Two-sided Cultural Niches: Topic Overlap, Geospatial Correlation, and Local Group Activities on Event-based Social Networks

Julia Hsin-Ping Hsu
hhsu2@gmu.edu
George Mason University
Fairfax, Virginia, USA

Heejoo Shin
hes002@ucsd.edu
University of California, San Diego
San Diego, California, USA

Noseong Park
noseong@yonsei.ac.kr
Yonsei University
Seoul, South Korea

Myeong Lee*
mlee89@gmu.edu
George Mason University
Fairfax, Virginia, USA

ABSTRACT

As event-based social networks (EBSNs) such as Meetup.com and Facebook Events gain popularity in managing local events (e.g., farmers' markets and social gatherings), two-sided cultural niches are created as event organizers and participants benefit from the platform while affecting each other. Among various factors, niche overlap, an ecological feature, has been studied as a key factor that shapes the success of online communities. While such ecological factors may also shape EBSN-based local groups' success, the context of EBSNs raises unique challenges in understanding the roles of cultural niches due to the informal nature of the local groups and their geographical embeddedness. In this paper, we examine the effects of Meetup groups' topic overlap and geospatial correlation on the activity levels of both organizers and participants, using one-year Meetup data for 500 cities in the United States. We find that (1) a group's topic overlap with other groups on EBSN is associated with its activity levels, and (2) local groups' geospatial correlation may moderate the effects of topic overlap for EBSN users, but inconsistently. The results provide a baseline understanding of EBSN-based groups from an ecological perspective.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; **Social networking sites**; • **Information systems** → **Web applications**.

KEYWORDS

Topic overlap, Meetup, Local groups, Geospatial correlation, Activity level

ACM Reference Format:
Julia Hsin-Ping Hsu, Heejoo Shin, Noseong Park, and Myeong Lee. 2023. Two-sided Cultural Niches: Topic Overlap, Geospatial Correlation, and Local Group Activities on Event-based Social Networks. In *The 11th International Conference on Communities and Technologies (C&T) (C&T '23), May 29–June 02, 2023, Lahti, Finland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3593743.3593758

*Corresponding author.

C&T '23, May 29–June 02, 2023, Lahti, Finland

ACM ISBN 979-8-4007-0758-2/23/05.
https://doi.org/10.1145/3593743.3593758

1 INTRODUCTION

Event-based social networks (EBSNs) such as Meetup.com and Facebook Events are used widely in many countries, providing people with means to organize and attend local events such as farmer's markets, local concerts, and neighborhood gatherings [19]. Both event organizers and participants can benefit from EBSN platforms by lowering the coordination cost while interacting with each other, comprising two-sided cultural niches [25, 26]. As EBSNs gain popularity, many studies have aimed to understand and predict the success of local groups on EBSNs, which involves both the organizer and participant perspectives.

Being *Event*-based social networks, on the one hand, a critical indicator of group success is the popularity of groups' events (i.e., whether people RSVP'd frequently for an event) [33]. Numerous studies have investigated the factors that contribute to the attractiveness of events to users, focusing primarily on the participant perspective [6, 21, 31, 34]. On the other hand, researchers have studied group success from the organizer perspective. Some have analyzed organizers' challenges and motivations for hosting events on Meetup.com through interview analysis [25]. Other quantitative studies have employed novel machine learning models to predict group success based on the number of events and group size (e.g., [17, 18, 20, 23]). These studies often leveraged supervised machine learning models with temporal, spatial, semantic and structural features. In both streams of research, the factors that affect or predict group success were mainly group- and event-level features such as density of event location, member loyalty, sentiment of event description, and text novelty.

1.1 Ecological Perspectives

However, ecological features have been less considered in the EBSN context. Instead, community-level ecological features, such as membership overlap [36] and topic overlap [35], have been extensively studied in non-EBSN contexts, including online communities and formal organizational settings. According to the organizational ecology theory [11], the density of a market niche promotes its legitimation, thereby attracting new members to the niche. Conversely, if the density of the niche intensifies, the members of the niche begin competing for resources, leading to a decrease in the entrance of newcomers. This theory has been developed in relation to social categorization theories because niches exist in the conceptual space where the audience (e.g., participants) and providers (e.g., event organizers) may hold different perspectives on organizations' positions [22]. Accordingly, scholars have examined the impact of organizations' audience-facing positions (e.g., tags or categories) on the success of the organizations [16].

Given this gap in the literature from an ecological viewpoint, the emergence of EBSNs in local communities presents theoretical and methodological challenges in studying the success of local groups. Theoretically, previous work on the organizational ecology of formal organizations and online communities provides limited understandings of (1) the effects of geographical embeddedness of organizations on their dynamics and (2) small-size, informal groups that may not have the same scaling aspirations as formal organizations and online communities. Local groups organized through EBSNs often have activities that are limited to specific geographical boundaries and tend to be small-scale to sustain members' interest-based activities rather than professional work.

Methodologically, prediction studies that aim to enhance the recommendation performances on EBSNs have paid less attention to ecological features in the feature engineering processes, as EBSN-based groups and events have not been studied extensively from an ecological perspective. The diversity of goals and activities among EBSN-based groups, ranging from technology-focused to travel- and business-focused groups that may or may not compete with each other for local participants, might present a challenge for AI researchers to engineer ecological features. These theoretical and methodological gaps necessitate the need to unpack the ecological dynamics of EBSN-based groups by considering both their geographical embeddedness and competition structures.

1.2 Event Topics: Niches in the Label Space

To understand the ecological dynamics of EBSN-based groups, it is necessary to study both the *feature space* and *label space* of the groups. This is because the label space (i.e., tag- or description-based identification of groups on EBSNs) plays a critical role in altering the feature space (i.e., organizers' and participants' actual activities), and vice versa [22]. As an initial effort to understand these dynamics and explore the ecological features in the context of geographically-embedded local groups, this paper focuses on the label space by drawing upon the theoretical frameworks of topic overlap and group success [35, 36].

