Maier et al. *EPJ Data Science* (2026) 15:30
https://doi.org/10.1140/epjds/s13688-026-00642-5

EPJ.org

EPJ Data Science

RESEARCH Open Access

Online interaction and identity cue adoption: a large-scale analysis of hashtag adoption on Twitter

Lena Maier[1*], Daniel Matter[1], Jason Jeffrey Jones[2] and Jürgen Pfeffer[1]

Handling Editor: Santo Fortunato

*Correspondence:
lena.maier@tum.de
[1]School of Social Science and Technology, Technical University of Munich, Richard-Wagner-Str. 1, Munich, 80333, Germany
Full list of author information is available at the end of the article

Abstract

With online interactions becoming an integral part of everyday social life, there is a need to better understand the relationship between social interaction and identity expression in digital environments. This study examines whether online self-presentation, specifically the adoption of identity-related hashtags in Twitter bios, is systematically associated with observable interaction patterns. Utilizing a large-scale dataset encompassing approximately 63 million Twitter profiles and 292 million interactions, we implement a matched quasi-experimental design comparing users who interacted with hashtag-bearing accounts to similar users who did not. Our results show that users who interact with others who feature particular hashtags in their bios subsequently adopt those hashtags at substantially higher rates. Adoption likelihood increases with the number of interaction partners displaying a given hashtag, though with diminishing marginal effects, and the magnitude of these associations varies across identity content categories, being strongest for fan communities and weakest for political hashtags. These patterns are consistent with theories of social influence and suggest that online self-presentation is systematically related to the social contexts in which users are embedded. However, given the observational design of this study, alternative explanations for the observed associations cannot be fully excluded. Future experimental research is needed to clarify the mechanisms underlying these associations and to examine their implications for community formation and the dynamics of collective identity in online environments.

Keywords: Online self-presentation; Online identity signaling; Social influence; Social contagion; Online social networks

1 Introduction

Social networking sites have become central venues for everyday social life, allowing individuals to connect with others and present themselves to others, much as they do offline. Such self-presentation can be understood as a form of identity signaling. Identity is typically conceptualized at three levels: the personal level, encompassing self-definitions based on individual characteristics, beliefs, and values; the relational level, reflecting roles and

Springer

positions assumed in interactions with others; and the collective level, referring to identification with broader social groups or categories [1–3]. Social media platforms such as Twitter, Instagram, or TikTok provide abundant opportunities to enact these identity layers through self-presentation in user profiles using means such as textual biographies (bios), profile pictures, and avatars [4–8]. In particular, bios enable self-labeling via words, emojis, and hashtags that communicate identity cues such as personal traits, social roles, or group memberships [6–8]. Previous research has shown that such online self-presentation can have far-reaching consequences, influencing, for example, how others engage with the presenting individual [9] and even feeding back into the individual's own self-concept and offline behavior [4, 10–12]. However, the existing body of work on the formation and implications of profile-based online self-presentation has primarily focused on individual motives. The socially dynamic mechanisms through which online self-presentation evolves remain less well understood.

Theories of social influence suggest that behaviors, attitudes, and even self-concepts can be shaped through interpersonal interaction, via mechanisms such as compliance, identification, and internalization [13–15]. When affect, attitudes, or behaviors spread between individuals, particularly in the absence of deliberate attempts to influence, this process is often described as social contagion [16]. Prior research has shown that i) social contagion occurring in offline interactions also persists in similar settings in online environments; ii) the magnitude of exposure via social interaction influences the level of contagion; and iii) the strength of these processes can vary depending on the nature of the behavior or attitude being transmitted.

In offline contexts, numerous empirical studies have documented such forms of social influence and contagion. For instance, in an experimental setting, Barsade [17] showed that emotional contagion occurs in groups, such that people continuously transmit their moods to others, influencing affect, judgments, and behavior. Similar mechanisms have been documented in online settings. In a large-scale field experiment on Facebook, Kramer et al. [18], for example, demonstrated that emotional states can spread via online interactions, leading people to experience the same emotions as an interaction partner, without awareness. Further, Bond et al. [19] showed that political self-expression by Facebook friends influenced individuals' own political expression and even their voting behavior.

As stated before, social influence processes may intensify with exposure: the more individuals an actor observes exhibiting a particular attribute, opinion, or behavior, the greater the likelihood of adoption [20, 21]. Several mechanisms may underlie such patterns: multiple exposures may provide cumulative evidence that a behavior is valuable or appropriate, a process known as social learning [22]; alternatively, observing many others adopt may signal normative expectations, triggering conformity to perceived group standards [23]. Additionally, the nature and social meaning of the behavior or attribute involved can further moderate patterns of social influence. Some behaviors may be more readily adopted after limited exposure, while others—particularly those that are more publicly visible or that carry potential social costs—may require greater reinforcement before adoption occurs [24, 25]. For instance, research on collective action has shown that behaviors involving political participation or public protest, which carry risks of social sanction or personal harm, tend to spread through networks differently than less risky behaviors [26]. Such

content-dependent variation suggests that the relationship between social exposure and adoption may differ across domains of identity expression.

Although prior research has demonstrated various mechanisms of social influence and contagion in online settings, it remains unclear whether and how these processes extend to individuals' self-presentation. In particular, little is known about whether individuals adjust their online self-presentation in response to the online self-presentation of others with whom they interact, and how such adjustments depend on the level of exposure and the content of the identity being expressed. Examining these dynamics can provide valuable insights into broader processes of social interaction and identity signaling in online environments, such as the formation of collective identities online or the emergence of sociopolitical polarization. The present study engages with these questions by examining whether profile-based self-presentation in the form of Twitter bios is systematically associated with observable interaction patterns. Specifically, drawing on the theoretical framework outlined above, we test the hypothesis that interaction with others who include a particular identity cue in their bio is associated with an increased likelihood that an individual subsequently adopts the same identity cue in their own bio (H1). We further examine whether this adoption likelihood increases with the observable number of interaction partners who already display the relevant identity cue in their self-presentation (H2). Finally, we assess whether adoption patterns differ across different types of identity content (H3).

By integrating perspectives on identity, self-presentation, and social influence, this study provides a systematic empirical analysis of profile-based self-presentation as it unfolds within online interaction networks. Tracking sequential changes in Twitter bios, we show that users' self-presentational choices are systematically associated with (i) the self-presentations of their observable interaction partners, (ii) the degree of observable interaction partners bearing a certain identity cue, and (iii) differences across types of identity content. Given the observational nature of the data, these relationships are best interpreted as descriptive associations that are compatible with social influence, but cannot rule out alternative explanations such as selection and unobserved confounding (this will be discussed in detail in Sect. 2.2.2). In doing so, the study shifts attention from individual motives to the relational and networked contexts in which online identity signaling takes place, and motivates future research using designs better suited to disentangling the underlying mechanisms.

2 Methods

2.1 Dataset

To test the three hypotheses, this study drew on a large-scale dataset comprising all public posts and the corresponding user profiles on Twitter (now X[1]) over two 24-hour snapshots: September 21, 2022 (day 1), and December 7, 2022 (day 2). The data collection procedure is documented in detail by [27]. The dataset contains approximately 63 million unique user profiles, including the information provided in their profile bios. These user profiles correspond to all Twitter accounts that either published original content or were referenced through retweets, replies, or quotes on at least one of the two days.

[1] At the time of data collection, the platform was still named Twitter. For consistency, we refer to the platform as Twitter throughout this article.

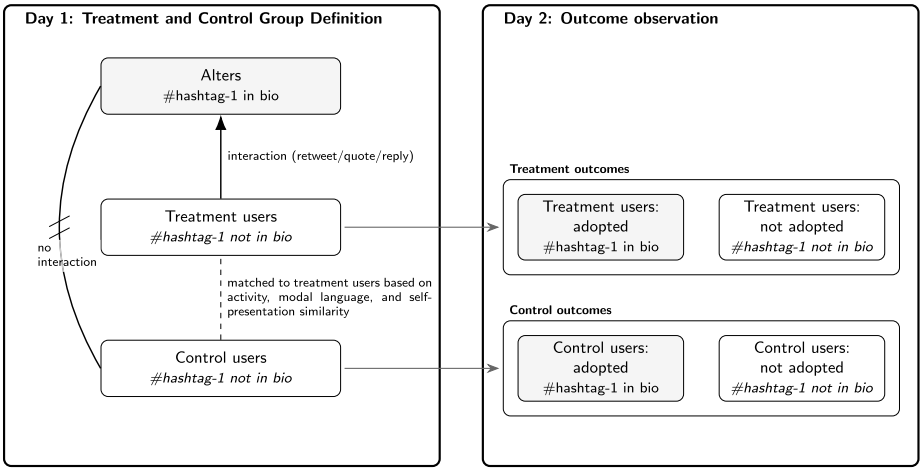

Figure 1 Quasi-experimental setup for hashtag adoption illustrated with example *#hashtag-1*

Retweets, replies, and quotes represent interactional ties between users. Within each 24-hour window, the dataset captures all public retweets, replies, and quotes, providing a high-coverage snapshot of observable interaction ties and contemporaneous self-presentation on the platform. In total, the dataset includes around 292 million interactions in the form of retweets, replies, and quotes.

