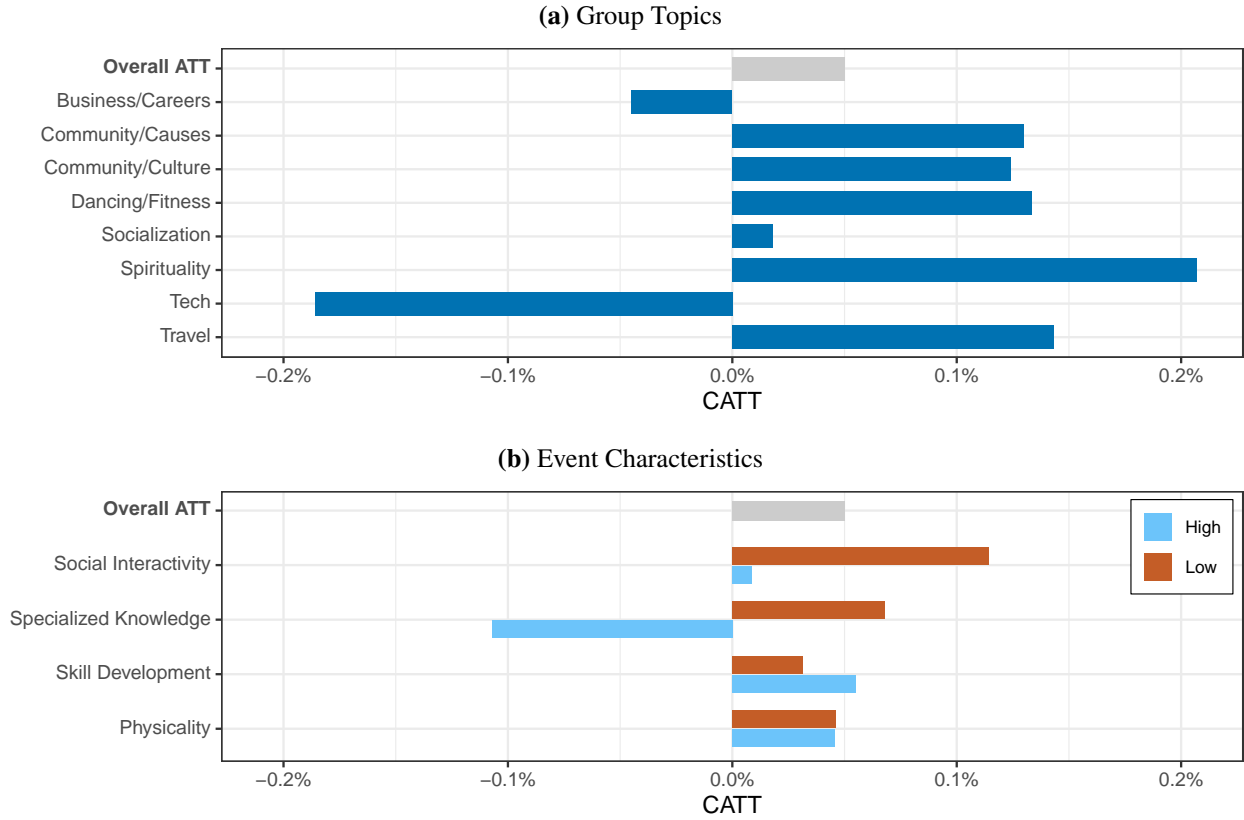


Figure 4: CATTs by Group Topics and Event Characteristics. Notes: CATTs refer to the causal effects of digitizing events among events that were digitized, conditional on particular a group topic (a) or event characteristic (b). The overall ATT is shown for reference.

5.1.2 Event characteristics

The theory-relevant event characteristics are also important determinants of how digitization affects event registrations. Figure 4b depicts conditional average treatment effects among subsets of digitized events sharing a common value of each of these binary variables (Equation (11)). Digitization has a negative effect on events requiring a higher degree of *specialized knowledge* (-0.11%). In contrast, digitization has a positive effect on events that place greater emphasis on *social interactivity* (0.11%) or *skill development* (0.06%), or require less *specialized knowledge* (0.07%). The relatively positive effect measured for events that emphasize *skill development* is consistent with recent results focusing exclusively on these types of events (e.g., studies of the digitization of educational activities in Bonsall 2023; Kofoed et al. 2024; Sconti 2022).

5.1.3 Combinations of event characteristics

The different digitization effects described thus far consider single-interest topics or theory-relevant variables in isolation. However, there are a substantial number of events with particular combinations of group topics and theory-relevant characteristics for which we can estimate digitization effects and provide new insights.

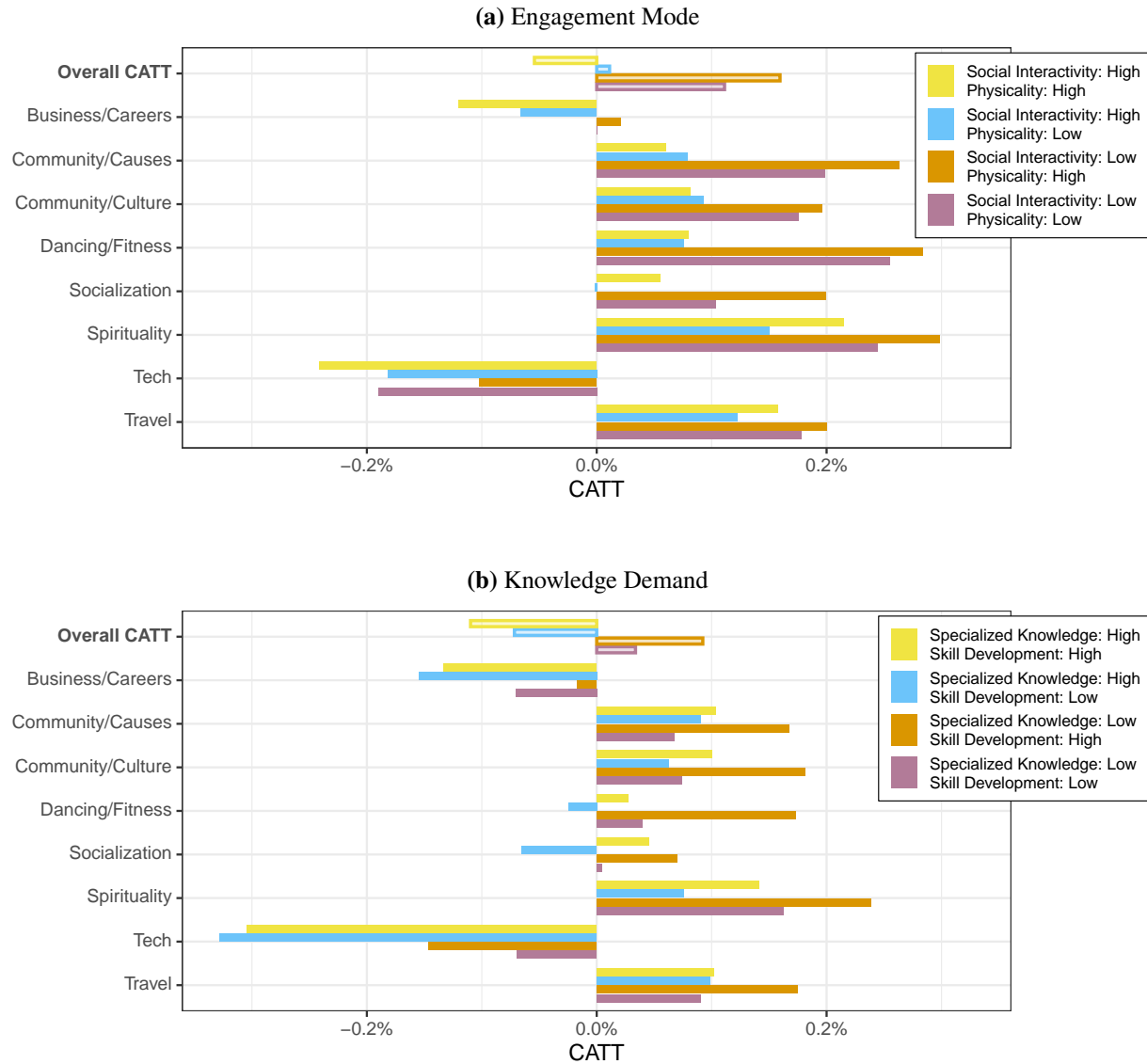


Figure 5: CATTs by Group Interest Topics and (a) Engagement Mode or (b) Knowledge Demand. Notes: Engagement mode is a combination of *social interactivity* and *physicality*; knowledge demand is a combination of *specialized knowledge* and *skill development*.

We therefore consider two higher-order pairings of event characteristics. The first characterizes the *engagement mode* of an event. This is defined by pairs of values of *physicality* and *social interactivity*, leading to the four combinations of these variables shown in Figure 5a. The second pairing of event characteristics represents the *knowledge demand* of the event, and is defined by values of *skill development* and *specialized knowledge*. These are shown in Figure 5b.

Focusing first on engagement mode, Figure 5a shows that the effect of digitization can vary significantly among events that all share a high degree of *physicality*, but differ in their level of *social interactivity*. Specif-

ically, digitization is especially beneficial for high-*physicality* activities that place less emphasis on *social interactivity*. Events of this type were organized by 372 groups and include online workshops on writing fiction, online workshops on sensory experiences involving food, livestream prayer events, and remote psychotherapy exercises. In contrast, high-*physicality* events that also emphasize *social interactivity* receive fewer registrations when digitized. Examples of such events include “hands-on and interactive live demos of AI applications”, interactive sessions for backpackers to prepare for outdoor trips where “breakfast, lunch, dinner, and snacks will be shown”, and events where attendees “make art, have engaging discussions, relax, and foster friendships through art”.

Figure 5b shows a similar contrast among events that all emphasize *skill development*, but which vary in the level of *specialized knowledge* required to participate. Specifically, digitization is beneficial for high *skill development* events that do not require *specialized knowledge* (i.e., suitable for beginners), but detrimental for those that presume *specialized knowledge*. These patterns suggest that digitization may be well-suited for introductory-level courses and workshops, but less effective (or even harmful) for courses requiring more advanced or specialized knowledge. One potential explanation for these results can be found in “media richness” theory, which posits that different communication sources (i.e., an in-person meeting or an online workshop) vary in their ability to transmit cues (e.g., facial expressions, text, and visual information) and in their effectiveness in conveying specialized knowledge. Moreover, “richer media” such as face-to-face communication can transmit many cues simultaneously (e.g., gestures and subtle facial expressions). Hence, compared to “poorer media” (e.g., a virtual presentation), richer media should be more effective when conveying information that may be unclear or confusing (Grewal et al., 2021). A complementary explanation could be that in-person and virtual communication differ in terms of input modalities (i.e., the combination of visual, olfactory, auditory, gustatory, and haptic systems), which potentially impairs the processing of more advanced or specialized information in virtual settings (Lim & Benbasat, 2002).

The positive effect of digitization on high *skill development* events that require little or no *specialized knowledge* may also help to reconcile some of the seemingly contradictory results in the literature on the digitization of education. For example, digitization has been found to lead to better academic outcomes among secondary school children (Gortazar et al., 2024), but worse outcomes among university students (Kofoed et al., 2024). A possible contributor to these discrepant results may lie in the amount of prior specialized knowledge required by the students.

5.1.4 Combinations of event characteristics and group topics

Given the importance of group interest topics in explaining treatment effect heterogeneity, Figure 5 also shows the CATTs for *engagement mode* and *knowledge demand* weighted by groups’ interest topic loadings. The average digitization effects by topic illustrated in Figure 4 are generally consistent across engagement modes. Only events organized by groups primarily catering to the topic of *Business/Careers* exhibit a large difference in CATTs depending on the engagement mode. *Business/Careers* events experience a -0.12% decrease in registrations when they emphasize both *physicality* and *social interactivity*. In contrast, digitization has no impact (0.001%) on registrations for *Business/Careers* events that place less emphasis on these

engagement modes.¹⁰ These results suggest that digitization is not detrimental to events organized by groups catering to *Business/Careers* topics per se, as registrations for *Business/Careers* events with low *social interactivity* are, at worst, unaffected by digitization. However, when *Business/Careers* events involve a high degree of *social interactivity*, in-person formats may be the best option for increasing registrations.

Turning to *knowledge demand*, Figure 5b shows that events organized by groups catering to *Socialization*, and requiring a higher degree of *specialized knowledge*, but at the same time *not* emphasizing *skill development*, exhibit a distinctly negative CATT of digitization (the CATT for *Socialization* is, on average, small and positive). Events of this type include online matchmaking and speed dating events for “intellectually-oriented single professionals with advanced degrees,” and an online “beginner backpacking collaborative leadership discussion” about common allergies and the management of health-related situations. The requirement of *specialized knowledge* also implies negative CATTs for events organized in *Tech*, *Dancing/Fitness*, and *Business/Careers* topics, especially when the event does not have a *skill development* goal.

5.1.5 Robustness Checks

The results of the robustness checks in Web Appendix G show that the digitization effects presented thus far are robust to the inclusion of unobserved confounders and present no evidence of systematic bias due to uncontrolled, idiosyncratic factors. In particular, the results of a sensitivity analysis suggest that an unobserved confounder would need to be, at minimum, 10 times stronger than the strongest covariate estimate (for the “weekend event” indicator) in order to alter conclusions qualitatively. Furthermore, a falsification test rules out the possibility that the effects are driven by idiosyncratic factors or random variation alone (Web Appendix G).

5.2 Time-varying Effects

Next, we assess the stability of topic- and feature-level digitization effects throughout 2024. As mentioned, we augment the main model specification to account for digitization effects that possibly vary by month. Figure 6 shows that digitization effects can vary significantly over time. This is most evident for events organized by groups catering to the interest topics of *Tech*, *Travel*, and *Business/Careers* (Figure 6a); and for events requiring *specialized knowledge* (Figure 6b). For other group topics (e.g., *Socialization*, *Dancing/Fitness*, and *Community*-related topics) and event characteristics (e.g., *skill development*), digitization effects do not appear to vary much over the course of 2024.

Events organized by groups catering to the *Tech* interest topic are especially notable. Whereas the CATT for *Tech* events is negative on average (−0.19%), the monthly CATTs are positive in January (0.29%), August

¹⁰The null effect for *Business/Careers* events with no *physicality* and no *social interactivity* is not due to the lack of such events. In 2024, 1,223 groups with above-average *Business/Careers* topic loadings organized 23,401 digitized events with no emphasis on *social interactivity* or *physicality*. In contrast, 904 digitized events organized by groups with above-average *Business/Careers* loadings do emphasize *social interactivity* and *physicality*. We can also rule out topic imbalance as an alternative explanation, as the organizing groups’ loadings on the *Business/Careers* topic are not dissimilar across events with more (versus less) emphasis on *physicality* and *social interactivity* (t-statistic = 0.29, p-value = 0.76).