We specifically focus on topic overlap in this paper, because event descriptions provided by organizers create cultural niches on EBSNs, which are the first encountering points for potential participants in navigating local groups. Given that organizers must position their groups carefully in the conceptual space of cultural ecology to appeal to participants, topic overlap (i.e., the extent to which a group's labels overlap with other groups) becomes an important ecological feature in the label space of EBSNs. To examine these dynamics, we present regression analysis results grounded in the theoretical frameworks of topic overlap. We examine the moderating effect of local groups' geospatial autocorrelation on the relationship between topic overlap and activity level in the EBSN context. Using local group and event data collected from Meetup.com for 500 U.S. cities in 2019, one of the most popular EBSN platforms in the United States, we computationally quantify key variables and conduct statistical analyses to understand the ecological dynamics of Meetup groups in hte label space.

We focus on the 2019 data, because it is the most recent data before the COVID-19 pandemic. Since 2020, people's offline activities have been significantly suppressed due to stay-at-home orders during the pandemic, which may provide weak signals for local group activities and their association with geospatial impact. The results provide implications for organizational theory, EBSN recommender system research, and EBSN system design by demonstrating varying effects of topic overlap on the two sides of the cultural niches (i.e., organizers and participants).

2 RELATED WORK

2.1 Ecological Features and Topic Overlap

In the domain of organizational ecology research, a wide range of factors that affect organizational success have been examined (e.g., [9]). These factors can be grouped into two categories: *intra-* and *inter-community* features. Intra-community features include, but are not limited to, membership size [7], communication activity [29], initial message volume [24], the roles of members [30], perceived activity levels [10], and social network features [27]. Inter-community features include membership overlap and topic overlap [28, 35, 36].

These intra- and inter-community features have been extensively studied in the contexts of either formal organizations or online communities. In the formal organizations' context, for example, competition (i.e., the extent to which organizations compete for resources) and complementarity (i.e., the extent to which organizations benefit from the existence of competitors) were studied as the main ecological forces that affect organizational success [5]. Studies have revealed that organizations initially benefit from their own experience but, over time, this experience may reduce their adaptability to the changing environment. In contrast, learning from competitors could help organizations improve their external capability [4, 13, 14]. This tension between competition and complementarity is closely related to "niche overlap," because the overlap between communities (or niches) plays roles both as the condition for resource spillover and a source of conflict.

2.2 Niche Overlap in Online Communities

These tensions and findings are consistent in the context of online communities. A study suggested that the existence of highly related online communities may increase users' activity levels [12]. Also, prior research has found that a focal community's ecological overlap with other online communities both at the membership (i.e., feature space) and topic levels (i.e., label space) may cause them not only to compete with each other but also to benefit from one another [5, 28, 35]. In other words, niche overlap plays an important role in shaping the success of online communities in *both* feature and label spaces [28, 36].

For example, Zhu and collaborators demonstrated how the success of an online community is shaped by its relationship with other communities using data from an enterprise online community platform [35]. They measured an online community's topic overlap with other communities and estimated the success of a community with its activity level. Their findings revealed a curve-linear relationship between topic overlap and activity level, suggesting that communities tend to complement each other more when their topic overlap is at a moderate level, leading to the highest activity level. Meanwhile, too much topic overlap may intensify the competition between communities, resulting in a lower activity level.

Overall, the literature suggests that an organization in a niche that overlaps with other niches in the label space needs to manage the tension between competition and complementarity well to maximize its success. However, in the context of EBSNs, it remains unclear how the geospatial embeddedness of a group within the ecology of similar groups and the informal nature of local groups would alter the roles of ecological forces between different cultural niches within a geographical area. We aim to unpack these dynamics by focusing on the ecological features of Meetup groups by asking:

> ***RQ***: *How is the topic overlap of Meetup groups associated with their activity levels, depending on the level of their geospatial autocorrelation?*

The outcomes of our analyses will be the foundation for (1) extending the theories of organizational ecology in the context of EBSNs and geographically-defined informal groups, (2) developing computational models to quantify topic overlap of local groups across cultural niches, and (3) enriching ecological features for EBSNs recommender systems.

3 DATA

We collected data for Meetup groups and events across 500 U.S. cities from January to December 2019 using the Meetup APIs.[1] The target 500 cities are defined by the Centers for Disease Control and Prevention (CDC) based on population and the inclusion of the big cities in all the U.S. States.[2] Because the use of EBSN platforms tends to be concentrated in urban areas, we followed the CDC's population-based selections of U.S. cities. The collected dataset contains 3,699,654 events organized by 97,965 Meetup groups. For each Meetup group, available fields include, but are not limited to, group name, description, category, location (latitude and longitude), organized events, and a set of user-generated tags.

The Meetup event data consists of event name, description, event date, hosting group's ID, location (e.g., latitude, longitude, venue name, and address), *how to find us*, and the number of RSVPs. Events that were canceled or received only one RSVP (i.e., events with no RSVPs other than the organizer himself/herself) were removed from the dataset. Figure 1 shows an example of a Meetup event information page, which contains information about the event, meetup location, and a list of members who have RSVP'd.

Groups that organized an excessive number of events within a day are excluded to ensure the quality and reliability of our analysis, because these groups tend to use the Meetup platform for purposes

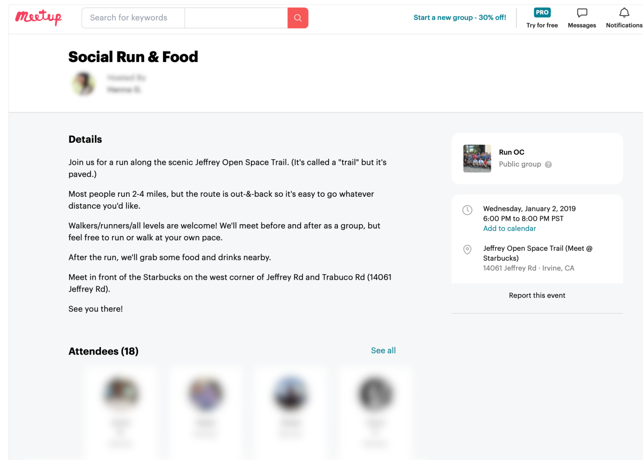

Figure 1: An example of an event information page on Meetup.com

other than sustaining local groups. Through a manual examination, we find that groups organizing more than two events per day on average tended to use the platform for advertisement, rather than sustaining local groups. The removal of such groups is justified as it allows us to focus on groups that are more likely to sustain local groups through organizing Meetup events.