2.2 Quasi-experimental setup

To examine whether profile-based self-presentation is associated with prior online interaction, we implemented a matched quasi-experimental design linking the adoption of new identity cues in users' Twitter bios to their observable interactions with other users. The setup is depicted in Fig. 1 and explained in detail in the following subsections.

Other users' bios are one potential source of new content for one's own self-presentation. On Twitter, individuals have several ways to view the content of other users' bios. On the web, the author's profile picture commands its own column to the top-left of each tweet, and hovering the cursor over the profile picture summons a mini-profile that includes the bio. The same hover-over mini-profile appears when one places the cursor over the user's display name or platform username. The bio is also prominent on the user page that follows from clicking (or tapping on the mobile app) on the profile picture, name, or username. However, direct measures of profile viewing are not available in our data. We therefore use user-to-user interactions—retweets, replies, and quotes—as an observable proxy for exposure to others' self-presentational content.

Building on this logic, users' bios were first observed on day 1, prior to any measured interaction, and then re-observed on day 2 to assess changes in bio identity cues following interactions that occurred on day 1. For this, we implemented 817 quasi-experiments for different identity cues in which we compared the adoption of a particular cue in treatment users' bios with the adoption by control users. Treatment users were defined as those who did not include the focal cue in their bio on day 1 prior to any interaction and who had at least one interaction with another user (referred to as an alter) on day 1 who had that cue in their bio. The control group consisted of users similar to the treatment users who also

did not possess the respective cue in their bio on day 1 prior to any interaction, but who did not interact on day 1 with any alter having that cue in their bio.

2.2.1 Measurement of identity cues

A prominent form of identity cues are hashtags. Hashtags consist of words or word groups preceded by a hash symbol (#) and serve as markers for particular themes or content. In user bios, hashtags frequently function as identity labels, indexing personal attributes (e.g., #vegan), relational roles (e.g., #dad), or collective affiliations (e.g., #LGBTQ). They thus operate both as signals of personal identity and as references to shared social meaning, often indicating a desire to affiliate with specific social groups [6, 8]. From a methodological perspective, hashtags are well-suited for cross-linguistic analysis. Their standardized format enables consistent extraction from bios written in diverse languages, allowing analyses that are not limited to English-speaking users. For this reason, we operationalized the adoption of identity cues as the adoption of hashtags in user bios. Hashtags were extracted using the sayrer/twitter-text Rust port [28], which implements the same detection logic applied internally by Twitter [29]. This ensured compatibility with platform definitions and robust language-independent extraction. To reduce noise and ensure sufficient prevalence, we restricted the analysis to hashtags used by at least 100 users on day 1 within their respective language community, resulting in 822 candidate hashtags. Of these, 817 hashtags yielded viable treatment and control groups after applying the matching procedure described below. Hence, we conducted 817 quasi-experiments, one for each of these hashtags.

2.2.2 Definition of treatment and control group

For each hashtag, treatment users were defined as those who (a) did not use the hashtag in their own bio on day 1 prior to any interaction, (b) had at least one observable interaction on day 1 with another user whose bio contained the focal hashtag, and (c) could be matched with at least one suitable control user (as detailed below). Interaction was defined as retweeting, quoting, or replying to a post authored by a user whose bio contained the respective hashtag.

Identification challenge To construct a comparable counterfactual for each treated user, we used a matched sampling design aimed at reducing confounding from homophily and other observable selection mechanisms. Homophily, the tendency of individuals to associate with others who share similar characteristics [30], poses a central identification challenge in observational studies of social influence because treatment through social interaction is not randomly assigned. In the context of this study, homophily implies that users who interact with hashtag-bearing accounts may already have latent interest in the associated topics, which could independently drive both the interaction and subsequent hashtag adoption. Consequently, observed associations between interaction and adoption may reflect pre-existing similarities rather than influence through social interaction [31]. Beyond homophily, additional confounders may include differential activity levels (more active users may be more likely both to interact and to modify their bios), exposure to external events (e.g., a game release simultaneously driving interaction with gaming accounts and adoption of gaming hashtags), and algorithmic curation (Twitter's recommendation system may expose similar users to similar content). To mitigate confounding arising from

these mechanisms, we matched treated users to control users who were similarly active and similar along a set of observable characteristics measured on day 1, but who were not treated on the focal hashtag through interaction. We then test our hypotheses using multivariate regressions that include these observable confounders as control variables (see Statistical analysis). This design reduces bias due to observable homophily and selection on measured covariates.

However, as shown by Shalizi and Thomas [32], homophily, shared environments, and influence are generically confounded in observational network data: without randomization, valid instruments, or strong structural assumptions, matching alone cannot fully disentangle social influence from these alternative explanations. Our data do not permit any of these approaches, as we lack experimental control for randomization, credible instruments for instrumental variable (IV) estimation, and the justification for the strong parametric assumptions required for structural models. In addition, our setting may involve further sources of unobserved confounding, such as common exposure to external events or platform-level recommendation effects, which cannot be completely addressed with the available data. We therefore interpret our estimates as conditional associations under reduced observable confounding, and discuss the remaining limitations and identifying assumptions in detail in Limitations.

Control group matching For each treatment user, control users were defined as those who (a) did not yet use the hashtag in their bio on day 1, (b) were not exposed on day 1 to the focal hashtag through interaction, and (c) were sufficiently similar to the treatment user based on the criteria described below. The pool of potential controls was first restricted to users who, like treatment users, exhibited observable activity on day 1 (posting, retweeting, quoting, or replying), ensuring comparability in basic engagement levels. In addition, matching was performed within modal language: control users were required to share the same modal language as the treated user.

To generate a candidate pool that captures multiple dimensions of user similarity beyond baseline activity and language, we employed a hybrid retrieval strategy that combines semantic and behavioral signals. First, we embedded free-text bio descriptions using a multilingual sentence embedding model. All bio embeddings were L2-normalized, and for each treated user we retrieved up to 1000 nearest neighbors by cosine similarity within the same language. To enable efficient retrieval over millions of users, we used Facebook AI Similarity Search (FAISS), a library for fast approximate nearest-neighbor search in high-dimensional spaces. This step identified users with broadly similar self-descriptions based on the semantic content of their bios. Second, to ensure adequate representation of users with overlapping hashtag usage patterns—who might not rank highly on embedding similarity alone if their bio text differs in other content—we additionally computed Jaccard similarity in bio hashtags for all candidate pairs. Jaccard similarity measures the overlap between two sets as the size of their intersection divided by the size of their union; here, it captures the proportion of shared hashtags between a treated user's bio and a candidate control's bio on day 1. We retained candidates with either high embedding similarity (top 1000 by cosine distance) or high Jaccard similarity (requiring at least a Jaccard similarity ≥ 0.1), ensuring all retained candidates had scores on both dimensions. The union of embedding-based and Jaccard-based candidates formed the general candidate pool for each treated user.

To further capture topical alignment between users and the focal hashtag, we constructed hashtag embeddings based on co-occurrence patterns within user bios. For each language, we built a hashtag co-occurrence matrix counting how frequently pairs of hashtags appeared together in the same bio, applied Positive Pointwise Mutual Information (PPMI) weighting, and performed truncated singular value decomposition (SVD) to obtain dense hashtag embeddings [33]. These embeddings represent hashtags that frequently co-occur in bios as being closer in a shared semantic space. Using these embeddings, we calculated a hashtag relevance score for each user–hashtag combination. For a given user and focal hashtag, the relevance score captures the maximum semantic similarity between the focal hashtag and any hashtag already present in the user's bio on day 1. Higher values indicate that the user's existing self-presentation is already closely aligned with the topic signaled by the focal hashtag, while lower values indicate little or no observable topical overlap. In the matching procedure, we minimized the absolute difference in relevance scores between treatment users and their control candidates to ensure comparability in pre-existing topical alignment. We additionally included the relevance score as a covariate in our regression models (see Statistical analysis) to control for residual differences in observable topical interest. The relevance score requires reliable hashtag embeddings, which in turn depend on sufficient co-occurrence data. For languages with sparse hashtag usage, reliable embeddings could not be estimated. As a result, relevance scores were available for 817 of 822 hashtags; our analyses focus on these 817 hashtags.