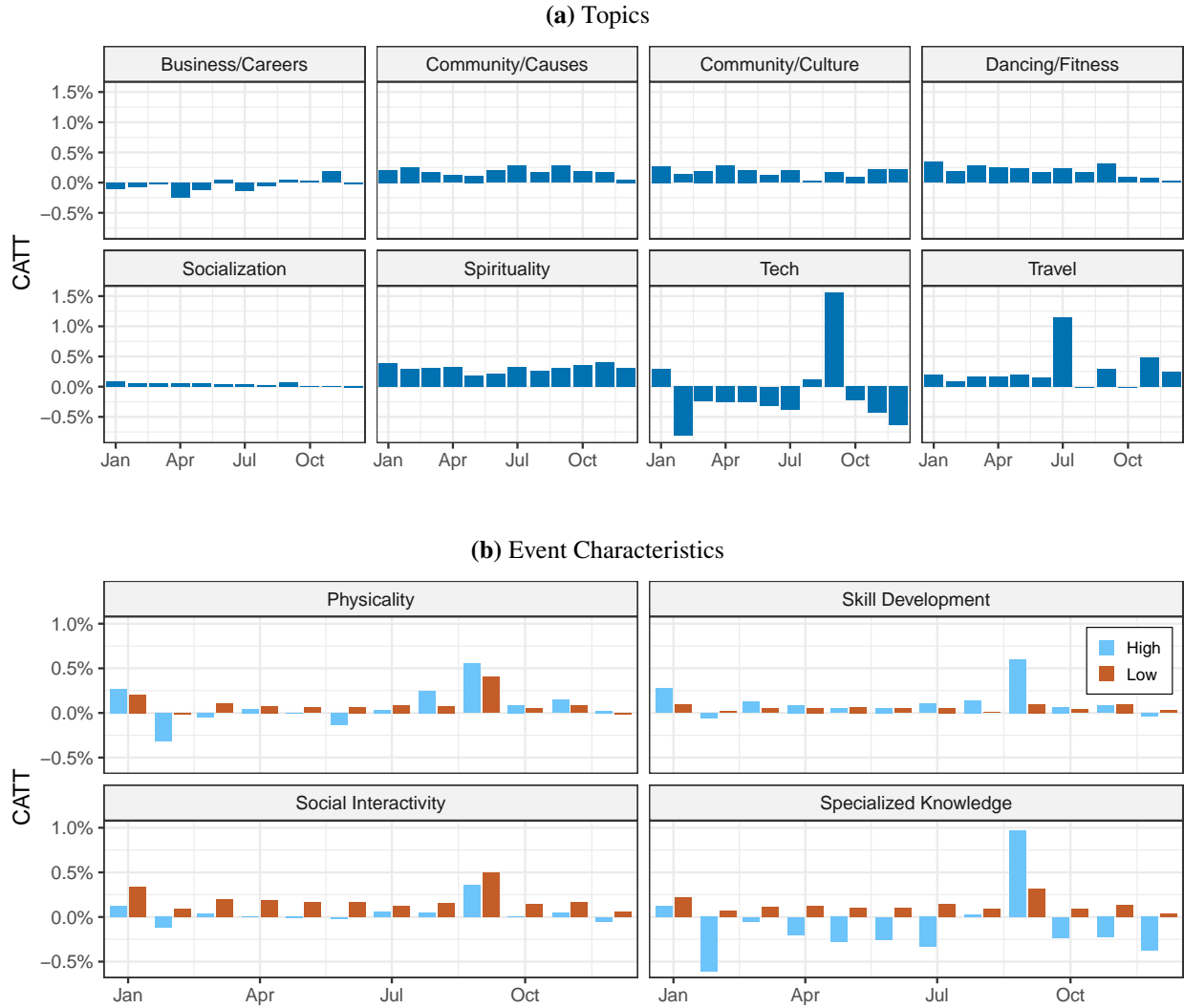


Figure 6: Month-Specific CATTs by (a) Topic and (b) Event Characteristic.

(0.13%), and September 2024 (1.55%). Similarly, digitization is positive for *Business/Careers* events in June (0.05%) and between September and November (0.04%–0.18%). The CATTs for events organized by groups catering to the *Travel* interest topic are mostly positive throughout 2024, with noticeable peaks in July (1.14%) and November (0.45%).

Comparing Figures 6a and 6b, a common pattern emerges in how CATTs vary over time for two types of events: those requiring *specialized knowledge*, and those organized by groups in the *Tech* category. Both are the most positive in September and the most negative in February. Figure 7a considers just the subset of events organized under by *Tech* groups and requiring a high degree of prior, *specialized knowledge*, and shows the same time-varying pattern with spikes in February and September. One potential explanation for this pattern is that *Tech* events are both more likely to require *specialized knowledge*, and more (less) or prevalent in September (February). However, Figure 7b shows that the monthly share of digitized events

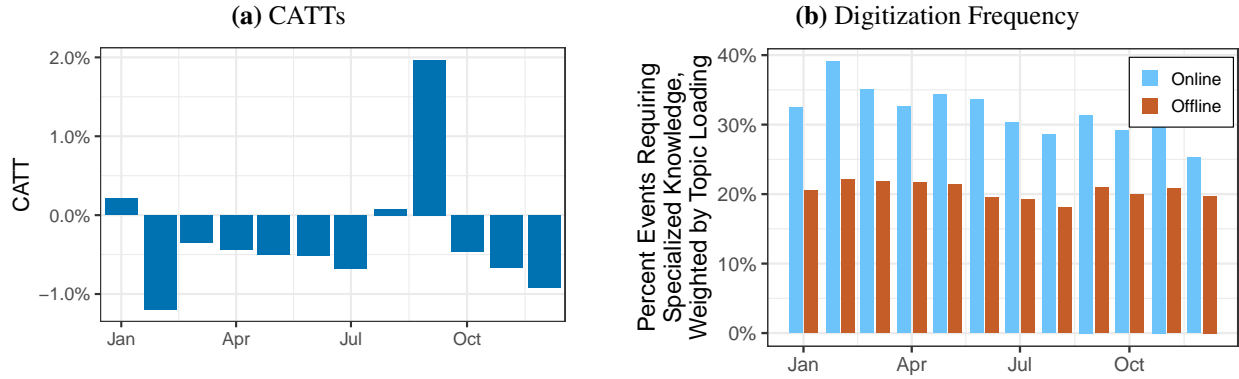


Figure 7: Monthly (a) CATTs and (b) Digitization Frequencies for Tech Events Requiring *Specialized Knowledge*. Notes: CATTs are weighted by the organizing group’s topic loading for *Tech*. See Web Appendix E for a visualization of the other topics.

that require *specialized knowledge* in *Tech* (weighted by the topic loadings of the corresponding events) is rather stable throughout the year. In other words, it is unlikely that a disproportionate incidence of digitized high-specialization events in February or September could fully explain the time-varying results.

The results of these analyses show that not only can the magnitude and direction of digitization effects vary with the characteristics of the events and the groups that organize them, but that they can also vary significantly over short periods. Current digitization strategies tend to account for the first two factors in isolation, suggesting, for example, that digitization is positive for activities organized around education and skill development (Lee et al., 2023), or that workers prefer in-person workplaces when they can socialize with their colleagues (Capossela, 2022). In light of the complex heterogeneity of digitization effects we measure, this focus on isolated factors and one-size-fits-all answers to digitization questions could be a reason for the sometimes fragmented and contradictory evidence on digitization effects currently available to practitioners. Instead, organizers of digitized events, as well as researchers studying digitization, should consider the interplay between their community, the events they organize, and even timing when evaluating digital transformation.

5.3 Digitization Effects and Opportunities for Digital Transformation

Based on the results just presented, we next identify several potential opportunities for digital transformation, with the goal of providing preliminary guidance to firms and policy-makers involved in the digital transformation process about how to successfully digitize an increasingly diverse range of activities. In particular, we discuss the common characteristics of groups and events with negative measured digitization effects, and the common characteristics of groups and events with low digitization rates, both of which we view as growth opportunities for event organizers and companies that facilitate digitization.

5.3.1 Growth opportunities for event organizers.

The first growth opportunity for organizers lies in the digitization of events with negative CATTs. As Figure 4 shows, digitization is negative for *Tech* events (−0.19%), followed by events requiring a high degree of *specialized knowledge* (−0.11%), and *Business/Careers* events (−0.05%). To offset these negative effects, event organizers can focus efforts on promoting these events to a wider group of potential attendees, improving technical aspects of how these events are delivered, and addressing the “pain points” that attendees report during technical and/or highly interactive digitized experiences.

The second growth opportunity arises from the observation that the relatively high degree of physicality associated with particular social experiences is not necessarily an obstacle to successful digital transformation. As Figure 4b shows, the level of *physicality* does not have a meaningful impact on CATTs. For instance, Table W3 shows that *Dancing/Fitness* groups organize the greatest proportion of events with a high degree of *physicality*, whereas Figure 5a shows that digitization is positive for such events. This could be because *Dancing/Fitness* groups excel at organizing digitized classes. At the same time, digitization is relatively negative for highly physical events that also emphasize *social interactivity*. This can be an important insight for event organizers. Following the COVID-19 pandemic, many highly physical activities, such as remote medical examinations, have been successfully digitized. Identifying and building on similar success cases, and informed by our results, organizers may wish to focus on improving the delivery of digitized physical experiences that also emphasize *social interactivity*, thereby leveraging an already existing interest in these types of activities to improve registration rates.

These growth opportunities also need to be evaluated in light of cost-related considerations. We began by noting that digitized events “are often more affordable, accessible, and efficient than in-person alternatives” for organizers and prospective participants. Indeed, industry reports indicate that in-person events can cost over 47% more than their digitized counterparts, and digitized events can save 60–75% of the costs associated with venue, travel, and on-site expenses (Uhlhorn, 2025). However, our results suggest that the financial advantages for organizers must be evaluated in context with the substantial heterogeneity in digitization effects. For instance, the results indicate that digitization generally increases participation in events focused on *Spirituality*, *Community/Culture*, and *skill development*, which may enhance the cost-saving opportunities associated with digitized formats. In contrast, digitization tends to negatively impact participation in *Business/Careers* and *Tech* events, as well as those emphasizing *specialized knowledge* and an interactive engagement mode, which may instead justify continued investment in more costly in-person formats. Ultimately, organizers should optimize their digitization strategies and budgets by aligning event formats with (i) their intended audiences, (ii) the objectives of the specific event, and (iii) the characteristics of the organization as a whole, balancing economic efficiency with digitization effectiveness.

5.3.2 Growth opportunities for companies that facilitate digitization.

Following the previous discussion regarding event organizers, companies that seek to enable a broader range of digital experiences can also focus on enhancing their products to better enable events requiring *specialized*

knowledge, physicality, and/or social interactivity. And as mentioned, events organized around technology and business interests appear to struggle with the adoption of digitized formats (and even when they adopt digitized formats, the effects on engagement are often negative). These types of events may have novel use cases that are not well supported by existing tools. Understanding these specialized use cases may provide the impetus for successful product innovation.

Our results also suggest a potential “disconnect” between the supply and demand for digital transformation. Figure 8 reports the share of digitized events across different engagement modes, knowledge demands, and interest topics. It shows that organizers do not often digitize highly physical events or events focused on *skill development*, such as DIY events, cooking sessions, art projects, walks, runs, and eating and drinking. Furthermore, the digitization rate is consistently lowest among events catering to *Socialization, Dancing/Fitness, Community/Causes, and Travel* interests. Yet, despite their lower prevalence, these digitized events were not necessarily unsuccessful. For example, very few *Travel* events that emphasize both *physicality* and *social interactivity* were digitized, yet these events have positive CATTs (Figure 5a).

Such “disconnects” in the context of digital transformation are at the heart of important policy discussions. For example, business leaders, academics, and policy-makers are concerned with the “great mismatch” resulting from diverging preferences for digitization between employers and employees (Robinson, 2023). The disconnect between supply and demand we observe points to a similar potential mismatch between the organizers and attendees of social events. Possible explanations for the observed patterns could align with theories on technology acceptance (Davis, 1989). For instance, COVID-19 shocks could have increased the perceived ease of use of digitization platforms, even for physical and socially interactive events. In turn, heterogeneity in the perceived usefulness of digitization across diverse events and topics might have led to differential adoption and participation. Alternatively, the patterns might arise based on inaccurate beliefs about the effectiveness of digitized formats for certain combinations of event features. The data do not allow us to draw definitive conclusions about the mechanisms, but nevertheless raise important questions that closely align with current policy debates.