In addition, groups that organized events only in a particular month of the year are also excluded from the regression analysis as well, as groups with a short period of activities provide limited implications for the understanding of the organizational dynamics and overall success of them. The dataset after these removals consists of 2,031,558 events organized by 75,834 Meetup groups across the 500 cities in the United States. Because the venue information of Meetup events is user-generated, there were records with inaccurate or missing geo-coordinates. In these cases, we used Google Geocoding APIs to convert venue addresses into latitude and longitude.[3]

4 APPROACH

4.1 DV: Activity Level

Following prior studies on EBSNs which used the number of events and the number of participants per event (i.e., RSVPs) as metrics of group success [18, 23], we calculate the activity level for each group using its total *number of monthly events* and average number of RSVPs per event per month. The number of events measures the level of activity from an organizer perspective, and the average number of RSVPs operationalizes groups' activity levels from a participant perspective.

4.2 IV: Topic Overlap

Meetup.com requires each group to select one of the pre-defined categories (e.g., Tech, Fitness). These pre-defined categories allow both group organizers and participants to distinguish between different topics and facilitate recommendations. Figure 2 shows the

[1] https://www.meetup.com/meetup_api/

[2] https://www.cdc.gov/places/about/500-cities-2016-2019/index.html

[3] https://developers.google.com/maps/documentation/geocoding

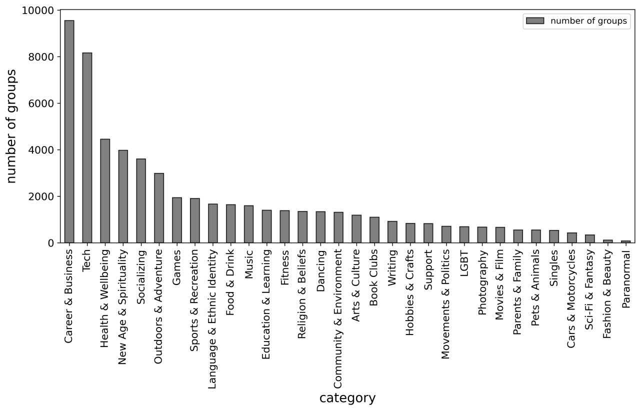

Figure 2: Number of groups for each pre-defined category

number of groups for each category. Although these predefined categories can be used as a proxy to identify cultural niches as each group *has to* choose a particular category out of 32 pre-defined categories, they may not be sufficient in identifying cultural niches, as (1) groups' activities may change over time, (2) some groups may belong to more nuanced categories that cannot be classified using the pre-defined ones, or (3) a couple of categories may actually represent a bigger category in the conceptual space (e.g., hypothetically, fitness and health might be actually one category).

Instead of using the pre-defined categories, we use *event description* as a means to capture cultural niches in the label space. Event organizers of Meetup groups use the platform to position themselves or advertise their events to potential participants by carefully writing event descriptions, which usually include the purposes of events and the details of their activities. Therefore, event descriptions can better capture groups' topics in a more nuanced and accurate way than the pre-defined categories do.

To quantify the topic overlap between different groups, we review prior approaches and propose an enhanced method. Zhu and colleagues quantified topic overlap using content from an enterprise online community platform by modeling them as a vector of TF-IDF (term frequency-inverse document frequency) scores [35]. The method involved calculating TF-IDF for all the postings, and determining topic overlap by summing the cosine similarity scores between the focal community's TF-IDF and the TF-IDF scores of all other communities. Another study employed a different approach by focusing on specificity per-word and derived the average of them in a sentence to represent overlap [32]. While these approaches provide meaningful quantification methods for measuring a focal community's topic overlap with other communities, they are susceptible to potential measurement biases and generalizability issues due to the unpredictable nature of word distributions within specific communities. To mitigate these biases, we employ TensorFlow Hub's Universal Sentence Encoder [8] to convert each group's event descriptions into a vector and use these vectors to calculate topic overlap. By utilizing pre-trained models, we can gain more comprehensive and standardized embeddings of natural languages for groups' topic-wise identity. This approach avoids potential measurement biases that could be caused by user-generated word distributions, as transformer-based pre-trained models capture the contextual information of sentences, and offers a more reliable method for quantifying topic overlap between groups.

After lemmatizing event descriptions, groups are represented by the processed event descriptions which are used as input for the Universal Sentence Encoder. We use the DAN (Deep Averaging Network)-based encoder as it shows better performance on simpler tasks [8].[4] DANs are a type of simple deep learning model where the vector average of one layer is propagated to the next layer after applying matrix multiplication, and classification is performed on the final layer's embedding [15]. Because TensorFlow Hub's Universal Sentence Encoder is already pre-trained based on a large amount of text data, the word embedding of the data presents less bias compared to those that were modeled based on the target data themselves. The universal sentence encoder processes each sample into a 512-dimensional vector. As our dataset contains 75,834 groups, this step results in a two-dimensional tensor of size 75,834×512.

Based on the tensor, we calculate the topic overlap of a group by summing the cosine similarity of sentence embeddings between the focal group and all the other groups. The angular form of cosine similarity is derived as below:

$$C(K,L) = \frac{E_k \cdot E_l}{\|E_k\|\|E_l\|} \tag{1}$$

where $C(K,L)$ denotes the cosine similarity of group K and L, and E_k and E_l refer to the embedding vector of group K and L, respectively, derived from the previous step. Thus, cosine similarity is the normalized dot product of two vectors. Building upon this cosine similarity matrix, the topic overlap of group g is calculated by summing the focal group g's cosine similarity with all the other groups, as shown below,

$$O_g = \sum_{n \in G, n \neq g} C_{gn} \tag{2}$$

where O is the matrix containing topic overlap, G is the set of Meetup groups, and C refers to the cosine similarity matrix calculated in the previous step.