Final control selection was performed separately for each hashtag, using the previously described candidate pools and their relevance scores for that hashtag. The matching procedure involves several design choices that can affect covariate balance and statistical power, including which similarity dimensions to prioritize when ranking candidates, how many controls to select per treated user (k), whether matching is performed with or without replacement, and whether to impose maximum distance thresholds (calipers) on specific covariates. Prior work emphasizes that no single design is universally optimal; selection should be guided by balance diagnostics and robustness considerations [34]. We therefore evaluated multiple designs varying along these dimensions.

For each design, candidates were ranked by a composite distance score—a weighted combination of selected similarity components (e.g., bio embedding similarity and relevance score difference)—and the k closest controls were selected for each treated user. We evaluated matching both with and without replacement. Matching with replacement allows controls to be reused across multiple treated users, maximizing treatment retention but potentially introducing dependence if individual controls are matched many times. Matching without replacement assigns each control to at most one treated user within a hashtag experiment; to prevent treated users with large candidate pools from exhausting available controls, we processed treated users in order of ascending candidate pool size, prioritizing those with fewer options. Treated users with fewer than k eligible controls after caliper filtering were excluded from analysis. Supplementary Note S1 presents the full overview of the design specifications and their robustness assessment. Supplementary Table S1 specifically presents an overview of the different matching design specifications that we implemented.

Matching quality was assessed using standardized mean differences (SMD) between treated users and matched controls on key covariates. Key covariates included relevance score, log-transformed activity rate (=total tweet count of the user/account age in days),

log-transformed follower count, log-transformed following count, and log-transformed account age in days. All covariates were measured on day 1 prior to any observed interaction. As a rule of thumb, $|SMD|$ below 0.10 is often interpreted as indicating good balance. We additionally evaluated the treatment retention rate, defined as the proportion of treated users successfully matched and, for designs using matching with replacement, the distribution of control reuse counts. Supplementary Table S2 presents the balance diagnostics across matching designs. Based on these diagnostics, we selected as our primary specification a design using k = 1 control users per treated user, matching without replacement, prioritizing control users with the smallest absolute difference in relevance scores compared to the respective treatment user, and imposing a caliper of 0.3 on log-transformed activity rate to ensure comparability in posting behavior. This chosen design achieved good covariate balance (mean $|SMD| = 0.1$) while maintaining high treatment retention (96.5%). We chose matching without replacement because matching with replacement resulted in high control reuse (mean control matched to 170 treated users), raising concerns about dependence in standard error estimation.

To assess robustness of association estimates, we re-estimated the main regression model for H1, which is described in Statistical analysis, under the alternative designs described in Supplementary Note S1. The treatment effect was significant and positive across all designs, with coefficients ranging from 1.79 to 2.08 (corresponding to odds ratios of 6.01 to 8.01). Effect sizes were smaller for k = 1 compared to k = 5, and smaller for designs with the activity caliper compared to those without, consistent with better covariate balance reducing residual confounding. Full results across designs are provided in Supplementary Table S3.

The final matched sample comprised 5,932,354 observations (2,966,177 matched pairs) across 817 hashtag experiments.

2.3 Identity content classification

To assess variation across identity content (H3), hashtags were grouped into eight thematic categories: *Politics*, *Celebrities*, *TV & music fandom*, *Sports fandom*, *Finance & Tech*, *Gaming fandom*, *Roles/traits/interests*, and *Other*. The categorization was performed inductively using a predefined dictionary (see Supplementary Note S2).

2.4 Statistical analysis

To examine whether interaction with hashtag-bearing users is associated with subsequent hashtag adoption, we combined descriptive comparisons with regression-based analyses applied to the matched samples constructed as described above. All analyses were conducted using the primary matching design selected in Sect. 2.2.2.

*2.4.1 Analyses for H*1

For H1, as an initial descriptive assessment, we compare adoption rates between treated users and their matched controls across all hashtag experiments. Differences in adoption rates are evaluated using two-sided Mann-Whitney U-tests. We rely on non-parametric tests at this stage because adoption rates are highly skewed across hashtag experiments and sample sizes are large, making distributional assumptions inappropriate.

Beyond descriptives, we estimate associations conditional on observable covariates and hashtag fixed effects. We estimate a logistic regression model predicting whether a user

adopted a given hashtag in their bio on day 2. For H1, the primary independent variable is a binary indicator of treatment, defined as interaction on day 1 with at least one alter whose bio contained the focal hashtag. The model includes covariates measured on day 1 capturing (1) user's activity and connectedness on Twitter, which is log-transformed activity score calculated as tweet count per account age in days, log-transformed follower count, log-transformed following count, and log-transformed account age in days, (2) characteristics of their self-presentation in their Twitter bio via an indicator for whether the user had any hashtag in their bio and the bio description length in number of characters, and (3) the semantic proximity of their self-presentation to the focal hashtag via the hashtag relevance score. Because activity score, follower count, following count, and account age are right-skewed, these variables are log-transformed to avoid influence of extreme observations. We use $ln(x + 1)$ for transformation so that observations with zero values could be retained while still compressing the upper tail of the distribution. Further, we account for unobserved heterogeneity across hashtags by including hashtag fixed effects, which absorb all between-hashtag differences (e.g., baseline popularity or topical salience). As a result, identification stems from within-hashtag comparisons between treated users and their matched controls, rather than from differences across hashtags. To assess sensitivity of results to model specification choices, we re-estimate the associational link between adoption and interaction under the alternative model specification without hashtag fixed effects. Treatment shows significant association in this specification as well; the detailed results of this alternative specification are presented in Supplementary Table S5.

2.4.2 Analyses for H2

For H2, to analyze dose-response relationships, we first graphically inspect the correlation of the number of distinct interaction partners whose bio contained the focal hashtag on day 1 with the adoption likelihood.

Next, similarly to H1, we estimate a logistic regression model that replaces the binary treatment indicator with the winsorized count of distinct interaction partners whose bios contained the focal hashtag on day 1. We apply winsorization to the interaction partner count to reduce noise from a small number of users with unusually high counts. To allow for nonlinear associations, this model also includes the squared term of the count of interaction partners. The model includes the same covariates and fixed-effect structure as the H1 model described above. We report alternative model specifications without winsorization, without the squared term of the count of interaction partners, and without hashtag fixed effects in Supplementary Tables S6-S8.

2.4.3 Analyses for H3

To assess heterogeneity across identity content categories (H3), we first analyze the distribution of adoption ORs for all hashtags per category using boxplots. Furthermore, we use pairwise two-sided Mann–Whitney U-tests with FDR correction to formally test whether the respective means differ.

To graphically examine variation in dose-response patterns, we plot the relationship between the number of interaction partners displaying the focal hashtag and adoption likelihood, separately for each hashtag category. To formally test the relationships, we estimate the same logistic regression models as for H1 and H2 separately for each hashtag category. For each category, we estimate both the treatment model and the dose-response

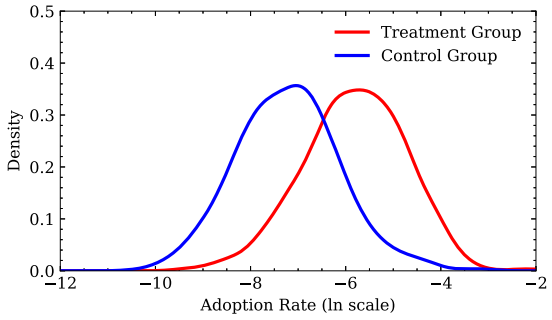

Figure 2 Kernel density estimates of adoption rates for treatment and control users. The x-axis is shown on a natural logarithmic scale

model using identical covariate sets and fixed-effect specifications, allowing effect sizes and dose-response patterns to vary across content types.