6 General Discussion and Conclusions

6.1 Digitization Effects

Digitization facilitates the organization of diverse events and activities, enabling the creation of social connections in innovative ways. However, the potential for digitization to foster engagement is not uniform across diverse activities, and the available evidence on the effect of digitization on social engagement is fragmented and sometimes contradictory. As business leaders and policy-makers evaluate the often complex implications of digitizing local communities, education, and workplaces, the essential question is whether this hinders people’s engagement in their communities. But even more precisely: Is digitization always a threat or an enhancement to social and community participation? Does its impact vary across categories of interest and types of events? Are digitization effects stable over time, providing insights into human behavior,

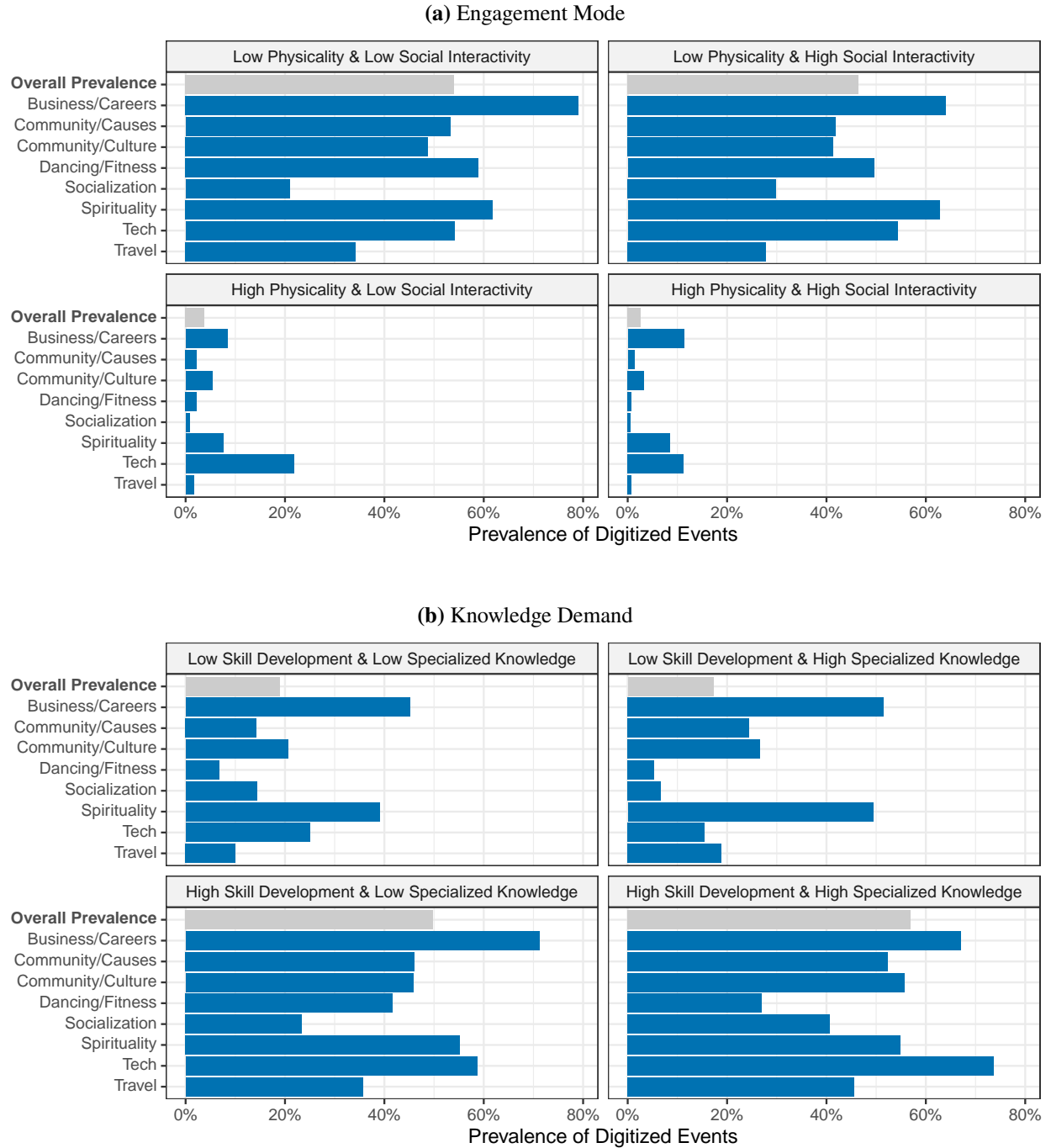


Figure 8: Digitization Frequency by Topic and Event Characteristics.

or are they transient, providing insights into current events?

To provide a comprehensive perspective on these issues, we investigate the heterogeneous effect of digitization on participation in thousands of communities and hundreds of thousands of events organized throughout 2024. We leverage a unique and highly detailed data set collected from the digital platform Meetup.

We exploit the granularity and attributes of both structured and unstructured data at the group, event, and individual levels in order to measure digitization effects that vary by topic, event features, and time.

We find that digitization has a meaningful, positive impact on engagement in social activities. Throughout 2024, digitized events gained an additional 141,441 registrations due to digitization and lost 74,011, resulting in a net gain of 67,430 prospective attendees and an average incremental gain of 0.13 registrations per digitized event. However, the results also suggest that digitization effects are highly heterogeneous. For example, digitization has a negative impact on participation in highly specialized and technical events, and events that cater to *Tech* and *Business/Careers* topics. Digitization effects also vary over time, sometimes rather dramatically, as in the case of *Tech* and *Business/Careers* events. Owing to the highly granular treatment effects we are able to recover, we suggest that the high degree of heterogeneity across topics, event characteristics, and even time can help reconcile some of the fragmented and contradictory evidence on digitization effects currently available to practitioners. Based on these results, we identify the main areas of opportunity and improvement for companies and vendors that enable the digital transformation of the economy, and suggest that the most impactful improvements to digitization technologies lie in the delivery of specialized knowledge and performance of physical tasks. Finally, we highlight the key managerial and policy implications of our results, particularly the emergence of a “mismatch” between the supply and demand for digital transformation in local communities (similar to the “great mismatch” discussed in the context of remote work, Robinson 2023).

6.2 Future Research

Our findings can be further extended by future research in various ways. First, our study measures participation intentions instead of event attendance due to data limitations. Event registrations are a widely used KPI to measure event success in the industry and correlate positively with actual attendance. However, industry reports suggest that between 30% and 50% of attendees who register for in-person and online events do not ultimately attend the activities. This intention-behavior gap would imply that, of the additional 67,430 net registrations attributed to digitization, between 33,715 and 47,201 would actually attend the events. Although this participation range is still economically significant, future studies might consider estimating the effects of digitization on actual attendance to overcome this limitation.

We also acknowledge that it is not possible to control for all potential confounders without running field experiments. We address this limitation to the best of our ability by combining extensive and granular data, theory-driven identification arguments grounded in the empirical context, doubly robust econometric modeling approaches, and robustness checks that directly address concerns about omitted variables. With these safeguards in place, we believe that the qualitative takeaways we present should not be driven by uncontrolled factors. Although large-scale field experimentation across diverse communities could mitigate these concerns, we encourage future research in field settings to carefully evaluate the problem of limiting participants’ access to a given event format. These issues are especially salient in the aftermath of a global pandemic, as preventing attendees from accessing a virtual event might cross an ethical boundary. Such considerations are especially important for event categories that generate substantial socioeconomic value

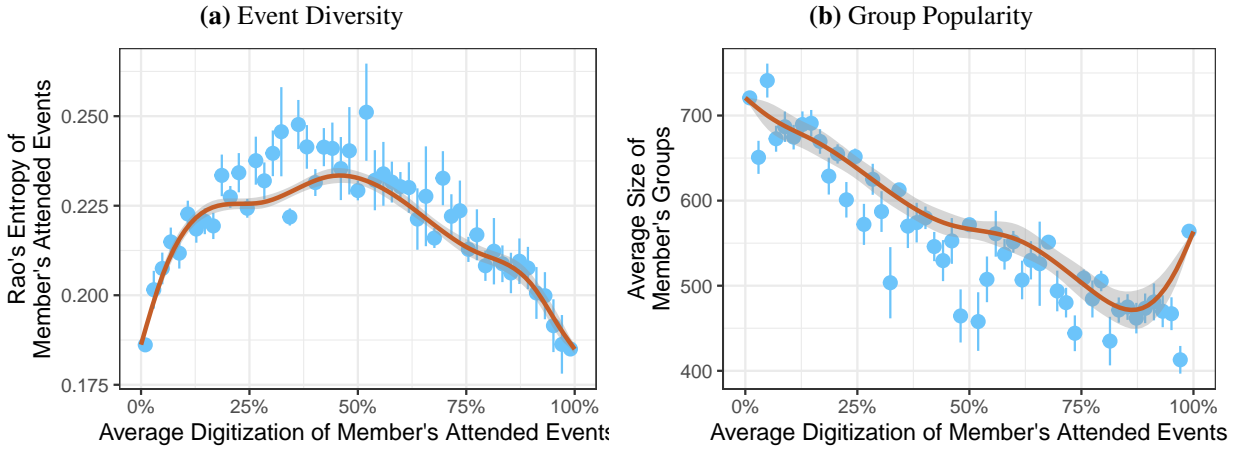


Figure 9: Average Digitization of Member's Attended Events, and Diversity and Popularity of Events and Groups.

for prospective attendees, such as career networking and social support groups. Excluding participants from a (counterfactually) higher value-generating event could seriously affect their personal and financial well-being. Conditional on resolving these ethical issues, field experiments could provide valuable insights into the boundaries and behavioral mechanisms of the digitization effects we measured.

Throughout the paper, we have focused on providing managerial insights for digitization firms, policy-makers, and organizers. However, previous research has shown that digitization could also directly benefit individual consumers. For instance, digitization promotes the adoption of more diverse goods (e.g., digital books allow consumers to read and enjoy more diverse genres), and access to “long-tail” products and services (e.g., the digitization of music allows for the discovery of relatively unknown or unpopular artists, Goldfarb and Tucker 2019). The increasing digitization of social events could create similar opportunities. Due to lower search and discovery costs, people who participate in more digitized events may also have better opportunities to join more diverse and niche groups and events (Goldfarb & Tucker, 2019).

Although the causal effects we estimate do not directly address this question, the granularity and richness of our data can provide initial evidence on whether diverse and “long-tail” consumption, which is prevalent in digital consumer markets, is also relevant for digitized social events. We briefly outline an exploratory analysis that we hope will motivate future investigation (see Web Appendix F for details on the metrics calculated for this analysis). First, we observe an inverse U-shaped relationship between digitization and participation in more diverse events. Figure 9a shows that members who participate in both digitized and in-person events tend to participate in a more diverse range of events. A similar, but weaker relationship holds for groups (Web Appendix F). This might suggest that—contra the typical long tail of digital consumption—members who prefer more digitization do not necessarily prefer proportionally more variety in social events. Instead, it might be that consumers who seek variety in digitization formats also seek variety in their communities.

In terms of participation in niche events and groups, and using group size as a proxy for popularity, Figure 9 shows that, on average, members who participated in more digitized events also participated in relatively

more niche groups, a pattern that is more closely aligned with long-tail consumption. The relationship between digitization and the popularity of events is, instead, closer to U-shaped (Web Appendix F). We present these patterns in the hope that future research can expand on these analyses and provide rigorous causal evidence on the consumer-level implications of digitization effects, which may or may not differ between traditional consumer markets and other aspects of the economy.

In closing, this study provides initial quantitative and generalizable insights into how digitization technologies can enhance social engagement in diverse local and distributed communities. Our findings have significant implications for all stakeholders involved in the economics of digitization, particularly policy-makers, companies enabling digital transformation, and individual members of local communities. We hope to inspire future research to further investigate, question, and ultimately expand the boundaries of human connection in an increasingly digitized world.

References

- Aarons-Mele, M. (2022). Working Through Your On-Camera Meeting Anxiety. *Harvard Business Review*.
- Angelici, M., & Profeta, P. (2024). Smart working: Work flexibility without constraints. *Management Science*, 70(3), 1680–1705.
- Arun, R., Suresh, V., Veni, M., & Narasimha, M. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. *Proceedings of the 14th Pacific-Asia Conference*, 391–402.
- Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25), 3083–3107.
- Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4), 962–973.
- Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm. *CREA Discussion Papers*, (13).
- Bettinger, E. P., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, 107(9), 2855–2875.
- Bloom, N., Han, R., & Liang, J. (2024). Hybrid working from home improves retention without damaging performance. *Nature*, 1–6.
- Bonsall, A. (2023, February). 3 types of meetings – and how to do each one well. <https://hbr.org/2022/09/3-types-of-meetings-and-how-to-do-each-one-well>
- Bremer, R. (2022, February). *How to lower the no-show of your events*. (tech. rep.). Momice.com.
- Brynjolfsson, E., Wang, C., & Zhang, X. (2021). The economics of it and digitization: Eight questions for research. *MIS quarterly*, 45(1), 473–477.
- Business Research Insights. (2024). Virtual Events Market Size | Global Report [2033].
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive lda model selection. *Neurocomputing*, 72(7-9), 1775–1781.
- Capossela, C. (2022). To Get People Back in the Office, Make It Social. *Harvard Business Review*.
- Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39–67.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1), 61–84.
- Di Pietro, G. (2023). The impact of covid-19 on student achievement: Evidence from a recent meta-analysis. *Educational research review*, 100530.
- D’Souza, J. (2025, January). Virtual events statistics by impacts, sponsorship, virtual & hybrid events, b2b industry and facts.
- EventUp Planner. (2024, August). *7 event KPIs that will prove you put on a great event* (tech. rep.). Eventupplanner.com.
- Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of economic literature*, 57(1), 3–43.

- Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2), 193–225.
- Gortazar, L., Hupkau, C., & Roldán-Monés, A. (2024). Online tutoring works: Experimental evidence from a program with vulnerable children. *Journal of Public Economics*, 232, 105082.
- Grand View Research. (2025). *Virtual events market size, share & trends analysis report* (tech. rep. No. GVR-4-68038-795-7).
- Grewal, R., Gupta, S., & Hamilton, R. (2021). Marketing insights from multimedia data: Text, image, audio, and video. *Journal of Marketing Research*, 58(6), 1025–1033.
- Haw, I.-M., Swink, M., & Zhang, W. (2023). The role of direct equity ownership in supply chains. *Journal of Operations Management*, 69(4), 586–615.
- Ingram, C., & Drachen, A. (2022). Impact of social distancing on face to face meetups for software practitioners during the covid-19 pandemic. *Proceedings of the ACM on Human-Computer Interaction*, 6, 1–22.
- Kofoed, M. S., Gebhart, L., Gilmore, D., & Moschitto, R. (2024). Zooming to class? experimental evidence on college students' online learning during covid-19. *AER: Insights*, 6(3).
- Krugmann, J. O., & Hartmann, J. (2024). Sentiment analysis in the age of generative ai. *Customer Needs and Solutions*, 11(1), 3.
- Le Mens, G., & Gallego, A. (2025). Positioning political texts with large language models by asking and averaging. *Political Analysis*, 33(3), 274–282.
- Lee, S., Lee, J.-H., & Jeong, Y. (2023). The effects of digital textbooks on students' academic performance, academic interest, and learning skills. *Journal of Marketing Research*, 60(4), 792–811.
- Lichand, G., Doria, C. A., Leal-Neto, O., & Fernandes, J. P. C. (2022). The impacts of remote learning in secondary education during the pandemic in brazil. *Nature Human Behaviour*, 6(8), 1079–1086.
- Lim, K. H., & Benbasat, I. (2002). The influence of multimedia on improving the comprehension of organizational information. *Journal of Management Information Systems*, 19(1), 99–127.
- Luo, B., & Ranjan, B. (2025). Express: Mapping spatial heterogeneity in retail advertising effectiveness. *Journal of Marketing Research*.
- Meetup. (2025). Media - Meetup 2025. <https://www.meetup.com/media/>
- Mills, C. (2023, August). *Adapt your In-Person B2B event strategy or lose audiences* (tech. rep.). Forrester.com.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Robinson, B. (2023). 'The Great Mismatch': Employers Firmer On Return-To-Office Policies In 2023. *Forbes*.
- Sarabipour, S. (2020). Virtual conferences raise standards for accessibility and interactions. *Elife*.
- Sconti, A. (2022). Digital vs. in-person financial education: What works best for generation z? *Journal of Economic Behavior & Organization*, 194, 300–318.
- Shapiro, B. T., Hitsch, G. J., & Tuchman, A. E. (2021). Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica*, 89(4), 1855–1879.
- Touré-Tillery, M., & Wang, L. (2022). The good-on-paper effect: How the decision context influences virtuous behavior. *Marketing Science*.
- Uhlhorn, T. (2025, January). *In person events: Your guide to planning & technology* (tech. rep.). spotme.com.