Finally, each group has a scalar value representing the total topic overlap it has with all the other groups. Because topic overlap quantifies a group's topic similarity with the other groups, it measures the extent to which a group's proposed activities and identity overlap with the other groups in the conceptual space of the cultural market (i.e., "label space" [22]) within the EBSN platform. Topic overlap, one of the key attributes that Meetup event organizers can carefully alter through composing their event descriptions, because the prevalence and popularity of similar groups in the adjacent region may affect the visibility and branding of their groups.

4.3 Moderating Factor: Geospatial Effects

Unlike previous research that only focused on online communities or formal organizations, we aim to understand how the relationship between groups' topic overlaps and their success on the EBSN platform are moderated by their geospatial characteristics. The specific methodology used to generate the geospatial variable is explained in the following.

[4] https://tfhub.dev/google/universal-sentence-encoder/4

4.3.1 Identifying online and venue-unknown events. Although most of the Meetup groups tend to organize physical events and usually provide physical event location information such as the venue name, latitude, and longitude, still some groups may also promote their virtual events on Meetup.com by providing additional information such as links to webinars. To investigate the moderating effects of the geospatial factor, online events that were not held at physical locations need to be excluded from our analysis. To achieve this, a Random Forests model is built to classify events into two categories: online and offline. We select 665 events from the dataset and annotate them by checking their names, descriptions and venue information manually. To train the model, a list of keywords (e.g., online, virtual, webinar, zoom) is defined and used to extract indicator features from the event name, venue name, and *how_to_find_us*.[5] We evaluate the model based on the random selection of the training and testing sets 100 times, independently. The average F1 score of 100 independent tests is 0.894 with minimal deviations.

The Random Forests model is trained on the annotated training data to identify online events from the entire dataset. Ultimately, about 1.6% of the entire events are classified as online events. In addition to online events, events that had vague venue information such as *from your home, your computer, my home*, which did not provide precise longitude and latitude, are excluded from the analysis. Keyword matching is employed to remove events that lack accurate geospatial information from the dataset. The final dataset used in examining the moderating effects of geospatial factors comprises 1,277,679 events organized by 58,356 groups. Figure 3 shows the distribution of the number of groups and events from the final dataset across the top 20 cities.

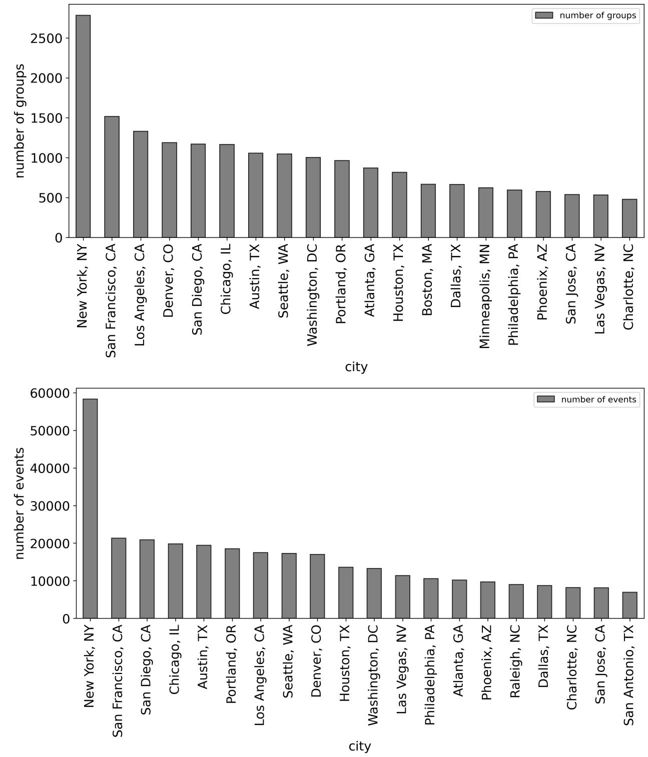

Figure 3: Number of groups and events in 20 cities

4.3.2 Spatial autocorrelation. Spatial autocorrelation is used to relate the value of the variable in a given location and the values of the same variable in the surroundings [2]. A positive autocorrelation value indicates that similar values are located near each other, whereas a negative spatial autocorrelation suggests that similar values tend to be located away from each other. In this study, we represent the location of each group with the centroid of its event locations. Subsequently, we utilize Local Indicators of Spatial Association (LISA) [1] (also known as Local Moran's I) to measure the relationship of organizing patterns between each group and its surrounding groups. The process for computing Local Moran's I for each group is explained in detail below.

(1) **Determining the threshold for the distance band:** According to [37], the average commute distance in the U.S. is between 24 km to 80 km, so we calculate the pairwise distance between groups and select the local minima within this range. The selection of the local minima is based on the idea that a decrease in the number of groups within a certain distance band may indicate the presence of a certain boundary. Figure 4 shows the pairwise distance distribution from 0 to 100 km, in which the local minima appears at 31 km. As a result, the threshold of the distance band is determined to be 31 km (about 19.3 miles).