As briefly discussed in Sect. 2.2.2, all reported estimates should be interpreted as conditional associations under reduced observable confounding. We will discuss limitations of our setup and potential alternative explanations for the observed patterns in the Discussion.

3 Results

For each hypothesis, we first present descriptive evidence, followed by the results of our multivariate regressions.

3.1 H1: interaction and adoption likelihood

H1 proposed that exposure to others who include a particular identity cue in their bio is associated with an increased likelihood of subsequently adopting the same cue.

Fig. 2 displays kernel density estimates of adoption rates for users in the treatment group and the control group of all hashtags, shown on a natural logarithmic scale. The comparison shows that, consistent with H1, aggregated adoption rates across all 817 hashtag experiments are higher in the treatment group compared to the matched control group. Mean adoption rates are 0.362% for treated users and 0.053% for controls. A two-sided Mann–Whitney U-test confirms that the difference between groups is statistically significant ($p < 0.001$).

These adoption rates are low in absolute terms, which reflects the specificity of our outcome measure. We track adoption of particular hashtags—not bio modifications in general. While approximately 36% of users in our sample modified their bio between day 1 and day 2, adopting any specific hashtag is a rare event; it requires not only editing one's bio but choosing to add that particular identity cue among limitless possible cues. The relevant comparison is therefore not the absolute adoption rate but the relative difference between treated and control users. Treated users adopted focal hashtags at nearly seven times the rate of matched controls, indicating a strong association between interaction and adoption despite the low base rates. Adoption rates per hashtag experiment are reported for a sample of ten hashtag experiments in Supplementary Table S4 and in detail for all hashtag experiments in Supplementary Data 1.

Table 1 Logistic regression testing the relationship between treatment and hashtag adoption

Variable	Coef. (SE)	OR [95% CI]
Interaction status	1.794*** (0.0299)	6.01 [5.66, 6.38]
Relevance score	1.995*** (0.0492)	7.35 [6.66, 8.12]
Log user activity rate	–0.3439*** (0.0109)	0.71 [0.69, 0.72]
Log follower count	0.0791*** (0.0081)	1.08 [1.06, 1.10]
Log following count	0.1016*** (0.0112)	1.11 [1.08, 1.13]
Has hashtag in bio (day 1)	0.2997*** (0.0367)	1.35 [1.25, 1.44]
Length of bio (day 1)	0.0011*** (0.0003)	1.00 [1.00, 1.00]
Log account age	–0.3493*** (0.0091)	0.71 [0.69, 0.72]
Fixed effects	Hashtag	
SE	Clustered by user	
Observations	5,540,940	
Pseudo R^2	0.150	

Notes: Dependent variable is hashtag adoption. Odds ratios are exponentiated coefficients with 95% confidence intervals. Model includes hashtag fixed effects; hashtags with no within-hashtag variation in adoption outcomes or single observations are dropped automatically (232 hashtags; 391,414 observations). Standard errors clustered by user. Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

We next estimate a multivariate logistic regression model estimating the association of hashtag adoption with interaction status, while controlling for users' observable characteristics and including hashtag fixed effects, as specified in Statistical analysis.

Table 1 presents the regression results, including a conversion of coefficients into odds ratios (OR) to ease interpretation. For outcomes with very low baseline probabilities—such as hashtag adoption in our data—odds ratios closely approximate risk ratios and can therefore be interpreted as proportional changes in adoption probabilities [35].

In line with H1, interaction status is positively associated with hashtag adoption and is significant at $p < 0.001$, meaning users who interacted with hashtag-bearing accounts had 6.01 times higher odds of adopting the focal hashtag (95% CI [5.66, 6.38]), controlling for covariates and hashtag fixed effects.

The relevance score, which captures observable topical alignment between users' existing hashtags on day 1 and the focal hashtag based on co-occurrence similarity, shows the strongest association with adoption in the model (OR = 7.35, 95% CI [6.66, 8.12]). This indicates that pre-existing topical interest is a major contributor to adoption patterns: users whose self-presentation was already semantically close to the focal hashtag on day 1 were considerably more likely to adopt it. At the same time, interaction status remains a substantial predictor even when controlling for this measure of topical alignment, suggesting that the association between interaction and adoption is not fully explained by pre-existing self-presentation.

Having any hashtag in the bio on day 1 shows a modest positive association with adoption (OR = 1.35, 95% CI [1.25, 1.44]), suggesting that some users have a general propensity for hashtag-based self-presentation independent of the specific focal hashtag. Users with a higher activity rate show a slightly lower likelihood of adoption (OR = 0.71, 95% CI [0.69, 0.72]), as do users with older accounts (OR = 0.71, 95% CI [0.69, 0.72]), possibly reflecting more stable self-presentation among long-term users.

The remaining covariates, while statistically significant, show odds ratios very close to 1.0, indicating that their practical impact on adoption is negligible; with a sample of over 5.5 million observations, even trivially small effects reach conventional significance thresholds. We therefore focus on the effect sizes.

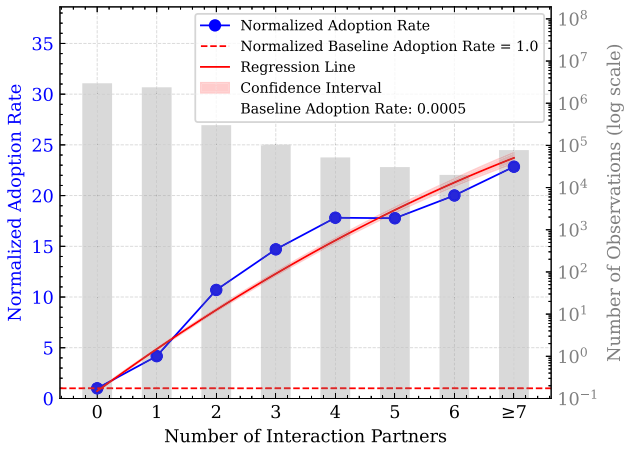

Figure 3 Normalized adoption rates as a function of the number of hashtag-bearing interaction partners. The dashed red line indicates the baseline adoption rate of 0.0005 in the control group, normalized to 1.0. The regression line and shaded area show the fitted quadratic model with 95% confidence intervals

To assess robustness, we re-estimate the model without hashtag fixed effects, allowing for across-hashtag comparisons. Results are consistent with the main specification (see Supplementary Table S5).

Together, these results support H1. Interaction with hashtag-bearing accounts is associated with a substantially higher likelihood of adopting the focal hashtag. Beyond this hypothesized association, the analysis also reveals that pre-existing topical alignment, captured by the relevance score, shows an even stronger association with adoption.

3.2 H2: dose–response relationship for adoption likelihood

H2 proposed that the likelihood of adoption increases with the observable number of interaction partners who already display the relevant identity cue.

Fig. 3 plots normalized adoption rates (relative to the zero-interaction baseline of 0.05%) against the count of hashtag-bearing interaction partners. The figure reveals a strong increase in adoption likelihood with increasing number of interaction partners, with diminishing marginal gains as the number of partners increases.

To formally test this relationship, we estimate a multivariate logistic regression model including the number of hashtag-bearing interaction partners and its squared term to capture potential nonlinearities. Because a small number of users had very high interaction partner counts, we winsorized this variable at the 95th percentile to reduce the influence of extreme observations. The model includes the same covariates as in H1, including the relevance score, as well as hashtag fixed effects to absorb any between-hashtag differences. Table 2 presents the results.

In line with H2, the number of observed interaction partners is positively associated with hashtag adoption (OR = 2.80, 95% CI [2.72, 2.89], $p < 0.001$), controlling for covariates and hashtag fixed effects. The squared term is negative and significant (OR = 0.91, 95% CI [0.91, 0.92]), indicating a concave relationship with diminishing marginal associations at higher numbers of interaction partners.

As in the H1 model, the relevance score remains a strong predictor of adoption (OR = 6.09, 95% CI [5.53, 6.69]). The remaining covariates show similar magnitudes to those in the H1 model.