- US Census Bureau. (2022, May). City and town population totals: 2020-2021. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>
- Zajdela, E. R., Huynh, K., Feig, A. L., Wiener, R. J., & Abrams, D. M. (2025). Face-to-face or face-to-screen: A quantitative comparison of conferences modalities. *PNAS nexus*, 4(1).

Web Appendix to “Digitization and Community Participation”

A	Details on Data Collection	1
B	LDA Topic Modeling of Interest Categories	2
C	Classification Task with LLMs	7
C.1	Prompts	7
C.2	Performance and Benchmarking	8
C.3	Summary Statistics	8
D	Covariate Balance Diagnostics	8
E	Additional Tables and Figures	9
E.1	Linear Probability Model – 2024 Estimates	9
E.2	Linear Probability Models – Monthly Estimates	12
E.3	Additional CATTs Figures	12
F	Metrics of Diversity and Popularity of Events	16
G	Robustness Checks	16
G.1	Regression Sensitivity Analysis	16
G.2	Placebo Treatment Assignment	18

Web Appendix A: Details on Data Collection

We collected data using Meetups’ GraphQL API schema (Meetup API, 2022). The data collection algorithm follows the hierarchical structure of decision-making on Meetup: first, we collected information about the groups, then data about their events, and finally, members’ registrations for the events of their groups. Data collection was performed in early 2025.

Querying GraphQL for groups: We constructed a GraphQL query using the geographical coordinates for the centroid of the top 10 largest cities in the U.S. by population, according to the latest available Census (in our case, the 2020 Census, US Census Bureau 2022). The cities are New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, San Antonio, San Diego, Dallas, and Jacksonville. In addition to the coordinates, the API query requires the inclusion of at least one keyword. For each query invoking a pair of coordinates, we used keywords closely related to Meetup’s “popular topics” (Meetup, 2020): Arts & Culture, Book Clubs, Career & Business, Cars & Motorcycles, Community & Environment, Dancing, Education & Learning, Fashion & Beauty, Fitness, Food & Drink, Games, Health & Wellbeing, Hobbies & Crafts, Language & Ethnic Identity, LGBT, Lifestyle, Movements & Politics, Movies & Film, Music, New

Age & Spirituality, Outdoors & Adventure, Paranormal, Parents & Family, Pets & Animals, Photography, Religion & Beliefs, Sci-Fi & Fantasy, Singles, Socializing, Sports & Recreation, Support, Tech, and Writing.

For each combination of locations and keywords, we asked the API to return all groups active within a 20-mile radius of the centroid, along with their public information.

Events collection: Using the group identifiers from the collection above, we queried the API to return the list of all events organized by the groups specified above during the calendar year 2024 and their public information.

Members and registrations collection: To obtain event registrations, we ran two separate queries using the event identifiers. One query returned the list of group members (at the time of the focal event) who registered for the event. The second query returned the list of group members (at the time of the focal event) for whom the registration status was “MISSING” (i.e., they did not register for the event). The union of registrations and missing registrations returns the list of members per group at the time of each event.

Web Appendix B: LDA Topic Modeling of Interest Categories

We performed LDA topic modeling to classify events into latent interest topics, based on the free text descriptions provided by the event organizers (Blei et al., 2003). This analysis aims to reduce the number of interest categories describing the events, starting from an upper bound of 21 original interest categories included in the raw Meetup data.

We select the number of latent topics for LDA models based on three metrics from Arun et al. 2010; Cao et al. 2009; Deveaud et al. 2014, using the R package “lda tuning” (Murzintcev, 2020). We chose an 8-topic model solution, based on both quantitative criteria and the objective of achieving dimension reduction (Figure W1). In particular, we noted that the Deveaud et al., 2014 metric has an inflection point at 8 topics, and then decreases slightly before stabilizing between 9 and 11 topics. We note similar patterns for the Cao et al., 2009 metric: the value of the metric drops sharply at eight topics. The nearest drops are only recorded at 12, 14, and 22 topics. However, a larger number of topics might not help mitigate the sparsity problem.

Figure W2 shows the top 20 words associated with each of the eight topics. These per-topic-per-word probabilities help us qualitatively characterize the topics in Section “Data” and interpret the main results of this study. Table W1 shows the average topic weights for each of the 21 predefined interest categories. Finally, Table W2 shows that the distribution of groups, events, and event digitization by interest topics is more balanced compared to the original 21 interest categories.

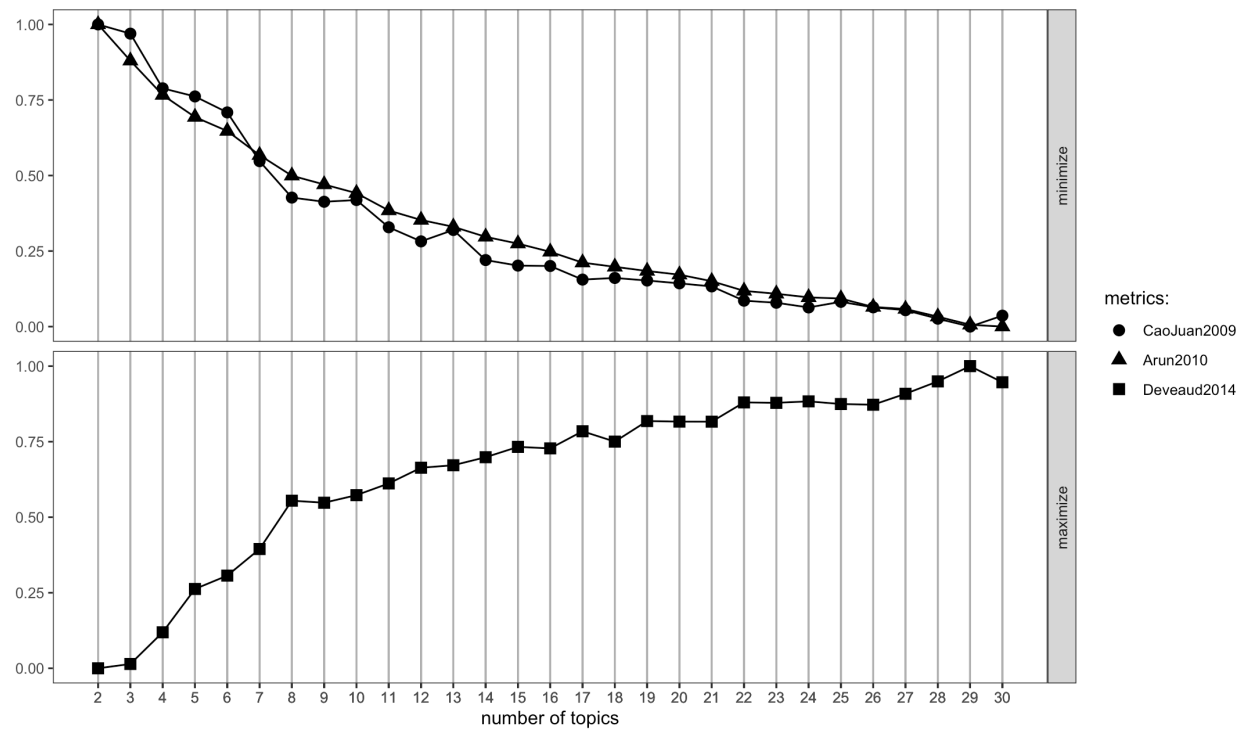


Figure W1: Recommended Topic Solutions

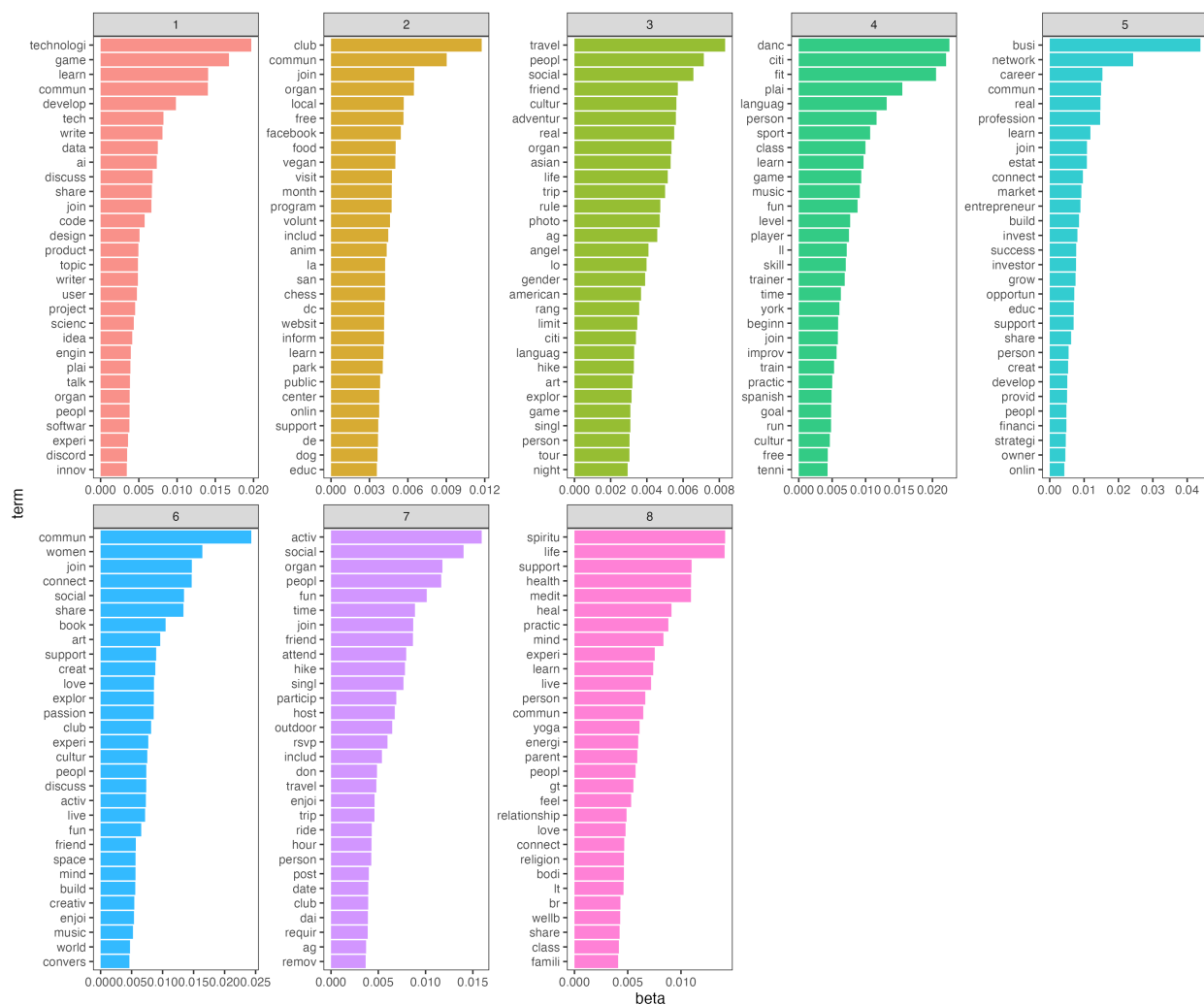


Figure W2: 8-Topic LDA Solution per-topic-per-word Probabilities (β , stemmed words).

Meetup Interest Category	Average Latent Topic Loadings							
	Business/ Careers	Community/ Causes	Community/ Culture	Dancing/ Fitness	Socializ- ation	Spiritu- ality	Tech	Travel
Art & Culture	0.05	0.06	0.52	0.12	0.07	0.06	0.06	0.07
Career & Business	0.74	0.06	0.06	0.02	0.03	0.03	0.06	0.01
Community & Environment	0.16	0.28	0.28	0.03	0.09	0.10	0.04	0.02
Dancing	0.01	0.06	0.12	0.63	0.09	0.04	0.01	0.05
Games	0.01	0.12	0.09	0.15	0.21	0.01	0.37	0.04
Health & Wellbeing	0.08	0.07	0.10	0.04	0.04	0.64	0.01	0.01
Hobbies & Passions	0.04	0.08	0.53	0.03	0.12	0.04	0.07	0.08
Identity & Language	0.02	0.09	0.29	0.31	0.12	0.04	0.03	0.10
Movements & Politics	0.05	0.21	0.36	0.01	0.10	0.11	0.08	0.07
Music	0.02	0.08	0.35	0.32	0.11	0.03	0.04	0.04
Not Specified	0.02	0.16	0.31	0.04	0.32	0.07	0.02	0.06
Parents & Family	0.06	0.09	0.15	0.07	0.12	0.48	0.02	0.03
Pets & Animals	0.02	0.52	0.14	0.04	0.19	0.06	0.00	0.02
Religion & Spirituality	0.03	0.07	0.18	0.02	0.05	0.62	0.02	0.02
Science & Education	0.07	0.17	0.20	0.05	0.07	0.23	0.19	0.02
Social Activities	0.05	0.04	0.28	0.06	0.45	0.04	0.02	0.06
Sports & Fitness	0.02	0.08	0.06	0.55	0.19	0.05	0.02	0.03
Support & Coaching	0.09	0.06	0.15	0.02	0.09	0.57	0.02	0.01
Technology	0.14	0.05	0.06	0.01	0.03	0.01	0.68	0.01
Travel & Outdoor	0.02	0.13	0.18	0.06	0.47	0.02	0.01	0.10
Writing	0.05	0.07	0.22	0.03	0.05	0.04	0.52	0.03

Table W1: Average Latent Topic Loadings per Interest Category. Note: Group organizers can select only one Meetup interest category.