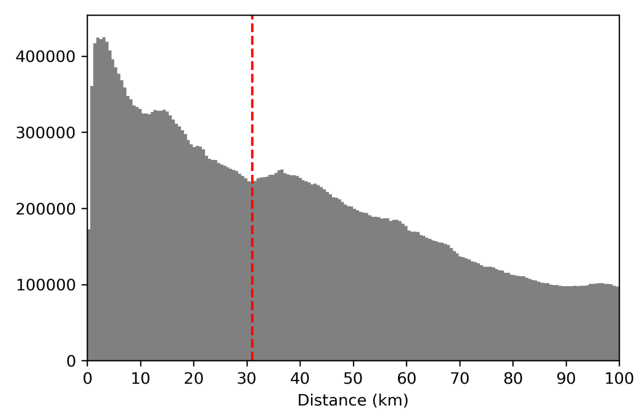

Figure 4: Distribution of distances between pairs of groups

(2) **Defining a spatial weights matrix:** The weight matrix is used to record which groups are close to one another. We define a spatial weights matrix based on the distance band, where the weight between group i and group j is calculated as:

$$w_{ij} = \begin{cases} \frac{1}{dist(i,j)} & \text{if } dist(i,j) \leq 31 \text{ km} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

(3) **Calculating Local Morans' I for each group:** We use Local Moran's I to compute the spatial autocorrelation of the

[5] *how_to_find_us* is a field that provides additional information of the meetup location in a plain text.

number of events, which provides a measurement of the relationship of organizing patterns between each group and its surrounding groups. Specifically, a high value of Local Moran's I indicates that groups have organized a similar number of events in the same region, suggesting the possibility of competitions between groups. The Local Morans's I of a group i is calculated as:

$$I_i = \frac{x_i - \overline{X}}{{S_i}^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_j - \overline{X}) \tag{4}$$

where x_i is the number of events of group i, $\overline{X}$ is the average number of events, $w_{i,j}$ is the spatial weight between group i and group j, n is the total number of groups, and

$${S_i}^2 = \frac{\sum_{j=1,j\neq i}^{n}(x_j - \overline{X}^2)}{n-1} \tag{5}$$

The summary statistics for variables are presented in Table 1.

4.4 Analytic Models

To understand the relationship between Meetup groups' topic overlap and activity levels, we use Ordinary Least-Square Linear Mixed-Effects Models implemented in R [3] where the main effects are nested within a county (a geographical boundary in the U.S. which is bigger than a city but smaller than a state) and Meetup category. We run two sets of models on Meetup groups, where both sets fit the 2nd-degree polynomial relationships between topic overlap and activity levels. The second-degree polynomial terms are based on the analytic models developed in the theories of organizational ecology, where the quadratic and linear terms of topic overlap indicate the competitive and legitimizing forces of the cultural niches [11, 28, 35]. One model predicts the average number of events per month, and the other model is for the average number of RSVPs per event for each group.

Both models include the interaction term between the geospatial autocorrelation of each group (i.e., Local Moran's I) and topic overlap to test the moderating effects of geospatial impact of similar surrounding groups. To control other known effects, we include control variables. They are the logarithm of group membership size, group age, the logarithm of the population of the group's county, ethnoracial heterogeneity of the county (i.e., Gini-Simpson index), and the proportion of households who have access to the internet. County-level control variables were collected or generated from the American Community Survey (ACS) 2019 data collected from U.S. Census Bureau.

We put random effects on county and category (thus, nested effects in the mixed-effects model), because the target Meetup groups are samples of the population (e.g. each category has more Meetup groups outside of the groups in 500 cities), and local groups' behavior and activities are often nested within the larger geo-political or geo-cultural boundaries (e.g., policies, regulations, or urban characteristics). The regression models are as follows:

$$\begin{aligned} a(g_i) = \beta_0 + \beta_1 t(g_i)^2 + \beta_2 t(g_i) + \\ \beta_3 m(g_i) + \beta_4 m(g_i) * t(g_i)^2 + \\ \sum_{g,x} c_{g,x} \end{aligned} \tag{6}$$

where g_i is group i, $a(g_i)$ is the activity level of group i, $m(g_i)$ is Local Moran's I of group i, cg, x is a list of control variables for each group and county x. To test the sensitivity of the models, we test two models (1) using only main effects as the baseline model and (2) with the interaction term to test the moderating effect of Local Moran's I. The results were consistent throughout the models, so we present only the results with all the terms included in the model. Table 2 presents the correlation matrix of fixed effects, indicating a minimal chance of multicollinearity issue in the model.

5 RESULTS

Table 3 presents the overall regression results, which suggest that the activity levels of Meetup groups are associated with topic overlaps, showing statistically significance. Because the regression models are based on second-degree polynomial terms, interpreting the results only based on the estimates is challenging. To understand the regression results better, we plot the fitted lines based on the predicted DVs and their standard errors in Figure 5. Each graph presents two lines depending on whether Local Moran's I is bigger than zero (i.e., *high*) or not (i.e., *low*) to visualize the moderating effects of groups' geospatial autocorrelation.

5.1 Baseline Results: Topic Overlap and Activity Level

The graph on the left of Figure 5 shows that organizers tend to create more events when topic overlap is either low or high; yet they tend to organize fewer events when topic overlap is at a moderate level. The graph on the right side presents that the number of RSVPs keeps increasing when the topic overlap is higher. This is a surprising result, because the results are contradictory to previous work on the impact of topic overlap in the online communities context (e.g., [36]). Theoretically, the effect of niche overlap is the opposite (showing a "bell-shape" instead of "U-shape") in predicting activity levels. In the literature, high topic overlap is a proxy of the intensified inter-niche competition, which usually hampers the entrance of newcomers and lowers the probability of groups' success. However, our findings show that (1) organizers tend to create more events when the topic overlap is either very high or very low, and (2) participants tend to attend Meetup events more when a group's topic overlap is higher (the second graph in Figure 5).

5.2 Moderating Effects of Spatial Autocorrelation

When it comes to geographical effects on the baseline relationship, the results show that there is a small effect of geographical embeddedness of Meetup groups from an organizer perspective (i.e., supply side of the cultural market), as the shapes of the lines in the left graph in Figure 5 are slightly different. This indicates that the geographical autocorrelation of Meetup groups moderates the

Table 1: Summary statistics of key variables.

Variables	Mean	SD	Min	Med	Max
Topic overlap	19798.010	4619.280	-2247.909	20371.024	31397.596
Monthly total # of events	1.825	3.545	0.000	1.000	73.000
Monthly avg. # of rsvps per event	11.014	18.960	2.000	12.000	870.000
Local Moran's I	0.021	0.451	-4.701	0.001	34.940

Table 2: Correlation matrix of fixed effects.