Table 2 Logistic regression testing the relationship between the number of interaction partners and hashtag adoption

Variable	Coef. (SE)	OR [95% CI]
No. of interaction partners	1.031*** (0.0154)	2.80 [2.72, 2.89]
(No. of interaction partners)^2	−0.0939*** (0.0022)	0.91 [0.91, 0.92]
Relevance score	1.806*** (0.0493)	6.09 [5.53, 6.69]
Log user activity rate	−0.4460*** (0.0110)	0.64 [0.63, 0.65]
Log follower count	0.0932*** (0.0079)	1.10 [1.08, 1.12]
Log following count	0.0832*** (0.0110)	1.09 [1.07, 1.11]
Has hashtag in bio (day 1)	0.3599*** (0.0363)	1.43 [1.33, 1.54]
Length of bio (day 1)	0.0011*** (0.0003)	1.00 [1.00, 1.00]
Log account age	−0.3526*** (0.0091)	0.70 [0.69, 0.72]
Fixed effects	Hashtag	
SE	Clustered by user	
Observations	5,540,940	
Pseudo R^2	0.156	

Notes: Dependent variable is hashtag adoption. The number of interaction partners is winsorized at the 95th percentile. Odds ratios are exponentiated coefficients with 95% confidence intervals. Model includes hashtag fixed effects; hashtags with no within-hashtag variation in adoption outcomes or single observations are dropped automatically (232 hashtags; 391,414 observations). Standard errors clustered by user. Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

To assess robustness, we re-estimate the model under alternative specifications: without winsorization, without the squared term, and without hashtag fixed effects. Across all specifications, the number of interaction partners remains significantly and positively associated with adoption likelihood, though effect sizes are smaller in models without winsorization or without the squared term (see Supplementary Tables S6-S8). This is expected as winsorization focuses estimation on the range where most observations fall, reducing the influence of extreme cases with many interaction partners, while the squared term allows the model to capture the steep initial increase in adoption likelihood that a linear specification would underestimate.

Together, these results support H2. The likelihood of hashtag adoption increases with the observable number of interaction partners displaying the focal hashtag, though the marginal association diminishes at higher exposure levels.

3.3 H3: variation across identity content categories

H3 proposed that adoption patterns differ across different types of identity content.

To test this, hashtags were classified into eight thematic categories: *Politics*, *Celebrities*, *TV & Music fandom*, *Sports fandom*, *Finance & Tech*, *Gaming fandom*, *Roles/traits/interests*, and *Other* (see Sect. 2.3).

Fig. 4 shows the distribution of odds ratios across all hashtag experiments within each category. Although the associations between interaction and adoption are consistently positive across all categories, their magnitudes vary. Celebrity hashtags exhibit the strongest median OR (Median OR = 7.36), followed by Gaming fandom hashtags (Median OR = 7.01). Political hashtags show the weakest associations (Median OR = 3.01). Pairwise two-sided Mann–Whitney U-tests with FDR correction indicate that *Politics* differs significantly from three, and *Celebrities* from five of the seven categories (all $p < 0.05$). Full results of the pairwise comparisons are reported in Supplementary Figure S1.

To descriptively examine whether also dose–response patterns vary across content types, Fig. 5 shows the normalized probability of hashtag adoption as a function of the number of hashtag-bearing interaction partners, separated by category.

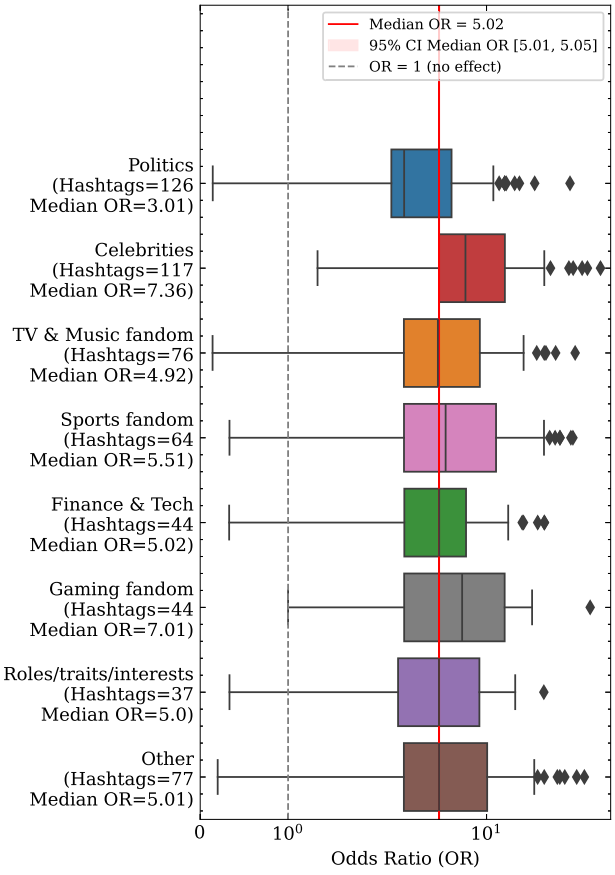

Figure 4 Distribution of odds ratios for hashtag adoption across hashtag experiments per identity content categories. The x-axis is shown on a symmetrical logarithmic scale. The solid vertical line indicates the overall median OR; the dashed vertical line indicates OR = 1. Hashtags with no within-hashtag variation in adoption outcomes or single observations are dropped from the OR calculation

To formally test these patterns, we estimate logistic regression models separately for each category. Table 3 presents selected odds ratios for two model specifications: the treatment model (Model A, with binary interaction status as predictor) and the dose–response model (Model B, with the winsorized number of interaction partners and its squared term as predictors). Both models include hashtag fixed effects and the same covariates as specified in section Statistical analysis. Full regression results are reported in Supplementary Table S9.

The results confirm substantial heterogeneity across categories. In Model A, odds ratios for interaction status range from 3.50 in *Politics* (95% CI [3.13, 3.93]) to 12.97 in *Gaming fandom* (95% CI [8.94, 18.83]). *Politics* shows the weakest association, with a confidence interval that does not overlap with any of the other categories. A similar pattern emerges in the dose–response model (Model B). The per-partner association is steepest for *Gaming fandom* (OR = 6.54, 95% CI [5.01, 8.54]) and *Sports fandom* (OR = 5.41, 95% CI [4.65, 6.28]), while *Politics* shows the flattest relationship (OR = 2.20, 95% CI [2.05, 2.37]). The squared term is smaller than one across all categories (ORs ranging from 0.82 to 0.93), confirming diminishing marginal associations at higher exposure levels regardless of content type.