Interest Categories	N. Groups	N. Events	% Digitized Events
<i>Latent (LDA)</i>			
Business/Careers	1 958	67 384	65.86
Community/Causes	742	41 763	32.74
Community/Culture	2 343	82 334	31.39
Dancing/Fitness	1 362	110 558	21.43
Socialization	1 790	155 273	15.08
Spirituality	1 894	108 081	48.13
Tech	1 745	58 476	46.61
Travel	271	21 144	24.42
<i>Categorical (Meetup)</i>			
Art & Culture	291	13 052	18.98
Career & Business	1 907	72 866	64.90
Community & Environment	228	8 216	23.29
Dancing	271	17 444	8.30
Games	464	33 963	6.67
Health & Wellbeing	824	55 632	53.34
Hobbies & Passions	824	28 824	41.88
Identity & Language	610	39 671	62.64
Movements & Politics	94	2 776	41.06
Music	328	11 600	30.95
Not Specified	335	15 365	3.16
Parents & Family	93	2 313	49.83
Pets & Animals	84	2 114	2.68
Religion & Spirituality	920	56 029	57.13
Science & Education	204	11 684	69.29
Social Activities	1 367	123 508	20.34
Sports & Fitness	776	68 807	0.41
Support & Coaching	293	12 435	68.69
Technology	1 396	23 971	56.99
Travel & Outdoor	622	34 406	3.70
Writing	174	10 337	54.49

Table W2: Distribution of Groups, Events, and Event Digitization by Interest Topics. Notes: Assignment of latent topics to groups and events is based on the latent topic with the highest loading. If two or more latent topics share the same (highest) loading, the group or event is assigned to all matching topics.

Web Appendix C: Classification Task with LLMs

We prompted the OpenAI model gpt-4o-mini-2024-07-18 to classify the 157,464 unique event descriptions in the list of events organized in 2024 via the OpenAI API in February 2025. We set the temperature parameter at 0, such that the classification process will most likely select the classification label with the highest probability. Event descriptions are split into tokens by the chat completion algorithm. The API requests entailed 418,693,957 combined input and output tokens.

C.1 Prompts

The prompts used in this task had a fixed format (“You are a highly intelligent assistant tasked with classifying event descriptions. You will be provided with an event description, and you will output a “Yes” or “No” answer to the following question.”) and varied in the classification instructions. The prompts were engineered following the 2025 OpenAI prompting guide (OpenAI, 2025). We ran separate tasks to measure the additional event characteristics: “social interactivity” (whether the event hinges on social interaction), “specialized knowledge” (whether the event covers specialized content and requires technical or prior knowledge of the topic at hand), “skill development” (whether the event’s focus is on skill development, training, and education), and “physicality” (whether the event entailed hands-on experiences that require a physical presence). Below, we report the classification instructions added to the base prompt for each separate task.

Social Interactivity Prompt: You are a highly intelligent assistant tasked with classifying event descriptions. You will be provided with an event description, and you will output a “Yes” or “No” answer to the following question: “Does the event offer opportunities for networking or interaction, has some networking elements, and/or puts strong emphasis on interactivity, interaction sessions and networking?” Please consider all elements of the description, and only answer “Yes” or “No”. Do not alter the event description in any way.

Specialized Knowledge Prompt: You are a highly intelligent assistant tasked with classifying event descriptions. You will be provided with an event description, and you will output a “Yes” or “No” answer to the following question: “Does the event cover topics or content demanding familiarity with the topic, advanced or specialized content that requires a higher level of expertise, use technical language, and/or reference to prerequisite knowledge?” Please consider all elements of the description, and only answer “Yes” or “No”. Do not alter the event description in any way.

Skill Development Prompt: You are a highly intelligent assistant tasked with classifying event descriptions. You will be provided with an event description, and you will output a “Yes” or “No” answer to the following question: “Does the event focus on or emphasize learning, skill development, training, and/or educational goals?” Please consider all elements of the description, and only answer “Yes” or “No”. Do not alter the event description in any way.

Physicality Prompt: You are a highly intelligent assistant tasked with classifying event descriptions. You will be provided with an event description, and you will output a "Yes" or "No" answer to the following question: "Does the event involve practical or physical 'hands-on' experiences that require a physical presence, such as workshops, DIY sessions, cooking classes, art projects, physical walks and runs, eating and drinking, and science experiments?" Please consider all elements of the description, and only answer "Yes" or "No". Do not alter the event description in any way.

C.2 Performance and Benchmarking

We evaluated the variability in the classification performed by the LLM on a subset of 1,000 event descriptions randomly sampled from the data. In particular, we performed each classification task three times on the subset of descriptions, resulting in 3,000 classifications in total. We then calculated the rate of agreement across classifications (inter-rater reliability) within each task and the proportion of descriptions that achieved full agreement across all four tasks.

The resulting inter-rater reliability rates are 99.1% (Social Interactivity), 99.3% (Specialized Knowledge), 99.1% (Skill Development), 98.4% (Physicality), and the share of descriptions that achieved full agreement is 96%. These results indicate that the LLM performance is stable across classification rounds.

We also evaluated the performance of OpenAI's gpt-4o-mini-2024-07-18 against Deepseek's "deepseek-reasoner" model DeepSeek-R1-0528 on the same 3,000 classifications (3 classification rounds for 1,000 descriptions) and found that the two models achieved highly similar classification results. Comparing the case-by-case classifications performed by gpt-4o-mini-2024-07-18 with DeepSeek-R1-0528, we found that the inter-rater reliability rates between the two models are 87.2% (Social Interactivity), 95.3% (Specialized Knowledge), 89.7% (Skill Development), 91% (Physicality), and the overall average inter-rater agreement is 95.3%. The high agreement rates suggest that the classification results are not sensitive to the use of a specific LLM.

C.3 Summary Statistics

Table W3 presents the prevalence of the four variables across events weighted by latent group topics.

Web Appendix D: Covariate Balance Diagnostics

The propensity score model includes all the relevant covariates and fixed effects, except individual-level fixed effects. The histogram of propensity scores by treatment group shows sufficient overlap across the entire range of propensity score bins, lending support to the conditional ignorability assumption (Figure W3).

We further assess whether the distributions of theory-relevant covariates are similar between the treated and control units in our sample, after accounting for the inverse propensity weights and the relevant fixed effects (Ho et al., 2011). To do this, we calculate the Standardized Mean Differences (SMDs), computed

Topic	% Events			
	Social Interactivity	Specialized Knowledge	Skill Development	Physicality
Business/Careers	70.24	12.54	68.48	10.13
Community/Causes	54.47	8.27	50.53	35.29
Community/Culture	62.73	5.74	41.10	31.64
Dancing/Fitness	56.83	6.89	46.98	59.81
Socialization	69.41	5.13	19.95	42.04
Spirituality	35.48	9.22	63.15	23.58
Tech	62.49	25.74	57.90	18.91
Travel	66.18	4.38	30.49	40.72

Table W3: Incidence of Event Characteristics by Latent Interest Topics. Note: Event feature incidences are weighted by the organizing group’s topic loadings.

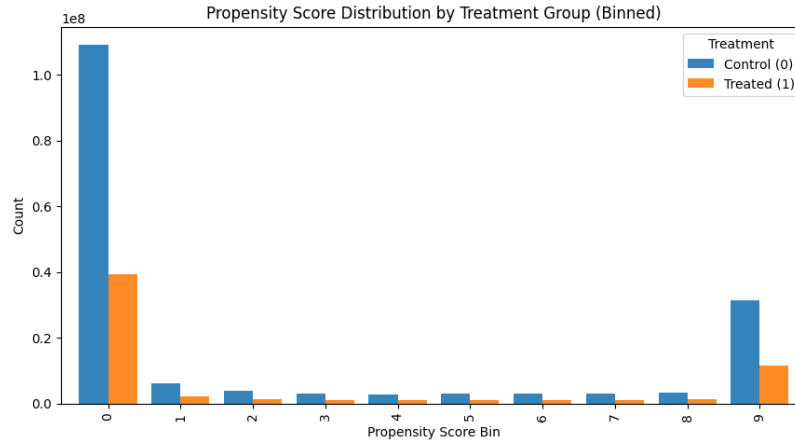


Figure W3: Propensity scores overlap – PS grouped in 10 bins for computational feasibility.

and plotted before inverse-propensity score weighting, of each covariate (and relevant interactions) between treatment and control units across propensity score strata (Ali et al., 2014; Belitser et al., 2011; Stuart et al., 2013). Figure W4 suggests that, before adjusting for inverse propensity score weights, several covariates and interactions exhibit SMDs greater than 0.1, and a few approaching or exceeding 0.25 – especially events that involve “hands-on” activities (Physicality), groups that involve “Socialization” topics, and the length of the event text descriptions.

The plot suggests that the conditional ignorability assumption is not plausible without covariate adjustment. In the analyses, we account for the imbalance through a doubly robust covariate adjustment in the outcome regression, as well as by including covariates, interaction terms, and fixed effects.

Web Appendix E: Additional Tables and Figures

E.1 Linear Probability Model – 2024 Estimates

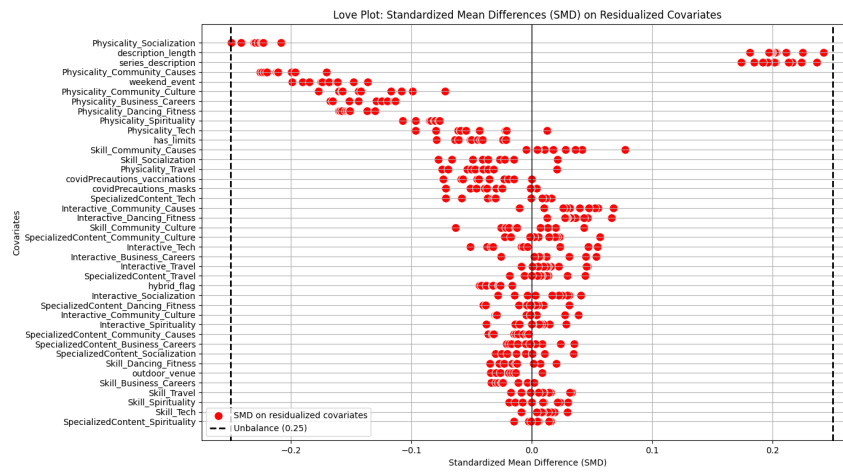


Figure W4: Standardized Mean Differences by Covariate across Propensity Score Bins (Love Plot)

Variable	Estimate	Std. Error
Online × Social Interactivity	-0.0008	0.0000
Online × Specialized Knowledge	-0.0006	0.0001
Online × Skill Development	0.0007	0.0000
Online × Physicality	0.0003	0.0001
Online × Tech	-0.0037	0.0001
Online × Community/Causes	0.0019	0.0001
Online × Travel	0.0017	0.0001
Online × Dancing/Fitness	0.0026	0.0001
Online × Business/Careers	-0.0006	0.0001
Online × Community/Culture	0.0025	0.0001
Online × Socialization	0.0001	0.0001
Online × Spirituality	0.0023	0.0001
Social Interactivity × Tech	0.0012	0.0001
Social Interactivity × Community/Causes	-0.0007	0.0001
Social Interactivity × Travel	-0.0004	0.0001
Social Interactivity × Dancing/Fitness	0.0019	0.0001
Social Interactivity × Business/Careers	0.0018	0.0001
Social Interactivity × Community/Culture	0.0012	0.0001
Social Interactivity × Socialization	0.0008	0.0000
Social Interactivity × Spirituality	0.001	0.0001
Specialized Knowledge × Tech	0.0003	0.0001
Specialized Knowledge × Community/Causes	0.0022	0.0001
Specialized Knowledge × Travel	-0.0031	0.0002
Specialized Knowledge × Dancing/Fitness	-0.0002	0.0001
Specialized Knowledge × Business/Careers	0.0002	0.0001
Specialized Knowledge × Community/Culture	-0.0005	0.0001
Specialized Knowledge × Socialization	-0.0009	0.0001
Specialized Knowledge × Spirituality	0.0003	0.0001
Skill Development × Tech	0.0009	0.0001
Skill Development × Community/Causes	0.0009	0.0001
Skill Development × Travel	-0.0005	0.0001
Skill Development × Dancing/Fitness	-0.0015	0.0001
Skill Development × Business/Careers	-0.001	0.0001
Skill Development × Community/Culture	-0.002	0.0001
Skill Development × Socialization	-0.0005	0.0000
Skill Development × Spirituality	-0.0006	0.0001
Physicality × Tech	-0.0009	0.0001
Physicality × Community/Causes	0.0012	0.0001
Physicality × Travel	0.0001	0.0001
Physicality × Dancing/Fitness	-0.001	0.0001
Physicality × Business/Careers	-0.0004	0.0001
Physicality × Community/Culture	0.0000	0.0001
Physicality × Socialization	-0.0005	0.0000
Physicality × Spirituality	0.001	0.0001
Masks required	0.0015	0.0001
Vaccination required	-0.0001	0.0001
Hybrid event	0.0009	0.0003
Outdoor venue	-0.0006	0.0000
Event series	-0.0005	0.0000
Description length	0.0000	0.0000
RSVP limits	0.0004	0.0000
Weekend event	0.0007	0.0000
ATT weight	0.0000	0.0000
R-squared	0.0001	0.0000
N.obs	231227896	0.0000

Table W4: 2024 Linear Probability Model Coefficient Estimates.