	Overlap	$Overlap^2$	Group size (log)	Group age	Population (log)	Gini-Simpson index	% of Internet Access
$Overlap^2$	0.047						
Group size (logged)	-0.258	-0.055					
Group age	0.049	-0.010	-0.538				
Population (logged)	-0.002	-0.002	-0.032	0.011			
Gini-Simpson index	0.009	0.002	-0.029	0.002	-0.552		
% of Internet Access	-0.005	-0.004	0.012	-0.014	-0.172	0.170	
Local Moran's I	0.025	-0.001	-0.012	-0.001	-0.002	0.002	0.000

Table 3: Results of regression analyses.

	Baseline models			Models w/ interactions		
DV: Avg. # of events	Est.	p		Est.	p	
IVs:						
Intercept	-1.62	0.000	***	-1.62	0.000	***
$Overlap^2$	**28.22**	0.000	***	**27.35**	0.000	***
$Overlap^1$	**30.75**	0.000	***	**30.13**	0.000	***
Local Moran's I	-	-		**0.22**	0.000	***
$Overlap^2 \times LocalMoran's\ I$	-	-		**23.19**	0.000	***
$Overlap^1 \times LocalMoran's\ I$	-	-		**-24.20**	0.004	**
CVs:						
Group size (logged)	0.18	0.000	***	0.18	0.000	***
Group age	0.00	0.004	**	0.0003	0.007	**
Population (logged)	0.02	0.037	*	0.02	0.034	*
Gini-Simpson index	-0.20	0.005	**	-0.20	0.004	**
% of Internet Access	0.60	0.001	***	0.60	0.001	***
DV: Avg. # of RSVP	Est.	p		Est.	p	
IVs:						
Intercept	-0.10	0.414		-0.10	0.415	
$Overlap^2$	**3.47**	0.000	***	**3.32**	0.000	***
$Overlap^1$	**18.90**	0.000	***	**18.81**	0.000	***
Local Moran's I	-	-		-0.01	0.510	
$Overlap^2 \times LocalMoran's\ I$	-	-		3.78	0.291	
$Overlap^1 \times LocalMoran's\ I$	-	-		1.52	0.791	
CVs:						
Group size (logged)	0.26	0.000	***	0.26	0.000	***
Group age	-0.00	0.000	***	-0.00	0.000	***
Population (logged)	0.00	0.626		0.00	0.631	
Gini-Simpson index	0.16	0.000	***	0.16	0.000	***
% of Internet Access	0.42	0.000	***	0.42	0.000	***

$^{***}p < 0.001$ $^{**}p < 0.01$ $^{*}p < 0.05$

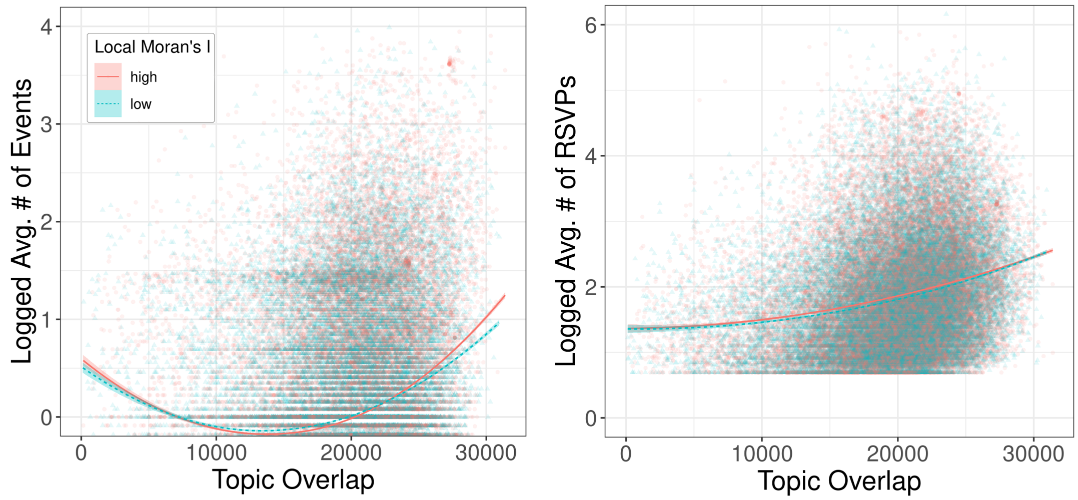

Figure 5: Relationship between topic overlap and groups' organizer-side activity level (left) and participant-side activity level (right).

effect of topic overlap positively (β=23.19, p<0.001), as presented in Table 3. According to Figure 5, the U-shape is narrower when spatial autocorrelation is high (see the red, solid line in the left graph of the figure). In other words, the geographical autocorrelation may intensify the effects of topic overlap on organizers' activities to organize events.

Specifically, while geospatial effects make minimal differences in organizers' activity levels when topic overlap is low or moderate, there is a meaningful difference when a group's topic overlap is high. Among the groups with highest topic overlaps, a focal group's degree of similarity with groups in the adjacent geographical area leads to an increase in the logarithm of the average number of events by 0.3. This corresponds to an average increment of 1.35 events per month for each group involved. While it may seem like a small difference, the impact of geospatial autocorrelation on this difference is substantial. Considering that the median average number of events per month for all Meetup groups is only 1 (Table 1), the increase in events due to geospatial influence is quite significant.

Unlike organizers whose activities may be affected by the geographical embeddedness of Meetup groups, participants' activity level are not affected by it, as the interaction term shows non-significant relationships in Table 3. The graph in the right side of Figure 5 confirms that there are not meaningful differences between participants' RSVPs between the groups with high and low Moran's I. In other words, participants' activities (i.e., RSVPing) are less affected by whether their groups of interest are clustered towards a certain geographical area or not.

6 DISCUSSION

Our findings provide implications for (1) organizational ecology researchers, (2) online community researchers, and (3) EBSN designers, because there are different ecological dynamics on EBSN platforms than online-only communities and formal organizations. Particularly, our discussion revolves around ecological forces of local Meetup groups, the effects of geospatial autocorrelation, and potential platform-specific dynamics.