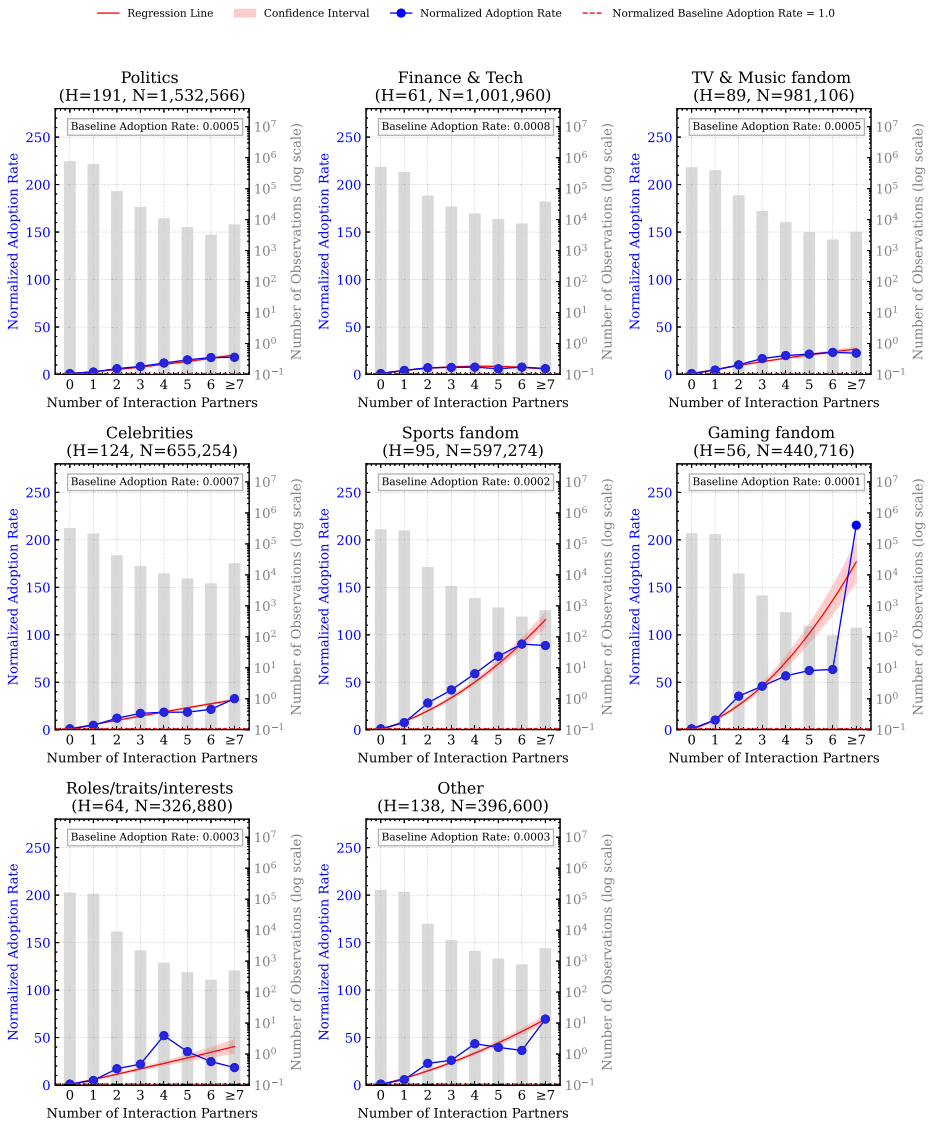

Figure 5 Dose–response relationships across identity content categories. Blue lines show normalized adoption rates relative to each category's baseline rate in the control group (dashed red line at 1.0). Regression lines and shaded areas show the fitted quadratic model with 95% confidence intervals. For each category, the number of considered hashtags (H) and the number of individual observations (N) are given. The graphs are ordered by number of observations in descending order

Notably, the magnitudes of the interaction and relevance-score associations do not vary in parallel across categories. In most categories, the relevance score OR exceeds the interaction status OR, consistent with the pooled models for H1 and H2. This pattern is most pronounced in *Politics*, where the relevance score OR is more than twice as large as the interaction status OR (7.61 vs. 3.50). In contrast, for *Celebrities* and *Other*, this pattern reverses: interaction status shows a larger OR than the relevance score (*Celebrities*: 8.83 vs. 5.89; *Other*: 7.50 vs. 4.23). This variation suggests that the relative importance of pre-existing topical alignment versus interaction differs across content domains. For political hashtags, observable prior interest appears to be the dominant predictor relative to interaction; for celebrity-related hashtags, interaction plays a comparatively larger role. Note,

Table 3 Logistic regression results by identity content category

Category (Observations)	Model A: Interaction status		Model B: Dose–response		
	Int. OR [95% CI]	Rel. OR [95% CI]	#Int. OR [95% CI]	$(\#\text{Int.})^2$ OR [95% CI]	Rel. OR [95% CI]
Politics (1,385,748)	3.50 [3.13, 3.93]	7.61 [6.19, 9.35]	2.20 [2.05, 2.37]	0.93 [0.92, 0.94]	6.18 [5.01, 7.63]
Finance & Tech (985,284)	5.10 [4.52, 5.76]	7.43 [5.99, 9.23]	2.59 [2.43, 2.77]	0.91 [0.90, 0.92]	7.29 [5.87, 9.04]
TV & Music fandom (963,856)	6.51 [5.69, 7.46]	9.48 [7.60, 11.82]	3.35 [3.10, 3.63]	0.89 [0.87, 0.90]	7.15 [5.73, 8.92]
Celebrities (648,172)	8.83 [7.51, 10.38]	5.89 [4.67, 7.42]	2.57 [2.41, 2.73]	0.93 [0.92, 0.94]	3.70 [2.93, 4.68]
Sports fandom (561,932)	10.38 [8.16, 13.21]	9.36 [6.00, 14.58]	5.41 [4.65, 6.28]	0.84 [0.82, 0.87]	6.89 [4.42, 10.74]
Gaming fandom (402,134)	12.97 [8.94, 18.83]	15.82 [9.80, 25.51]	6.54 [5.01, 8.54]	0.82 [0.77, 0.87]	13.85 [8.56, 22.39]
Roles/traits/interests (278,420)	5.52 [4.11, 7.42]	7.71 [5.35, 11.09]	3.60 [2.83, 4.57]	0.86 [0.82, 0.90]	6.49 [4.49, 9.38]
Other (315,394)	7.50 [5.74, 9.79]	4.23 [3.05, 5.88]	3.18 [2.81, 3.61]	0.91 [0.89, 0.92]	4.23 [3.05, 5.89]

Notes: Model A: Interaction status effect. Model B: Dose–response effect. All models include hashtag fixed effects and covariates. Odds ratios (OR) with 95% confidence intervals. Int. = Interaction status; Rel. = Relevance score; #Int. = Number of interaction partners (winsorized). All coefficients are significant at $p < 0.001$. Standard errors clustered by user. Categories ordered by number of observations.

categories with smaller sample sizes, particularly *Gaming fandom* and *Roles/traits/interests*, exhibit wider confidence intervals, reflecting greater uncertainty in the estimates.

Together, these results support H3. The magnitude of association between interaction and adoption varies meaningfully across identity content categories. We discuss potential explanations for this variation in section Discussion.

4 Discussion

The present study examines associations between online social interactions and individuals' self-presentation, focusing on the relationship between interaction with hashtag-bearing accounts and subsequent adoption of those hashtags into users' own Twitter bios. Since self-presentation in digital contexts not only shapes how others interact with an individual [9], but may also affect the individual's own self-perception [4, 10–12], understanding the dynamics behind the formation of online self-presentation provides valuable insights into broader processes of social interaction and identity signaling in online environments. By conceptualizing profile-based self-presentation as an exposure–response process, we test whether interaction with others is associated with a higher likelihood of adopting identity cues. The findings document robust associations between interaction patterns and identity cue adoption. These patterns are consistent with social influence processes; however it should be noted, that the nature of our observational design cannot rule out alternative mechanisms.

Our results show that users who interact with others displaying a given identity cue exhibit substantially higher rates of adopting that cue into their own self-presentation compared to matched control users (H1). Across a large and diverse sample of hashtags, treated users show considerably higher adoption rates than controls. This association holds after controlling for observable user characteristics and expressed topical interest, and is con-

sistent across alternative model specifications. These findings are compatible with theories of social influence emphasizing how interpersonal contact can transmit behaviors and attitudes [13–15]. Notably, the relevance-score—a measure we introduced to capture pre-existing topical alignment of a user's self-presentation with a focal hashtag—shows a stronger association with adoption than interaction status does. This shows that existing self-presentation can be a powerful predictor of adoption itself: users whose existing self-presentation already aligns with the focal hashtag are considerably more likely to adopt it, regardless of interaction. At the same time, interaction status remains a substantial and statistically significant predictor after accounting for this alignment, suggesting that observable interaction patterns carry information about adoption likelihood beyond what is captured by prior self-presentation.

An additional dimension of our analysis concerns the role of the amount of interaction partners for identity cue adoption likelihood. We observe that adoption rates increase with the number of observed interaction partners who display the focal hashtag, with diminishing marginal gains at higher exposure levels (H2). This dose-response pattern aligns with prior literature documenting positive relationships between the number of social contacts exhibiting a behavior and the likelihood of adoption [20, 36]. It must be noted here, that the observed amount of interaction partners results from a one-day-snapshot observation and thus can only capture how many exposures a user experienced at that observed day. Detailed implications of this are discussed in Limitations.

Beyond the overall association between interaction and identity cue adoption, we find substantial heterogeneity in this association across identity content categories (H3). In the regression models, *Gaming fandom* and *Sports fandom* exhibit the strongest associations with interaction status, while political hashtags show the weakest. The relative magnitude of the interaction status and relevance score associations also varies across categories. For most content types, the relevance score OR exceeds the interaction status OR, consistent with the pooled models. This pattern is most pronounced for political hashtags, where the relevance score OR is more than twice as large as the interaction status OR. In contrast, for celebrity-related hashtags, interaction status shows a stronger association than the relevance score. This suggests that both the association between self-presentation and adoption, and the association between interaction and adoption, differ across categories and do so to varying degrees. These category-level differences admit multiple interpretations. For political content, for example, the relatively low association of interaction with a focal hashtag may partly reflect the contentious nature of political discourse online [37]: if political interactions frequently involve disagreement or conflict rather than affiliation, interaction may not translate into adoption in the same way that pre-existing political commitments do. In contrast, fan communities organized around celebrities, gaming, sports, or entertainment may show stronger interaction effects because engagement in these domains is often more affiliative than contentious. A deeper investigation of the mechanisms underlying content-specific patterns is beyond the scope of this study and represents a productive direction for future research.