E.2 Linear Probability Models – Monthly Estimates

	January	February	March	April	May	June
Online x Social Interactivity	−0.0017 (0.0002)***	−0.0020 (0.0002)***	−0.0009 (0.0002)***	−0.0013 (0.0002)***	−0.0015 (0.0002)***	−0.0018 (0.0002)***
Online x Specialized Knowledge	−0.0022 (0.0003)***	−0.0028 (0.0003)***	0.0001 (0.0003)	−0.0016 (0.0003)***	−0.0025 (0.0003)***	−0.0020 (0.0003)***
Online x Skill Development	0.0015 (0.0002)***	0.0015 (0.0002)***	0.0011 (0.0002)***	0.0013 (0.0002)***	0.0006 (0.0002)***	0.0006 (0.0002)**
Online x Physicality	−0.0007 (0.0003)*	0.0007 (0.0003)*	−0.0006 (0.0003)*	−0.0005 (0.0003)	−0.0008 (0.0003)**	−0.0014 (0.0003)***
Online x Tech	0.0045 (0.0007)***	−0.0127 (0.0007)***	−0.0053 (0.0007)***	−0.0045 (0.0006)***	−0.0034 (0.0007)***	−0.0044 (0.0006)***
Online x Community/Causes	0.0014 (0.0007)*	0.0056 (0.0007)***	0.0015 (0.0006)*	0.0008 (0.0006)	0.0019 (0.0006)***	0.0045 (0.0006)***
Online x Travel	0.0014 (0.0006)*	0.0010 (0.0006)	0.0013 (0.0007)	0.0013 (0.0007)*	0.0030 (0.0009)***	0.0024 (0.0011)*
Online x Dancing/Fitness	0.0056 (0.0007)***	0.0049 (0.0007)***	0.0054 (0.0007)***	0.0046 (0.0007)***	0.0048 (0.0007)***	0.0039 (0.0007)***
Online x Business/Careers	−0.0036 (0.0006)***	0.0005 (0.0006)	−0.0007 (0.0005)	−0.0039 (0.0006)***	−0.0008 (0.0005)	0.0020 (0.0006)***
Online x Community/Culture	0.0039 (0.0005)***	0.0041 (0.0005)***	0.0035 (0.0004)***	0.0054 (0.0004)***	0.0045 (0.0004)***	0.0032 (0.0004)***
Online x Socialization	0.0007 (0.0003)*	0.0015 (0.0003)***	−0.0001 (0.0003)	0.0003 (0.0003)	0.0010 (0.0003)***	0.0012 (0.0003)***
Online x Spirituality	0.0044 (0.0004)***	0.0032 (0.0004)***	0.0033 (0.0004)***	0.0037 (0.0004)***	0.0025 (0.0004)***	0.0030 (0.0004)***
Social Interactivity x Tech	0.0022 (0.0004)***	0.0030 (0.0004)***	0.0008 (0.0004)*	0.0037 (0.0004)***	0.0011 (0.0003)**	0.0007 (0.0004)
Social Interactivity x Community/Causes	0.0019 (0.0005)***	−0.0005 (0.0005)	−0.0003 (0.0004)	−0.0007 (0.0004)	0.0029 (0.0004)***	0.0009 (0.0004)*
Social Interactivity x Travel	−0.0018 (0.0004)***	0.0002 (0.0003)	0.0001 (0.0003)	−0.0003 (0.0003)	−0.0002 (0.0002)	−0.0001 (0.0003)
Social Interactivity x Dancing/Fitness	−0.0007 (0.0003)*	0.0035 (0.0003)***	0.0025 (0.0003)***	0.0016 (0.0003)***	0.0031 (0.0003)***	0.0028 (0.0003)***
Social Interactivity x Business/Careers	0.0035 (0.0005)***	0.0021 (0.0005)***	0.0024 (0.0004)***	0.0031 (0.0005)***	0.0031 (0.0005)***	0.0017 (0.0004)***
Social Interactivity x Community/Culture	0.0018 (0.0003)***	0.0055 (0.0003)***	0.0024 (0.0003)***	0.0017 (0.0003)***	0.0025 (0.0003)***	0.0023 (0.0003)***
Social Interactivity x Socialization	0.0013 (0.0002)***	0.0007 (0.0002)***	0.0006 (0.0001)***	0.0014 (0.0001)***	0.0010 (0.0001)***	0.0017 (0.0001)***
Social Interactivity x Spirituality	−0.0005 (0.0004)	0.0020 (0.0004)***	0.0013 (0.0003)***	0.0021 (0.0004)***	0.0014 (0.0004)***	0.0012 (0.0004)***
Specialized Knowledge x Tech	−0.0001 (0.0004)	0.0028 (0.0004)***	0.0010 (0.0004)**	0.0007 (0.0004)	0.0051 (0.0004)***	0.0025 (0.0004)***
Specialized Knowledge x Community/Causes	0.0044 (0.0008)***	0.0013 (0.0007)*	0.0018 (0.0006)**	0.0031 (0.0006)***	0.0036 (0.0006)***	0.0031 (0.0006)***
Specialized Knowledge x Travel	−0.0065 (0.0011)***	−0.0025 (0.0008)**	−0.0007 (0.0010)	−0.0050 (0.0010)***	−0.0019 (0.0011)	−0.0017 (0.0010)
Specialized Knowledge x Dancing/Fitness	0.0001 (0.0003)	−0.0009 (0.0003)**	0.0002 (0.0003)	−0.0006 (0.0003)	−0.0003 (0.0003)	0.0005 (0.0004)
Specialized Knowledge x Business/Careers	0.0038 (0.0006)***	0.0032 (0.0006)***	0.0001 (0.0006)	0.0020 (0.0006)***	0.0019 (0.0005)***	−0.0004 (0.0005)
Specialized Knowledge x Community/Culture	0.0017 (0.0008)*	−0.0006 (0.0006)	−0.0028 (0.0006)***	0.0024 (0.0007)***	0.0004 (0.0006)	−0.0005 (0.0006)
Specialized Knowledge x Spirituality	0.0009 (0.0006)	0.0003 (0.0006)	0.0014 (0.0005)**	−0.0014 (0.0005)**	0.0031 (0.0006)***	0.0015 (0.0005)**
Skill Development x Tech	0.0001 (0.0004)	0.0009 (0.0004)*	0.0017 (0.0004)***	0.0036 (0.0004)***	0.0029 (0.0004)***	0.0017 (0.0004)***
Skill Development x Community/Causes	0.0006 (0.0005)	0.0034 (0.0005)***	0.0003 (0.0005)	0.0007 (0.0004)	−0.0000 (0.0005)	−0.0005 (0.0004)
Skill Development x Travel	0.0012 (0.0004)**	−0.0011 (0.0004)*	0.0020 (0.0004)***	−0.0008 (0.0004)*	−0.0004 (0.0004)	−0.0009 (0.0004)*
Skill Development x Dancing/Fitness	−0.0031 (0.0003)***	−0.0050 (0.0003)***	−0.0043 (0.0003)***	−0.0039 (0.0003)***	−0.0021 (0.0003)***	−0.0000 (0.0003)
Skill Development x Business/Careers	−0.0007 (0.0006)	−0.0017 (0.0006)**	0.0004 (0.0006)	−0.0029 (0.0005)***	0.0006 (0.0005)	−0.0012 (0.0005)*
Skill Development x Community/Culture	−0.0016 (0.0004)***	−0.0034 (0.0004)***	−0.0015 (0.0003)***	−0.0036 (0.0003)***	−0.0019 (0.0003)***	−0.0020 (0.0003)***
Skill Development x Socialization	−0.0008 (0.0002)***	−0.0006 (0.0002)**	−0.0006 (0.0002)***	−0.0002 (0.0002)	−0.0011 (0.0002)***	−0.0008 (0.0002)***
Skill Development x Spirituality	−0.0011 (0.0003)**	−0.0005 (0.0003)	−0.0007 (0.0003)*	−0.0012 (0.0003)***	−0.0004 (0.0003)	−0.0006 (0.0003)
Physicality x Tech	−0.0004 (0.0005)	−0.0056 (0.0004)***	0.0001 (0.0004)	0.0015 (0.0004)***	0.0021 (0.0004)***	0.0023 (0.0004)***
Physicality x Community/Causes	−0.0027 (0.0006)***	−0.0001 (0.0006)	0.0008 (0.0005)	0.0009 (0.0005)	0.0003 (0.0005)	0.0029 (0.0005)***
Physicality x Travel	0.0000 (0.0003)	0.0001 (0.0003)	−0.0013 (0.0003)***	−0.0005 (0.0003)	0.0009 (0.0003)***	0.0005 (0.0003)
Physicality x Dancing/Fitness	0.0004 (0.0005)	−0.0012 (0.0005)*	−0.0009 (0.0004)*	−0.0001 (0.0004)	−0.0005 (0.0004)	−0.0016 (0.0004)***
Physicality x Business/Careers	−0.0117 (0.0008)***	−0.0058 (0.0008)***	−0.0060 (0.0008)***	−0.0024 (0.0008)**	−0.0029 (0.0007)***	0.0011 (0.0007)
Physicality x Community/Culture	0.0010 (0.0004)**	0.0007 (0.0003)	−0.0003 (0.0003)	−0.0006 (0.0003)*	−0.0017 (0.0003)***	0.0003 (0.0003)
Physicality x Socialization	−0.0008 (0.0002)***	−0.0009 (0.0002)***	−0.0005 (0.0001)**	−0.0006 (0.0001)***	−0.0000 (0.0001)	−0.0005 (0.0001)***
Physicality x Spirituality	0.0018 (0.0004)***	0.0008 (0.0004)	0.0008 (0.0004)*	0.0008 (0.0004)*	0.0017 (0.0004)***	0.0025 (0.0004)***
Masks	0.0015 (0.0003)***	0.0031 (0.0003)***	0.0024 (0.0003)***	0.0026 (0.0003)***	0.0014 (0.0003)***	0.0019 (0.0003)***
Vaccinations	0.0001 (0.0003)	−0.0011 (0.0003)***	−0.0002 (0.0002)	−0.0006 (0.0003)*	0.0008 (0.0003)**	−0.0003 (0.0003)
Hybrid	0.0028 (0.0022)	0.0361 (0.0022)***	−0.0037 (0.0015)*	0.0180 (0.0012)***	0.0573 (0.0015)***	0.0097 (0.0020)***
Outdoor Venue	−0.0012 (0.0002)***	−0.0004 (0.0002)	−0.0018 (0.0002)***	−0.0012 (0.0002)***	−0.0008 (0.0002)***	0.0007 (0.0002)***
Event Series	−0.0006 (0.0001)***	−0.0009 (0.0001)***	−0.0004 (0.0001)***	−0.0007 (0.0001)***	−0.0003 (0.0001)***	−0.0004 (0.0001)***
Description Length	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***
RSVP Limits	0.0005 (0.0001)***	0.0001 (0.0001)	0.0005 (0.0001)***	0.0005 (0.0001)***	0.0006 (0.0001)***	−0.0000 (0.0001)
Weekend Event	0.0010 (0.0001)***	0.0012 (0.0001)***	0.0008 (0.0001)***	0.0009 (0.0001)***	0.0007 (0.0001)***	0.0007 (0.0001)***
ATT Weights	0.0000 (0.0000)*	0.0000 (0.0000)***	0.0000 (0.0000)	0.0000 (0.0000)***	−0.0000 (0.0000)	0.0000 (0.0000)*
Num. obs.	15931970	16315067	17939833	17935406	18421674	18480371
Num. groups: week nr.	5	5	5	5	5	5
Num. groups: group id	6536	6725	7002	7189	7110	7233
Num. groups: member id	422773	434935	438920	440489	441515	441607
R ² (full model)	0.3359	0.3333	0.3295	0.3258	0.3277	0.3190

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table W5: January – June 2024 Monthly LPM Coefficient Estimates.

E.3 Additional CATTs Figures

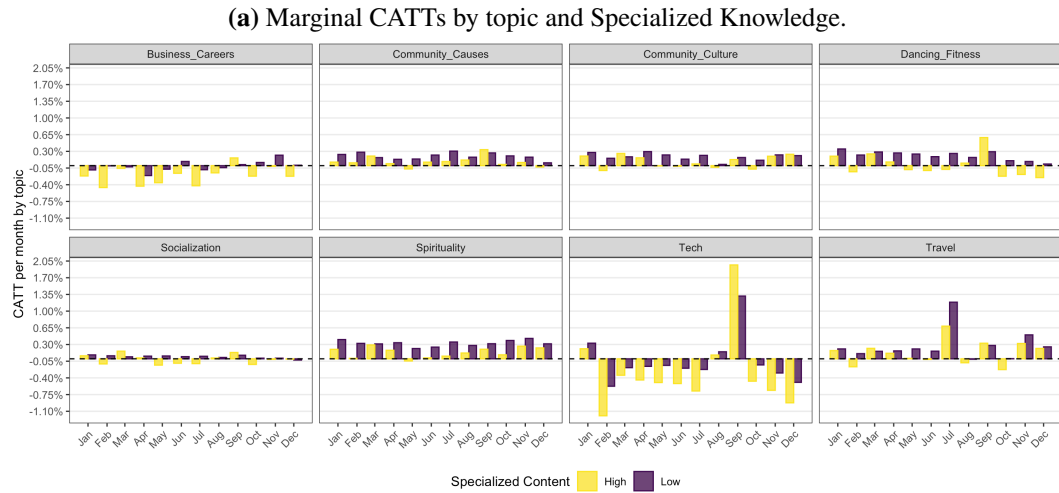
Standard errors for CATTs in Figure W6 are obtained by i) simulating 30 coefficient estimates from a normal distribution centered on the point estimate, with standard errors obtained from the monthly regression model (using the `arm` R package, Gelman and Su 2024); ii) calculating the relevant treatment effects conditional on the simulated coefficients; and iii) quantifying uncertainty in the simulated treatment effect predictions. Owing to the large sample sizes, the resulting uncertainty bounds for treatment effects are typically very