6.1 Searchability over Competition?

Similar to the ecological studies reviewed in the Related Work section, it was expected that group organizers on EBSNs might have to balance the tension between competition and complementarity to succeed in the cultural niches of local groups. However, our findings on the opposite directions of the regression coefficients compared to prior studies (e.g., [36]) suggest that Meetup organizers' tension management strategies may have to differ from those of other online communities, because there might be hidden factors or unknown platform dynamics shaping the impact of topic overlap on users' activity levels on EBSNs.

There can be different interpretations for the opposite directions of the regression results compared to that of the literature (Figure 5). Organizers might be affected by the perceived competition in the label space of Meetup, but in a different way than in other ecological contexts. One possible interpretation is that a Meetup group's topic overlap with other groups could operationalize visibility and searchability (i.e., the probability of being listed in top rankings from search queries) rather than niche overlap in the label space of the cultural market. It is possible that event descriptions are heavily utilized in the mechanism of recommender systems embedded in the EBSN platform. In that case, overlapping with many other groups in the label space may increase the probability of being exposed to potential participants rather than being unique in the cultural market. Because the algorithms of the Meetup recommender systems are unknown, future studies may have to examine the relationship between topic overlap and searchability to confirm the meaning of topic overlap on EBSN platforms.

This hypothesis regarding the reflection of searchability on topic overlap could be related to the participant-side findings. Participants might be less affected by the competition structure of the cultural niches; instead, their participation may benefit from the increased visibility based on groups' diversification strategies of event descriptions. In other words, participants may join groups based on the increased searchability of the groups driven by the richness of event descriptions.

6.2 Market Saturation and Potentially Hidden Moderators

It is also possible that topic overlap has not been saturated in the cultural market of Meetup.com. In theory, activity levels increase when topic overlap increases to some degree, and then start decreasing as topic overlap further intensifies, presenting a bell-shaped curve in the graph. In the context of Meetup.com, it is possible that the patterns observed among the "high topic overlap" groups are actually those that are observed before the topic overlap reaches its peak. In other words, inter-niche competition might be weak on EBSN platforms, compared to other contexts such as formal organizations and online-only communities, which present only part of the ecological relationship. This might be the case for both organizers and participants, as both of the graphs show either opposite or linear shapes that are usually observed in low topic overlap scenarios in other contexts. Again, to confirm this interpretation, further studies are needed to measure the inter-niche competition on EBSNs.

Another possibility is that there are important moderators not included in the regression model. While we ensured that key variables identified in prior studies were included in the regression models, it is possible that there are unique moderators that are specific to EBSN platforms. For example, category-wide and city-wide factors such as the mobility of the groups, availability of public transportation, or the volume of events per category might have affected the directions of the relationships. While these possibilities are treated as random-effect variables in the regression model, the possibility of new moderating factors needs to be examined directly as fixed-effect variables. This opens up new opportunities for organizational ecology researchers to further explore hidden moderators in shaping EBSN dynamics.

6.3 Differential Geographical Impacts

Although topic overlap indicates idiosyncratic patterns in the ecological dynamics of EBSNs, it is clear that our analysis sheds light on the role of geographical embeddedness of EBSN-based groups in understanding their inter-community competition and activity levels. The difference in the effect size of topic overlap in predicting the average number of events indicates that organizers are not only aware of similar groups on the EBSN platform but also sensitive to those in their neighborhoods. This implies that, even in the context of informal or semi-formal local groups, competition among group organizers exists and tends to be geographically bounded. This finding provides important implications for EBSN designers, as they can implement protocols or strategies to encourage event organizers to enrich their descriptions to attract more participants.

Meanwhile, the findings that there are minimal differences in participants' activity levels depending on spatial autocorrelation indicate that the geographical concentration of similar groups is not salient to or perceived by Meetup participants, hardly creating a sense of geographically-embedded market. In other words, it might be the case that, for participants, a sense of cultural niches might come from only the online space on Meetup.com, but not from local groups' locations. Similar to the organizer perspective, however, participants' perception of cultural niches could benefit further from category-specific analysis, because their perception of cultural niches might vary depending on the topic of local events (e.g., travel groups vs. social gatherings).

Both organizational ecology and online community researchers can be informed by this result by taking geographically-associated factors into account when understanding online dynamics. Additionally, AI and recommender systems researchers can benefit from this study by understanding the roles of ecological factors in group success, considering them in a geographical manner as well. Practically, ecological features such as topic overlap and the derived competition-related features can be used for further engineering machine learning models to increase the generalizability of prediction models moving forward.

6.4 Limitations and Future Research Opportunities

As with other empirical studies, this study has limitations. Because there are other label space features and word embedding models that could be used for quantifying topic overlap, thorough sensitivity tests are necessary to ensure the proposed relationships. Also, choosing the year of 2019 as the study period could present limitations in understanding the evolving dynamics on EBSNs, because the platform features, their users, and external factors such as the COVID-19 pandemic keep changing people's and organizational behaviors. Longitudinal analysis of the reported dynamics is necessary in future studies to precisely understand how external factors such as the pandemic have affected the ecological dynamics. Finally, the study can further benefit from causal inference methods using longitudinal data over multiple years, as the ultimate goal of the theoretical development will be on understanding causality rather than correlation.

Despite the limitations, EBSN and social computing designers can take the ecological features into account in designing the system. Clearly, event description and their association with geographical locations matter in both organizers' and participants' activities. EBSN designers who seek to increase the number of users and their engagement on the platform may have to consider (1) the reappropriation of event descriptions in recommeder systems and user interfaces for both organizers and users, (2) the presentation methods of group identification and event description to potential participants, and (3) their dynamic use based on geographical contexts. Considering these aspects will help social computing and EBSN designers create more engaging and impactful hyperlocal environments.

ACKNOWLEDGMENTS

This work is partly supported by the 4-VA@Mason and the National Science Foundation under Grant No. 2217706. We thank 4-VA@Mason, the NSF DASS program, anonymous reviewers, and the conference chairs.

REFERENCES

[1] Luc Anselin. 1995. Local indicators of spatial association—LISA. *Geographical analysis* 27, 2 (1995), 93–115.