Taken together, these findings make two contributions to existing literature. First, our results indicate that profile-based self-presentation, which in previous research has often been examined in terms of individual motives, is systematically associated with interaction with others in the network. Second, the study contributes to ongoing debates on collective identity formation in digital environments by illuminating how micro-level inter-

actions and meso- and macro-level identity patterns are associated. The patterns we observe align with social influence theories, which describe how individuals' self-definitions may be shaped in relation to others through mechanisms such as group affiliation, normative expectations, observational learning, or perceived consensus [13–15, 20, 23, 38, 39]. Specifically, our findings indicate that users who interact with others who display particular identity cues are more likely to adopt those cues themselves, and this likelihood increases with the number of observed interaction partners. To the extent that the observed associations in parts reflect social influence processes, the implications of our findings can be considered at two levels. At the individual level, prior research has shown that self-presentation shapes how others engage with individuals and their content. Adaptations in profile presentation may therefore have downstream effects on others' behaviors and interactions. Prior work has further indicated that online self-presentation can feed back into an individual's own self-concept and offline behavior [4, 10–12]. This perspective aligns with Kelman's classic distinction between compliance, identification, and internalization as forms of influence [13]. Even expressions initially adopted for social reasons may over time become internalized. From this perspective, contemporary interaction patterns may be linked not only to subsequent interactions but also, over time, to how users understand and present themselves. At the collective level, systematic associations between interaction and identity cue adoption could contribute to identity clustering within network communities. If users who interact tend to adopt similar identity markers, and if platform algorithms connect users with similar interests, the result may be reinforcement of identity-based groupings. This pattern resonates with discussions of echo chambers, where algorithmic recommendations and homophilous interaction combine to concentrate similar users [40].

These interpretations remain subject to important limitations, which we address in the following subsection.

4.1 Limitations

The most central limitation concerns identification. As formalized by Shalizi and Thomas [32], homophily and contagion are generically confounded in observational network data: under general conditions, any pattern consistent with influence can equally be generated by latent similarity between individuals. Credible causal identification of social influence would require experimental manipulation of exposure, valid instrumental variables, or structural models with strong assumptions [32]. Our quasi-experimental design, despite matching and covariate adjustment, does not satisfy these requirements. Our matching procedure and the inclusion of the hashtag relevance score reduce confounding from observable sources of similarity, but cannot address confounding from unmeasured characteristics such as psychological dispositions, offline contexts, or interests not yet reflected in users' profiles, that may independently predict both interaction and adoption. In other words, the higher adoption rate among treated users could arise solely from latent interests that also increase the likelihood of interacting with hashtag-bearing accounts. Similarly, the observed dose–response pattern could reflect that users with stronger underlying interests both interact with more hashtag-bearing accounts and are more likely to adopt the corresponding identity cue, even in the absence of any causal influence from exposure.

A related consideration is the role of platform recommendation systems. Twitter's algorithm surfaces content based on inferred user interests, which means the interactions we

observe are partly a product of algorithmic curation. This is not a separate confounding mechanism but rather a pathway through which homophily operates. The algorithm infers users' latent preferences and then creates opportunities for interaction with like-minded others. Users inclined toward a particular identity may be shown relevant content, engage with it, and independently update their self-presentation, with the algorithm facilitating the interaction but the underlying interest driving both. Our matching procedure provides partial protection against this concern, as treated and control users with similar observable characteristics should be subject to similar algorithmic curation. However, to the extent that the algorithm acts on signals we cannot observe, it may amplify the latent homophily problem discussed above. Disentangling algorithmic mediation from interpersonal influence would require experimental manipulation of recommendation systems, which is beyond the scope of this study.

Shared environmental exposure presents another related concern. External events such as game releases, political developments, or cultural moments might independently prompt both interaction with relevant accounts and adoption of related identity cues, with no causal connection between the two. The differences we observe across hashtag categories could therefore also be driven by such common shocks that affect some hashtags but not others. The weaker effects observed for political hashtags, for instance, may partly reflect that political engagement often involves disagreement rather than affiliation, but may also reflect that political hashtag adoption is more strongly driven by external political events than by interpersonal influence. Similarly, the strong effects for gaming hashtags could reflect strong community norms around identity display in gaming communities, or coordinated responses to game releases which drive interaction and adoption at the same time. That said, the consistency of positive associations across a multitude of hashtags spanning diverse content categories, including stable traits and ongoing interests not tied to discrete events, suggests the patterns are not fully reducible to coincidental environmental triggers.

An additional limitation, next to the general confounding problem of homophily, is that our data comprise two discrete snapshots separated by 77 days, and we have no visibility into what occurred during this interval. This temporal gap has several implications for interpretation. First, we observe interactions on a single day rather than cumulative exposure history. Users classified as having one interaction partner on day 1 may have had many more interactions with hashtag-bearing accounts during the subsequent weeks; users classified as having many partners may have had no further contact. As a result, the observed dose–response patterns may misrepresent the true relationship between exposure and adoption, potentially attenuating, flattening, or otherwise distorting the shape of the underlying association. The dose–response patterns we report, therefore, reflect a single-day snapshot of exposure patterns rather than the full history of encounters that may have preceded adoption. Second, control users—who by definition did not interact with hashtag-bearing accounts on day 1—may have done so during the interval. If some controls were subsequently exposed through interaction and then adopted, they would appear in our data as unexposed adopters, attenuating the observed difference between treatment and control groups. In this case, our estimates would be conservative, and the true association could be stronger than observed. Third, intervening factors unrelated to the day 1 interaction may have driven adoption. External events, offline experiences, or platform exposure through channels other than direct interaction could have indepen-

dently prompted users to adopt hashtags during the 77-day window. Our matching procedure provides some protection against this concern: matched controls are similar to treated users on observable characteristics and presumably inhabit similar information environments, yet show substantially lower adoption rates. However, treated users, who by definition engaged with hashtag-relevant content, may be more attentive or responsive to external events in those domains than observably similar controls, a difference our matching cannot fully address.

Finally, our analysis derives from a single platform with particular affordances for self-presentation. Whether similar patterns hold for other forms of identity expression or on platforms with different profile structures remains unknown.

4.2 Future directions

While we cannot establish that social influence causes the observed associations between interaction and identity cue adoption, the associations suggest that online self-presentation is embedded in social contexts rather than determined solely by individual preferences. These documented patterns motivate several directions for future research. Experimental designs that manipulate exposure to others' identity presentations could establish whether the associations we observe in fact reflect social influence mechanisms. Longitudinal data with finer temporal resolution could further clarify the dynamics of identity change and help disentangle the interplay between online self-presentation and online social interaction. The variation in the association strength of interaction and adoption across content categories also warrants further investigation. Understanding why interaction appears more predictive for some identity domains (e.g., celebrity fandom) than others (e.g., politics) could help identify the conditions under which social exposure is more or less consequential for identity expression. Qualitative research could complement our quantitative approach by exploring how users experience and interpret encounters with others' self-presentations, and what motivates decisions to adopt particular identity cues. Such work might reveal mechanisms that sole quantitative data cannot distinguish.

4.3 Conclusion

This study documents large-scale associations between online interaction patterns and identity cue adoption in user profiles. Users who interact with others displaying particular hashtags subsequently show substantially higher adoption rates for those hashtags compared to matched non-interacting users. This association strengthens as the observed number of interaction partners rises, and its magnitude varies across content domains. These patterns persist after controlling for observable user characteristics including pre-existing topical alignment of a user's self-presentation with a focal hashtag, which itself is as a strong predictor of adoption as well. The findings suggest that online self-presentation is not independent of the social contexts in which users are embedded. The identity cues users display in their profiles are systematically related to the identities displayed by those they engage with; users' self-presentational choices appear to unfold in relation to their interaction patterns rather than in isolation. By shifting attention from individual motives to the relational and networked contexts in which identity signaling takes place, this study contributes to an emerging understanding of how self-presentation in digital environments is situated within broader social and platform structures. Future research employing experimental or quasi-experimental designs with more longitudinal data could

help clarify the mechanisms underlying these patterns and their implications for phenomena such as community formation, and the dynamics of online collective identities.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-026-00642-5.

Additional file 1. The Supplementary Data 1 file lists the full outcomes and statistics for all 817 hashtag experiments. (XLSX 65 kB)
Additional file 2. The Supplementary Information file contains Supplementary Tables S1-S9, Supplementary Figure S1 and Supplementary Notes S1 and S2. (PDF 217 kB)

Acknowledgements
Not applicable.