	July	August	September	October	November	December
Online x Social Interactivity	-0.0004 (0.0002)**	-0.0006 (0.0002)***	0.0000 (0.0002)	-0.0008 (0.0002)***	-0.0007 (0.0002)***	-0.0009 (0.0002)***
Online x Specialized Knowledge	-0.0028 (0.0003)***	-0.0014 (0.0003)***	-0.0014 (0.0003)***	-0.0026 (0.0003)***	-0.0010 (0.0003)***	-0.0006 (0.0003)*
Online x Skill Development	0.0018 (0.0002)***	0.0010 (0.0002)***	0.0007 (0.0002)***	0.0009 (0.0002)***	0.0005 (0.0002)**	0.0009 (0.0002)***
Online x Physicality	-0.0000 (0.0003)	0.0014 (0.0003)***	-0.0021 (0.0003)***	0.0014 (0.0003)***	0.0009 (0.0003)**	0.0009 (0.0003)**
Online x Tech	-0.0071 (0.0006)***	0.0018 (0.0006)**	0.0253 (0.0006)***	-0.0042 (0.0006)***	-0.0081 (0.0007)***	-0.0119 (0.0007)***
Online x Community/Causes	0.0035 (0.0006)***	0.0026 (0.0006)***	0.0021 (0.0006)***	0.0036 (0.0005)***	0.0015 (0.0006)**	-0.0001 (0.0006)
Online x Travel	0.0191 (0.0008)***	-0.0014 (0.0009)	0.0027 (0.0008)***	-0.0015 (0.0009)	0.0082 (0.0010)***	0.0036 (0.0007)***
Online x Dancing/Fitness	0.0029 (0.0006)***	0.0019 (0.0006)**	0.0002 (0.0006)	0.0014 (0.0006)*	0.0018 (0.0006)**	0.0018 (0.0007)**
Online x Business/Careers	-0.0025 (0.0005)***	-0.0018 (0.0006)***	-0.0046 (0.0005)***	0.0007 (0.0005)	0.0036 (0.0006)***	0.0008 (0.0006)
Online x Community/Culture	0.0026 (0.0004)***	-0.0007 (0.0004)	-0.0020 (0.0004)***	0.0020 (0.0004)***	0.0047 (0.0004)***	0.0060 (0.0005)***
Online x Socialization	-0.0004 (0.0003)	0.0006 (0.0003)*	-0.0004 (0.0003)	0.0000 (0.0003)	-0.0007 (0.0003)**	-0.0008 (0.0003)**
Online x Spirituality	0.0028 (0.0004)***	0.0032 (0.0004)***	0.0036 (0.0004)***	0.0046 (0.0004)***	0.0052 (0.0004)***	0.0037 (0.0004)***
Social Interactivity x Tech	0.0007 (0.0004)	0.0007 (0.0004)	0.0118 (0.0004)***	0.0011 (0.0004)**	0.0032 (0.0004)***	-0.0014 (0.0005)**
Social Interactivity x Community/Causes	-0.0004 (0.0004)	0.0008 (0.0004)*	-0.0007 (0.0004)	-0.0017 (0.0004)***	0.0009 (0.0004)*	-0.0016 (0.0004)***
Social Interactivity x Travel	0.0003 (0.0003)	0.0001 (0.0003)	-0.0002 (0.0003)	-0.0001 (0.0003)	-0.0008 (0.0003)*	0.0006 (0.0003)
Social Interactivity x Dancing/Fitness	0.0016 (0.0003)***	0.0006 (0.0003)*	0.0014 (0.0003)***	0.0029 (0.0003)***	0.0035 (0.0003)***	0.0027 (0.0003)***
Social Interactivity x Business/Careers	0.0028 (0.0004)***	0.0033 (0.0005)***	-0.0020 (0.0004)***	0.0030 (0.0004)***	0.0004 (0.0005)	0.0033 (0.0005)***
Social Interactivity x Community/Culture	0.0013 (0.0003)***	0.0027 (0.0003)***	0.0002 (0.0003)	0.0023 (0.0003)***	0.0019 (0.0003)***	0.0024 (0.0003)***
Social Interactivity x Socialization	0.0014 (0.0001)***	0.0011 (0.0001)***	0.0005 (0.0001)***	0.0008 (0.0001)***	0.0012 (0.0001)***	0.0020 (0.0001)***
Social Interactivity x Spirituality	0.0006 (0.0004)	0.0015 (0.0004)***	0.0017 (0.0003)***	0.0004 (0.0003)	0.0012 (0.0003)***	0.0015 (0.0003)***
Specialized Knowledge x Tech	0.0046 (0.0004)***	0.0028 (0.0004)***	0.0119 (0.0004)***	0.0046 (0.0004)***	0.0029 (0.0004)***	0.0038 (0.0004)***
Specialized Knowledge x Community/Causes	0.0052 (0.0006)***	0.0054 (0.0006)***	0.0029 (0.0006)***	0.0021 (0.0005)***	-0.0013 (0.0006)*	-0.0014 (0.0006)*
Specialized Knowledge x Travel	-0.0017 (0.0013)	0.0008 (0.0011)	-0.0000 (0.0008)	-0.0070 (0.0013)***	-0.0036 (0.0009)***	-0.0042 (0.0011)***
Specialized Knowledge x Dancing/Fitness	-0.0023 (0.0004)***	-0.0016 (0.0004)***	-0.0011 (0.0004)**	0.0002 (0.0004)	0.0001 (0.0004)	-0.0008 (0.0004)
Specialized Knowledge x Business/Careers	-0.0007 (0.0005)	-0.0005 (0.0006)	-0.0021 (0.0005)***	-0.0006 (0.0005)	-0.0018 (0.0005)**	-0.0033 (0.0007)***
Specialized Knowledge x Community/Culture	0.0000 (0.0006)	-0.0009 (0.0006)	-0.0004 (0.0005)	0.0008 (0.0005)	-0.0011 (0.0005)*	-0.0013 (0.0006)*
Specialized Knowledge x Spirituality	0.0007 (0.0006)	0.0016 (0.0005)**	0.0023 (0.0005)***	0.0022 (0.0005)***	0.0014 (0.0005)**	0.0014 (0.0005)**
Skill Development x Tech	0.0000 (0.0004)	0.0015 (0.0004)***	-0.0015 (0.0004)***	0.0021 (0.0004)***	0.0003 (0.0004)	0.0043 (0.0004)***
Skill Development x Community/Causes	0.0004 (0.0004)	0.0006 (0.0005)	0.0024 (0.0004)***	0.0000 (0.0004)	0.0004 (0.0004)	0.0006 (0.0005)
Skill Development x Travel	0.0039 (0.0004)***	-0.0008 (0.0004)	-0.0012 (0.0003)***	-0.0009 (0.0003)**	-0.0009 (0.0004)*	-0.0019 (0.0004)***
Skill Development x Dancing/Fitness	0.0015 (0.0003)***	0.0003 (0.0003)	0.0027 (0.0003)***	0.0018 (0.0003)***	-0.0015 (0.0003)***	-0.0050 (0.0003)***
Skill Development x Business/Careers	-0.0039 (0.0005)***	-0.0026 (0.0005)***	-0.0019 (0.0004)***	-0.0001 (0.0005)	-0.0007 (0.0005)	-0.0025 (0.0005)***
Skill Development x Community/Culture	-0.0034 (0.0003)***	-0.0030 (0.0003)***	-0.0034 (0.0003)***	-0.0040 (0.0003)***	-0.0019 (0.0003)***	-0.0021 (0.0003)***
Skill Development x Socialization	-0.0012 (0.0002)***	-0.0004 (0.0002)*	-0.0011 (0.0002)***	-0.0005 (0.0002)***	-0.0002 (0.0002)	-0.0004 (0.0002)*
Skill Development x Spirituality	-0.0012 (0.0003)***	0.0002 (0.0003)	-0.0004 (0.0003)	-0.0004 (0.0003)	-0.0004 (0.0003)	-0.0012 (0.0003)***
Physicality x Tech	-0.0016 (0.0004)***	-0.0008 (0.0004)	0.0044 (0.0004)***	-0.0020 (0.0004)***	-0.0002 (0.0005)	0.0022 (0.0005)***
Physicality x Community/Causes	0.0048 (0.0005)***	0.0009 (0.0005)	0.0028 (0.0005)***	0.0027 (0.0005)***	0.0018 (0.0005)***	0.0005 (0.0006)
Physicality x Travel	-0.0013 (0.0003)***	-0.0007 (0.0003)*	0.0005 (0.0003)	-0.0005 (0.0002)*	0.0004 (0.0003)	0.0014 (0.0003)***
Physicality x Dancing/Fitness	-0.0011 (0.0004)**	-0.0018 (0.0004)***	-0.0020 (0.0004)***	-0.0016 (0.0004)***	-0.0014 (0.0004)***	-0.0018 (0.0004)***
Physicality x Business/Careers	-0.0005 (0.0007)	-0.0010 (0.0007)	-0.0012 (0.0006)	-0.0015 (0.0006)*	0.0053 (0.0007)***	-0.0042 (0.0007)***
Physicality x Community/Culture	0.0002 (0.0003)	0.0003 (0.0003)	0.0005 (0.0003)	0.0010 (0.0003)***	-0.0004 (0.0003)	-0.0011 (0.0003)***
Physicality x Socialization	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0008 (0.0001)***	-0.0010 (0.0001)***	-0.0012 (0.0001)***	-0.0011 (0.0002)***
Physicality x Spirituality	0.0008 (0.0004)*	0.0010 (0.0004)**	0.0009 (0.0004)*	0.0007 (0.0004)	0.0009 (0.0004)*	0.0019 (0.0004)***
Masks	0.0027 (0.0008)***	0.0015 (0.0010)	0.0024 (0.0008)**	-0.0020 (0.0011)	-0.0230 (0.0018)***	-0.0016 (0.0016)
Vaccinations	-0.0009 (0.0008)	-0.0005 (0.0010)	-0.0016 (0.0008)	0.0034 (0.0011)**	0.0238 (0.0018)***	0.0016 (0.0016)
Hybrid	0.0059 (0.0019)**	0.0248 (0.0017)***	0.0028 (0.0011)*	0.0172 (0.0015)***	0.0203 (0.0019)***	0.0025 (0.0015)
Outdoor Venue	0.0018 (0.0002)***	0.0010 (0.0003)**	0.0004 (0.0003)	-0.0027 (0.0004)***	0.0024 (0.0004)***	0.0044 (0.0005)***
Event Series	-0.0004 (0.0001)***	-0.0007 (0.0001)***	-0.0008 (0.0001)***	-0.0007 (0.0001)***	-0.0006 (0.0001)***	-0.0000 (0.0001)
Description Length	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***
RSVP Limits	0.0002 (0.0001)*	0.0006 (0.0001)***	0.0007 (0.0001)***	0.0004 (0.0001)***	0.0005 (0.0001)***	0.0004 (0.0001)***
Weekend Event	0.0004 (0.0001)***	0.0005 (0.0001)***	0.0005 (0.0000)***	0.0006 (0.0000)***	0.0008 (0.0001)***	0.0008 (0.0001)***
ATT Weights	0.0000 (0.0000)*	0.0000 (0.0000)	0.0000 (0.0000)***	0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)***
Num. obs.	18628416	18292245	18400369	18757891	17117385	16081499
Num. groups: week nr.	5	5	6	5	5	6
Num. groups: group id	6892	7128	7388	7686	7502	7192
Num. groups: member id	435408	437187	443778	443284	439814	429790
R ² (full model)	0.3217	0.3250	0.3306	0.3304	0.3321	0.3368

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table W6: July – December 2024 Monthly LPM Coefficient Estimates.

narrow compared to differences across values of covariates. Hence, we mainly focus on point estimates when reporting estimated treatment effects in the manuscript.



(b) Monthly share of events with high emphasis on Specialized Knowledge, weighted by topic loadings.

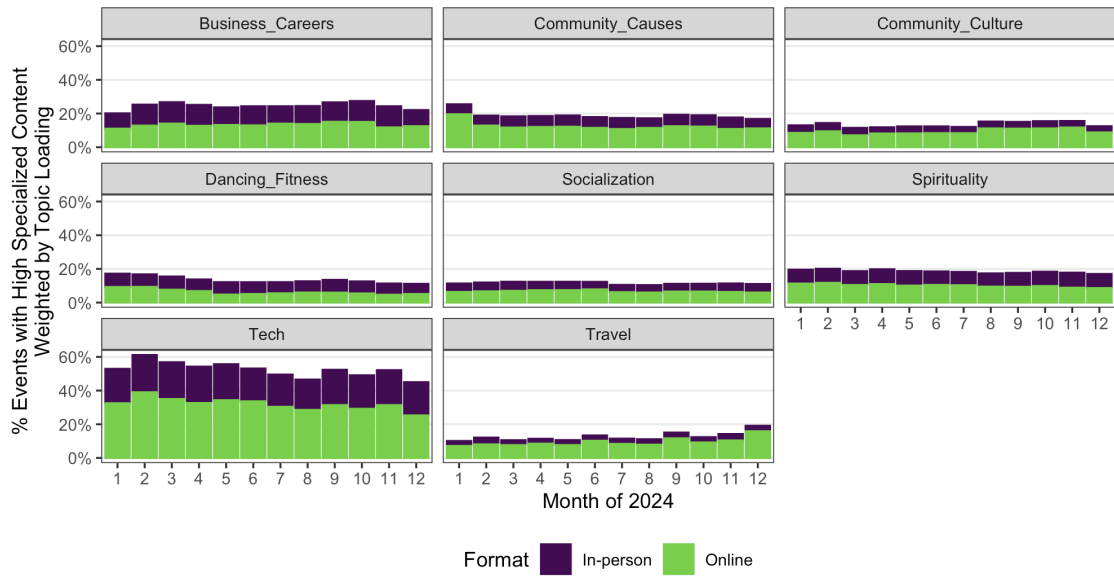


Figure W5: Monthly marginal CATTs and frequencies of specialized events weighted by topic loadings.

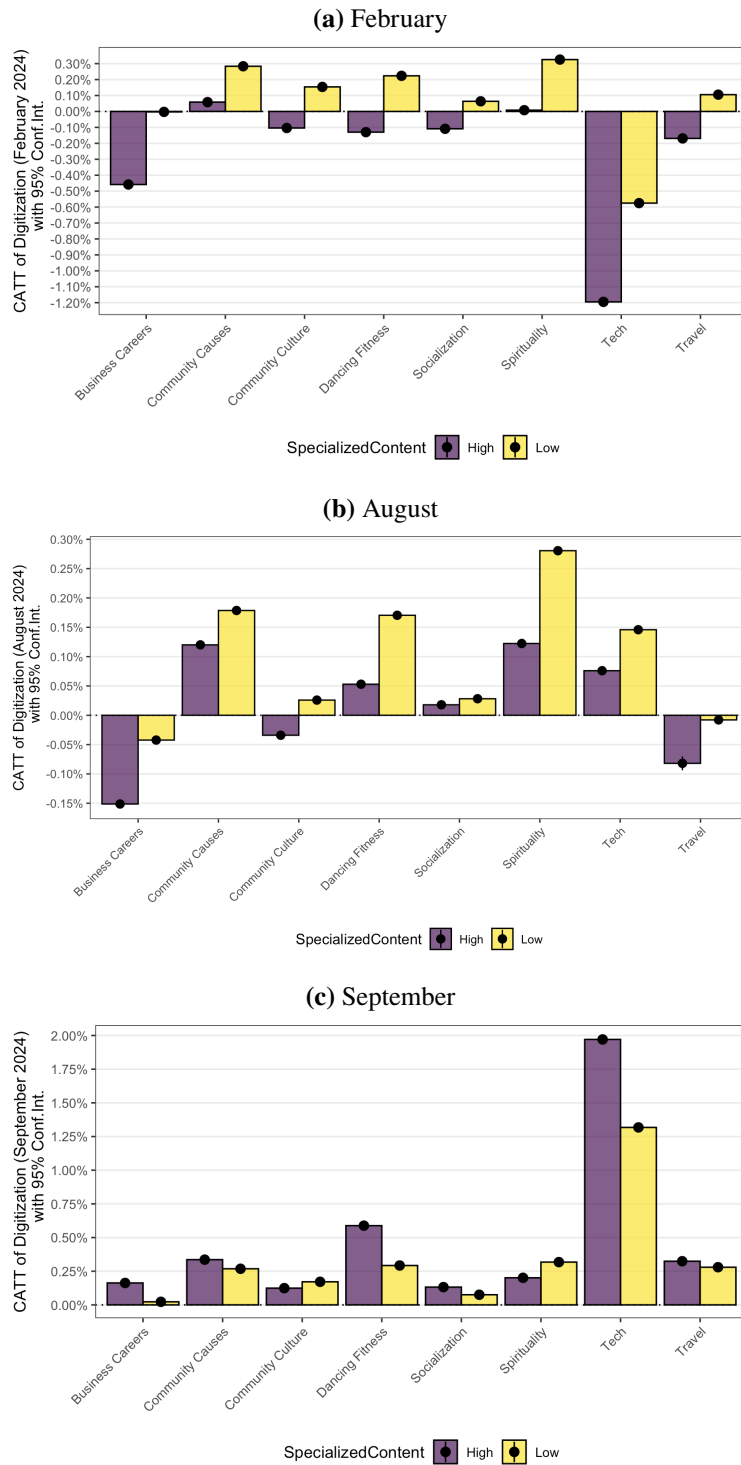


Figure W6: Marginal CATTs by topic and Specialized Knowledge, with 95% confidence intervals.