[2] Luc Anselin. 2013. *Spatial econometrics: methods and models*. Vol. 4. Springer Science & Business Media.

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).

[4] Joel AC Baum and Paul Ingram. 1998. Survival-enhancing learning in the Manhattan hotel industry, 1898–1980. *Management science* 44, 7 (1998), 996–1016.

[5] Joel AC Baum and Andrew V Shipilov. 2006. *Ecological approaches to organizations*. Sage handbook for organization studies.

[6] Ayan Kumar Bhowmick, Soumajit Pramanik, Sayan Pathak, and Bivas Mitra. 2021. On the Role of Micro-categories to Characterize Event Popularity in Meetup. In *ICWSM*.

[7] Brian S Butler. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research* 12, 4 (2001), 346–362.

[8] Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder. arXiv:1803.11175 [cs.CL]

[9] Tiago Cunha, David Jurgens, Chenhao Tan, and Daniel Romero. 2019. Are all successful communities alike? Characterizing and predicting the success of online communities. In *WWW*.

[10] Laura Dabbish, Rosta Farzan, Robert Kraut, and Tom Postmes. 2012. Fresh Faces in the Crowd: Turnover, Identity, and Commitment in Online Groups. In *CSCW*.

[11] Michael T Hannan and John Freeman. 1989. *Organizational ecology*. Harvard university press.

[12] Jack Hessel, Chenhao Tan, and Lillian Lee. 2016. Science, askscience, and badscience: On the coexistence of highly related communities. In *ICWSM*.

[13] Paul Ingram and Joel AC Baum. 1997. Chain affiliation and the failure of Manhattan hotels, 1898-1980. *Administrative science quarterly* (1997), 68–102.

[14] Paul Ingram and Joel AC Baum. 1997. OPPORTUNITY AND CONSTRAINT: ORGANIZATIONS'LEARNING FROM THE OPERATING AND COMPETITIVE EXPERIENCE OF INDUSTRIES. *Strategic management journal* 18, S1 (1997), 75–98.

[15] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *ACL-IJCNLP*. 1681–1691.

[16] Balázs Kovács and Michael T Hannan. 2015. Conceptual spaces and the consequences of category spanning. *Sociological science* 2 (2015), 252–286.

[17] Myeong Lee and Brian S Butler. 2020. Cultural activity diversity and community characteristics: An exploratory study. In *Sustainable Digital Communities: 15th International Conference, iConference 2020, Boras, Sweden, March 23–26, 2020, Proceedings 15*. Springer, 32–49.

[18] Guangyu Li, Yong Liu, Bruno Ribeiro, and Hao Ding. 2018. On group popularity prediction in event-based social networks. In *ICWSM*.

[19] Xingjie Liu, Qi He, Yuanyuan Tian, Wang-Chien Lee, John McPherson, and Jiawei Han. 2012. Event-based social networks: linking the online and offline social worlds. In *ACM SIGKDD*. 1032–1040.

[20] Xiang Liu and Torsten Suel. 2016. What makes a group fail: Modeling social group behavior in event-based social networks. In *IEEE BigData*. IEEE, 951–956.

[21] Di Lu, Rosta Farzan, and Claudia Lopez. 2017. To go or not to go! What influences newcomers of hybrid communities to participate offline. In *Proceedings of the 8th International Conference on Communities and Technologies*. 159–168.

[22] Elizabeth G Pontikes and Michael T Hannan. 2014. An ecology of social categories. *Sociological science*. 1 (2014), 311–343.

[23] Soumajit Pramanik, Midhun Gundapuneni, Sayan Pathak, and Bivas Mitra. 2016. Predicting group success in Meetup. In *ICWSM*.

[24] Daphne R Raban, Mihai Moldovan, and Quentin Jones. 2010. An empirical study of critical mass and online community survival. In *CSCW*.

[25] Stephen Ricken, Louise Barkhuus, and Quentin Jones. 2017. Going online to meet offline: Organizational practices of social activities through meetup. In *Proceedings of the 8th International Conference on Communities and Technologies*. 139–148.

[26] Marc Rysman. 2009. The economics of two-sided markets. *Journal of economic perspectives* 23, 3 (2009), 125–43.

[27] Sho Tsugawa and Sumaru Niida. 2019. The impact of social network structure on the growth and survival of online communities. In *ASONAM*. 1112–1119.

[28] Xiaoqing Wang, Brian S Butler, and Yuqing Ren. 2013. The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science* 24, 2 (2013), 414–431.

[29] Wanli Xing and Fei Gao. 2018. Exploring the relationship between online discourse and commitment in Twitter professional learning communities. *Computers & Education* 126 (2018), 388–398.

[30] Wanli Xing, Sean Goggins, and Josh Introne. 2018. Quantifying the effect of informational support on membership retention in online communities through large-scale data analytics. *Computers in Human Behavior* 86 (2018), 227–234.

[31] Tong Xu, Hengshu Zhu, Hao Zhong, Guannan Liu, Hui Xiong, and Enhong Chen. 2018. Exploiting the dynamic mutual influence for predicting social event participation. *IEEE Transactions on Knowledge and Data Engineering* 31, 6 (2018), 1122–1135.

[32] Justine Zhang, William Hamilton, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Community identity and user engagement in a multi-community landscape. In *ICWSM*.

[33] Shuo Zhang and Qin Lv. 2017. Event organization 101: Understanding latent factors of event popularity. In *ICWSM*.

[34] Shuo Zhang and Qin Lv. 2018. Hybrid EGU-based group event participation prediction in event-based social networks. *Knowledge-Based Systems* 143 (2018), 19–29.

[35] Haiyi Zhu, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E Kraut. 2014. Selecting an effective niche: an ecological view of the success of online communities. In *CHI*.

[36] Haiyi Zhu, Robert E Kraut, and Aniket Kittur. 2014. The impact of membership overlap on the survival of online communities. In *CHI*.

[37] Pengyu Zhu. 2013. Telecommuting, household commute and location choice. *Urban Studies* 50, 12 (2013), 2441–2459.