Author contributions
L.M.: Conceptualization, methodology, analysis and figures, writing of the original draft. D.M.: Data acquisition, analysis, figures. J.J.J.: Conceptualization, theoretical framework, writing—revision and editing of manuscript. J.P.: Data acquisition, conceptualization, writing—revision and editing of manuscript, supervision. All authors discussed the results and approved the final version of the paper.

Funding information
Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Data availability
We used an existing dataset of publicly available Twitter (now X) user profile metadata that was previously compiled and made available. The dataset comprises two 24-hour snapshots from September and December 2022, as described and referenced in [27]. In accordance with Twitter's terms of service and to protect user privacy, the combined dataset cannot be shared publicly. However, access to the full dataset may be granted by the authors upon request for academic, non-commercial research purposes, subject to institutional approval and compliance with applicable data protection regulations.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]School of Social Science and Technology, Technical University of Munich, Richard-Wagner-Str. 1, Munich, 80333, Germany. [2]Department of Sociology, Stony Brook University, Stony Brook, 11794, NY, USA.

Received: 24 September 2025 Accepted: 11 March 2026 Published online: 20 March 2026

References

1. Sedikides C, Gaertner L, O'Mara Kunz E (2011) Individual self, relational self, collective self: hierarchical ordering of the tripartite self. Psychol Stud 56:98–107. https://doi.org/10.1007/s12646-011-0059-0
2. Vignoles VL, Schwartz SJ, Luyckx K (2011) Introduction: toward an integrative view of identity. In: Handbook of identity theory and research. Springer, New York, pp 1–27
3. Tajfel H, Turner JC (2004) The social identity theory of intergroup behavior. In: Jost JT, Sidanius J (eds) Political psychology: key readings, 0 edn. Psychology Press, New York, pp 276–293. https://doi.org/10.4324/9780203505984-16. https://psycnet.apa.org/doi/10.4324/9780203505984-16
4. Huang J, Kumar S, Hu C (2021) A literature review of online identity reconstruction. Front Psychol 12. https://doi.org/10.3389/fpsyg.2021.696552
5. Rogers N, Jones JJ (2021) Using Twitter bios to measure changes in self-identity: are Americans defining themselves more politically over time? J Soc Comput 2(1):1–13. https://doi.org/10.23919/JSC.2021.0002
6. Barron ATJ, Bollen J (2022) Quantifying collective identity online from self-defining hashtags. Sci Rep 12(1):15044. https://doi.org/10.1038/s41598-022-19181-w
7. Jones JJ (2023) Ipseology - a new science of the self. Jason Jeffrey Jones Productions, Port Jefferson. https://jasonjones.ninja/ipseology-a-new-science-of-the-self-book/
8. Yang L, Sun T, Zhang M, Mei Q (2012) We know what @you #tag: does the dual role affect hashtag adoption? In: Proceedings of the 21st international conference on world wide web. ACM, Lyon, pp 261–270. https://doi.org/10.1145/2187836.2187872. https://dl.acm.org/doi/10.1145/2187836.2187872
9. Taylor S, Muchnik L, Kumar M, Aral S (2022) Identity effects in social media. Nat Hum Behav 7:1–11. https://doi.org/10.1038/s41562-022-01459-8
10. Subrahmanyam K, Šmahel D (2011) Constructing identity online: identity exploration and self-presentation. Springer, New York, pp 59–80. https://doi.org/10.1007/978-1-4419-6278-2_4
11. Manago AM, Graham MB, Greenfield PM, Salimkhan G (2008) Self-presentation and gender on MySpace. J Appl Dev Psychol 29(6):446–458. https://doi.org/10.1016/j.appdev.2008.07.001

12. Lang G (2012) Think Twice before You Post: the Impact of Online Self-Presentation on the Self-Concept. PhD thesis. https://dx.doi.org/10.2139/ssrn.2122809
13. Kelman HC (1958) Compliance, identification, and internalization three processes of attitude change. J Confl Resolut 2(1):51–60. https://doi.org/10.1177/002200275800200106
14. Cialdini R, Goldstein N (2004) Social influence: compliance and conformity. Annu Rev Psychol 55:591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015
15. Turner JC (1991) Social influence. Open University Press, Milton Keynes
16. Levy DA, Nail PR (1993) Contagion: a theoretical and empirical review and reconceptualization. Genet Soc Gen Psychol Monogr 119(2):233–284
17. Barsade SG (2002) The ripple effect: emotional contagion and its influence on group behavior. Adm Sci Q 47(4):644–675. https://doi.org/10.2307/3094912
18. Kramer ADI, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci USA 111(24):8788–8790. https://doi.org/10.1073/pnas.1320040111
19. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. Nature 489(7415):295–298. https://doi.org/10.1038/nature11421
20. Latane B (1981) The psychology of social impact. Am Psychol 36(4):343–356
21. MacCoun RJ (2012) The burden of social proof: shared thresholds and social influence. Psychol Rev 119(2):345–372. https://doi.org/10.1037/a0027121
22. Bandura A, Walters RH (1977) Social learning theory, vol 1. Englewood cliffs Prentice Hall, Englewood Cliffs
23. Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. J Abnorm Soc Psychol 51(3):629–636. https://doi.org/10.1037/h0046408
24. Barash V, Cameron C, Macy M (2012) Critical phenomena in complex contagions. Soc Netw 34(4):451–461. https://doi.org/10.1016/j.socnet.2012.02.003
25. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. Am J Sociol 113(3):702–734. https://doi.org/10.1086/521848
26. Fink C, Schmidt A, Barash V, Cameron C, Macy M (2016) Complex contagions and the diffusion of popular Twitter hashtags in Nigeria. Soc Netw Anal Min 6(1):1. https://doi.org/10.1007/s13278-015-0311-z
27. Pfeffer J, Matter D, Jaidka K, Varol O, Mashhadi A, Lasser J, Assenmacher D, Wu S, Yang D, Brantner C, Romero DM, Otterbacher J, Schwemmer C, Joseph K, Garcia D, Morstatter F (2023) Just another day on Twitter: a complete 24 hours of Twitter data. In: Proceedings of the international AAAI conference on web and social media, vol 17. ICWSM, Limassol, pp 1073–1081. https://doi.org/10.1609/icwsm.v17i1.22215
28. Sayre R (2021) twitter-text (Rust Implementation). https://github.com/sayrer/twitter-text. GitHub repository
29. Twitter, Inc. (2024) twitter-text. https://github.com/twitter/twitter-text. GitHub repository
30. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27(1):415–444. https://doi.org/10.1146/annurev.soc.27.1.415
31. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci USA 106(51):21544–21549. https://doi.org/10.1073/pnas.0908800106
32. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. Sociol Methods Res 40(2):211–239. https://doi.org/10.1177/0049124111404820
33. Levy O, Goldberg Y (2014) Neural word embedding as implicit matrix factorization. In: Proceedings of the 28th international conference on neural information processing systems, vol 2. MIT Press, Cambridge, pp 2177–2185
34. Stuart EA (2010) Matching methods for causal inference: a review and a look forward. Stat Sci, 25(1). https://doi.org/10.1214/09-STS313
35. Hosmer DW, Lemeshow S (2000) Applied logistic regression, 1st edn. Wiley, New York. https://doi.org/10.1002/0471722146. https://onlinelibrary.wiley.com/doi/book/10.1002/0471722146
36. Centola D (2010) The spread of behavior in an online social network experiment. Science 329(5996):1194–1197. https://doi.org/10.1126/science.1185231
37. Berger J, Heath C (2008) Who drives divergence? Identity signaling, outgroup dissimilarity, and the abandonment of cultural tastes. J Pers Soc Psychol 95(3):593–607. https://doi.org/10.1037/0022-3514.95.3.593
38. Moussaïd M, Kämmer JE, Analytis PP, Neth H (2013) Social influence and the collective dynamics of opinion formation. PLoS ONE 8(11):78433. https://doi.org/10.1371/journal.pone.0078433
39. Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 84(2):191–215. https://doi.org/10.1037/0033-295X.84.2.191
40. Diaz Ruiz C, Nilsson T (2023) Disinformation and echo chambers: how disinformation circulates on social media through identity-driven controversies. J Public Policy Mark 42(1):18–35. https://doi.org/10.1177/07439156221103852

Publisher's note