Web Appendix F: Metrics of Diversity and Popularity of Events

To perform our exploratory analysis of individual-level benefits of digitization, we calculate a series of useful indicators. First, for each member, we calculate the share of digitized events they attended in 2024. To calculate this share, we consider only events to which members responded positively (as opposed to all events to which they were exposed). Using the same subset of events, we calculate individual-level Rao’s quadratic entropy (Purdon et al., 2022; Zielinski, 2022). This metric describes the average diversity of events/groups a member attended in 2024, ranging from 0 (the member attended very homogeneous events and groups) to 1 (the member attended very diverse events and groups). Second, for each event, we describe event popularity as the average number of positive RSVPs per event. Finally, for each group, we measure group popularity as the number of group members. We take the inverse of these popularity scores such that the inverse popularity score will be more positive when a group or event is relatively less popular (i.e., “long tail” group or event).

Web Appendix G: Robustness Checks

G.1 Regression Sensitivity Analysis

We evaluate the sensitivity of our regression estimates to the inclusion of unobserved confounders, as described in Cinelli and Hazlett 2020. The intuition behind the regression sensitivity analysis is to quantify (i) “the minimum strength of association that unobserved confounding would need to have, both with the treatment and with the outcome, to change the research conclusions”, and (ii) “how strongly confounders explaining all the residual outcome variation would have to be associated with the treatment to eliminate the estimated effect” (Cinelli & Hazlett, 2020). We adopt the default convention of reporting the strength of confounding necessary to bring the observed effects to zero using a statistical significance level of 5%. More precisely, we compute the following sensitivity statistics that summarize the overall robustness of the estimates to unobserved confounding:

Partial R^2 ($R^2_{Y \sim D|X}$). The partial R^2 of the treatment with the outcome measures the percentage of variation of the outcome explained by the treatment, after taking into account the part explained by the remaining covariates. The variation in the outcome explained by the treatment determines how strong unobserved confounders would need to be in order to nullify the observed effect. In practice, if confounders explained 100% of the residual variance of the outcome, they would need to explain at least [Partial R^2]% of the residual variance of the treatment in order to fully account for the observed estimated effect.

Robustness Value ($RV_{q=1}$). The robustness value for the point estimate describes the minimum strength of association (measured in terms of partial R^2) that unobserved confounders would need to have, both with the treatment and with the outcome, to nullify the effect. This means that unobserved confounders would need to explain at least [RV]% of the residual variance of both the treatment and the outcome to bring your estimated effect to 0.

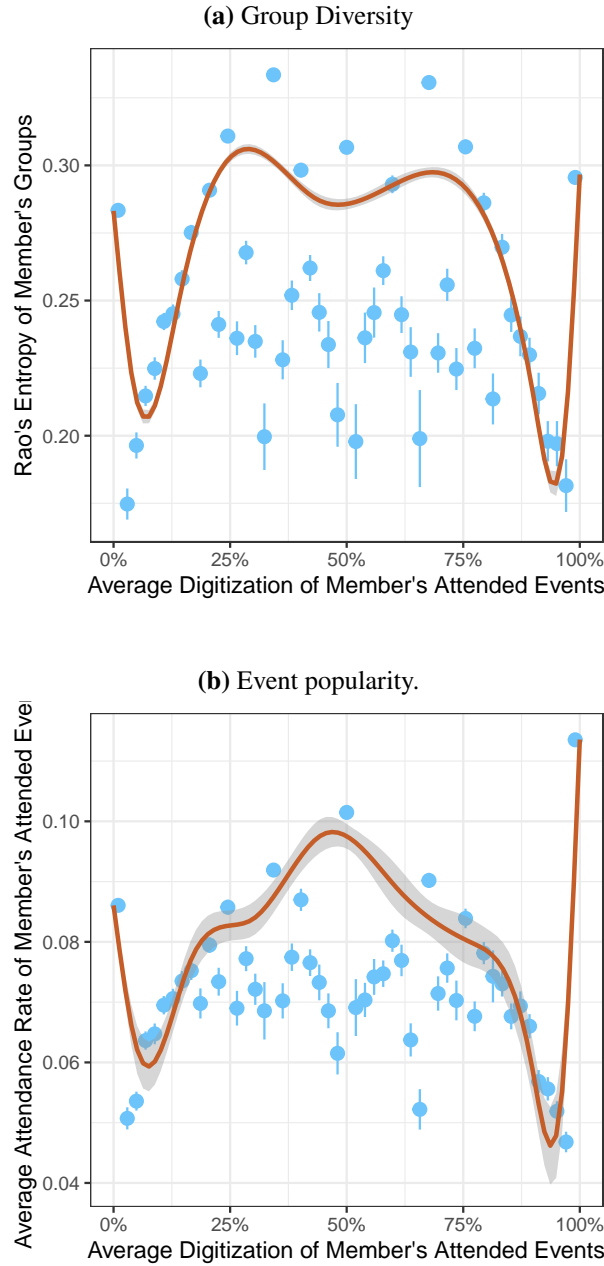


Figure W7: Average Digitization of Member's Attended Events, and Diversity and Popularity of Events and Groups (Group Diversity and Event Popularity Plots).

Robustness Value of t-value ($RV_{q=1, \alpha=0.05}$). The robustness value for the t-value describes the minimum strength of association (measured in terms of partial R^2) that brings the point estimate not to exactly zero but rather into a range where it is no longer statistically different from 0. This means that unobserved confounders would need to explain at least [RV-t]% of the residual variance of both the treatment and the outcome for the null hypothesis that the true treatment effect is equal to 0 to not be rejected at the significance level of 0.05.

Table W7 reports the results of the sensitivity analysis. We note that the partial R^2 of the strongest observed covariate in our main model (i.e. the covariate corresponding to the “Weekend event” indicator) is 0.0022%. Hence, out of the twelve digitization effects, five (“Community/Culture”, “Dancing/Fitness”, “Spirituality”, “Tech”, “Travel”) are especially robust to unobserved confounders, unless an omitted variable is (i) as strongly tied to the treatment as our best observed predictor and (ii) explains significantly more variance in event registrations than any covariate we measure. For example, unobserved confounders would need to explain at least 0.18% of the residual variance of both the “Community/Culture” digitization treatment and the RSVP outcome to bring the estimated effect of digitizing “Community/Culture” events to 0. Considering that our strongest observed covariate explains 0.0022% of the residual variance, an unobserved confounder would need to be 81 times stronger than the strongest observed variable to nullify the effect.

The remaining six significant coefficients (“Business/Careers”, “Community/Causes”, “Social Interactivity”, “Physicality”, “Skill Development”, and “Specialized Knowledge”) may be more sensitive to omitted variable bias. Particularly, for “Community/Causes”, “Social Interactivity”, “Skill Development”, “Specialized Knowledge”, and “Travel” effects, a confounder would need to be 30–60 times stronger than our strongest observed covariate to erase statistical significance. For digitization effects in “Business/Careers” and high “Physicality” events, a hidden confounder would only need to be about 10 times stronger than the “weekend” status covariate to alter statistical significance. Finally, the digitization effect across “Socialization” events is not significant even before considering threats from omitted variable bias.

Outcome: Y_{ie}						
Treatment:	Est.	S.E.	t-value	$R^2_{Y \sim D X}$	$RV_{q=1}$	$RV_{q=1, \alpha=0.05}$
<i>Online × Business/Careers</i>	-0.001	0.000	-5.464	0%	0.04%	0.02%
<i>Online × Community/Causes</i>	0.002	0.000	16.767	0%	0.11%	0.1%
<i>Online × Community/Culture</i>	0.002	0.000	30.049	0%	0.2%	0.18%
<i>Online × Dancing/Fitness</i>	0.003	0.000	21.57	0%	0.14%	0.13%
<i>Online × Social Interactivity</i>	-0.001	0.000	-23.111	0%	0.2%	0.1%
<i>Online × Physicality</i>	0.0003	0.000	5.263	0%	0.03%	0.02%
<i>Online × Skill Development</i>	0.001	0.000	19.819	0%	0.1%	0.1%
<i>Online × Socialization</i>	0.0001	0.000	1.627	0%	0.01%	0%
<i>Online × Specialized Knowledge</i>	-0.001	0.000	-12.388	0%	0.1%	0.1%
<i>Online × Spirituality</i>	0.002	0.000	30.103	0%	0.2%	0.18%
<i>Online × Tech</i>	-0.004	0.000	-28.12	0%	0.18%	0.17%
<i>Online × Travel</i>	0.002	0.000	11.318	0%	0.07%	0.06%
df = 231227843						

Table W7: Sensitivity analysis results.

G.2 Placebo Treatment Assignment

In addition to assessing the sensitivity of the causal estimates, we also perform a placebo analysis in which the treatment is randomly assigned to individuals following a Bernoulli distribution. This is also called “falsification test” in Haw et al. (2023). The placebo treatment analysis should rule out the possibility that idiosyncratic factors drive the effects and that significant results are spurious, as approximately 1 in 20 estimates may appear significant due to random variation alone at a 5% significance level.

The results of the sensitivity and placebo analyses, combined, indicate the overall robustness of the findings in the face of the threat of unobserved confounding. Figure W8 shows that the only coefficient

exhibiting significance at 5% confidence level under placebo treatment assignment corresponds to the “Specialized Knowledge” term, while the rest of the coefficients are not significantly different from 0. Even if the coefficient corresponding to the digitization effect in “Specialized Knowledge” events is statistically significant, its magnitude is very small and close to zero. While notable, this isolated significant placebo result may represent normal random variability rather than evidence of systematic bias.

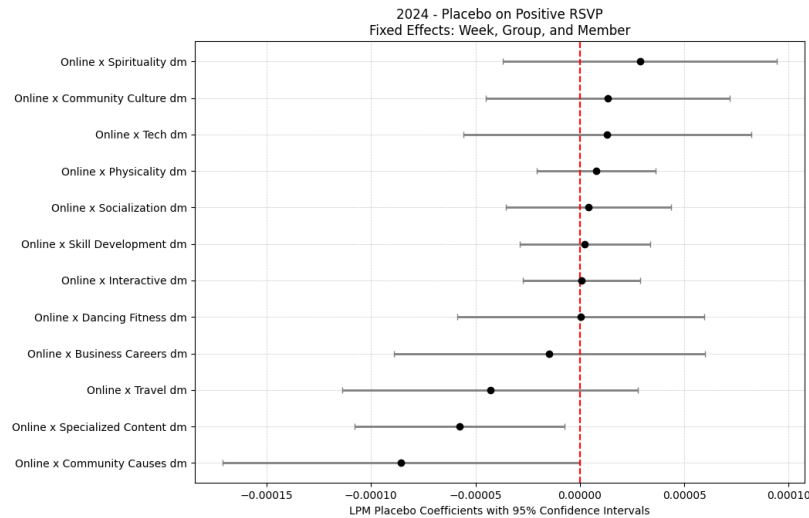


Figure W8: LPM coefficient estimates – Placebo treatment assignment.

References

- Ali, M. S., Groenwold, R. H. H., Pestman, W. R., Belitser, S. V., Roes, K. C. B., Hoes, A. W., de Boer, A., & Klungel, O. H. (2014). Propensity score balance measures in pharmacoepidemiology: A simulation study. *Pharmacoepidemiology and drug safety*, 23(8), 802–811.
- Arun, R., Suresh, V., Veni, M., & Narasimha, M. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. *Proceedings of the 14th Pacific-Asia Conference*, 391–402.
- Belitser, S. V., Martens, E. P., Pestman, W. R., Groenwold, R. H. H., De Boer, A., & Klungel, O. H. (2011). Measuring balance and model selection in propensity score methods. *Pharmacoepidemiology and drug safety*, 20(11), 1115–1129.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive lda model selection. *Neurocomputing*, 72(7-9), 1775–1781.
- Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39–67.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1), 61–84.

- Gelman, A., & Su, Y.-S. (2024). *Arm: Data analysis using regression and multilevel/hierarchical models* [R package version 1.14-4]. <https://CRAN.R-project.org/package=arm>
- Haw, I.-M., Swink, M., & Zhang, W. (2023). The role of direct equity ownership in supply chains. *Journal of Operations Management*, 69(4), 586–615.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>
- Meetup. (2020). Popular topics (Archived). <https://web.archive.org/web/20200101035155/https://www.meetup.com/topics/>
- Meetup API. (2022). API Docs Schema | MeetUp. <https://www.meetup.com/api/schema/#graphQl-schema>
- Murzintcev, N. (2020). *Ldatuning: Tuning of the latent dirichlet allocation models parameters* [R package version 1.0.2]. <https://CRAN.R-project.org/package=ldatuning>
- OpenAI. (2025, April). GPT-4.1 Prompting Guide. https://cookbook.openai.com/examples/gpt4-1_prompting_guide
- Purdon, A., Mole, M. A., Selier, J., Kruger, J., Mafumo, H., & Olivier, P. I. (2022). Using the rao's q diversity index as an indicator of protected area effectiveness in conserving biodiversity. *Ecological Informatics*, 72, 101920.
- Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of clinical epidemiology*, 66(8), S84–S90.
- US Census Bureau. (2022, May). City and town population totals: 2020-2021. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>
- Zielinski, A. (2022). Impact of model settings on the text-based rao diversity index. *Scientometrics*, 127(12), 7751–7768.