Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM 2021)

On the Role of Micro-categories to Characterize Event Popularity in Meetup

Ayan Kumar Bhowmick[1], Soumajit Pramanik[2], Sayan Pathak[3] and Bivas Mitra[1]
[1]Department of Computer Science & Engineering, IIT Kharagpur, India
[2]Department of Electrical Engineering and Computer Science, IIT Bhilai, India
[3]Microsoft Research, Redmond, WA, USA
ayankb@iitkgp.ac.in, soumajit@iitbhilai.ac.in, sayanpa@microsoft.com, bivas@cse.iitkgp.ac.in

Abstract

Event-based social networking platforms such as *Meetup* have recently witnessed a huge growth. However, with the rise in the volume of groups and events, making individual events attractive has become increasingly challenging for its organizers. As a result, we find that events hosted by groups of same category at similar venues and similar times, also widely differ in their popularity. Data study reveals that the topics specified in textual descriptions of events may be key to their popularity. In this paper, we introduce a novel concept of topical *micro-categories* in the context of EBSNs for accurately characterizing events, such that events belonging to the same *micro-category* exhibit similar popularity profile. We develop a principled method to detect such *micro-categories* from the textual descriptions of individual events. Our experiments reveal the significance of the detected *micro-categories* in determining the popularity of associated *Meetup* events and groups. We also investigate the effectiveness of the *micro-categories* in a real-world application scenario by developing a recommendation model; this model recommends relevant *micro-categories* to a group for hosting its future events with enhanced popularity. Notably, our model achieves an average *NDCG* score of around 0.75 showing a straight 5% improvement over the best performing competing method.

1 Introduction

In recent years, there has been a rapid growth in popularity of event-based social networks (EBSNs) such as *Meetup* (Liu et al. 2012), *Douban* and *Plancast* (Wang et al. 2015). *Meetup* users may join various online groups of their interest and participate in physical events, hosted by those groups at various venues. *Meetup* groups' survivability heavily depends on the success of the events hosted by the respective *Meetup* groups, which is frequently measured as the volume of participants attending those events. Attendance of the hosted *Meetup* events depends on multiple factors, such as (a) topic of the event (reflected by its textual description), (b) group hosting the event, (c) venue (physical location) of the event and (d) time of the event. Among these factors, event topics are primarily determined by the *Category* of the *Meetup* groups hosting those events. *Category* field indicates the broad interest of a *Meetup* group (say, 'Career/Business', 'Tech', 'Health/Wellbeing', 'Socializing' etc.) and topic of the events hosted by those groups. Hence, the role of the *Meetup* group category in regulating the popularity of the hosted events calls for an in-depth investigation.

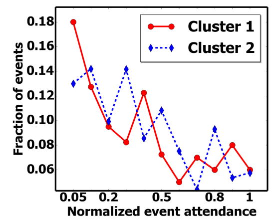
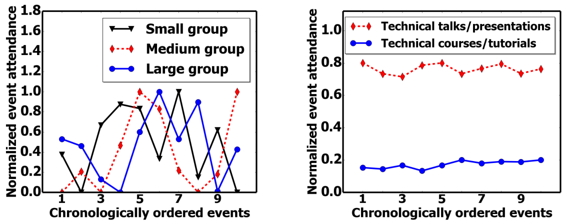


Figure 1: Attendance distribution of two sample clusters of events from the 'Tech' category; each cluster denotes a set of events hosted at similar locations and similar times. Attendance is normalized considering the attendance distribution of all events in 'Tech' category.

Our study reveals that popularity of *Meetup* events often widely varies, even if they are hosted by the (i) same type (category) of groups, (ii) located at very similar venues and (iii) have similar hosting time. For instance, in Fig. 1, we cluster all the *Meetup* events (say, Cluster 1) hosted by the 'Tech' category *Meetup* groups in Chicago, arranged during the morning session of February 2016 at the 'Hyde Park' area of Chicago. Albeit all these events share the same category, hosting venue and timing, we observe a wide variation in their attendance distribution. Similar behavior can be observed for all the 'Tech' *Meetup* events, hosted during the evening session of April 2017 at the 'Lincoln Park' area of Chicago (say, Cluster 2). On the other hand, as shown in Fig. 2(a), even events hosted by the same group can also widely vary irrespective of the size of the group. Interestingly, Fig. 2(b) shows that all the events related to 'Technical talks/presentations' attracts higher attendance than the events related to 'Technical courses/tutorials', despite the fact that all of them are hosted by the *Meetup* groups from 'Tech' category. This anecdote clearly highlights the limitation of the vanilla *Meetup Category* to provide a proper characterization of the hosted *Meetup* events. Essentially, each *Meetup* event exhibits certain intrinsic characteristics (as fo-



(a) (b)

Figure 2: Impact of topical inclination on event attendance. (a) Normalized attendance of events hosted by few sample groups from the 'Tech' category with small (below 33.33^{th} percentile), medium (above 33.33^{th} & below 66.67^{th} percentiles) and large (above 66.67^{th} percentile) size respectively, (b) Normalized attendance of few sample events from the 'Tech' category related to the topics 'Technical talks/presentations' and 'Technical courses/tutorials'.

cused sub-topics), which reflects the objectives and activities of a specific event. We hypothesize that fine-grained characterization of *Meetup* events may play an important role to perceive their capacity to attract event participants.

Attempts have been made in state-of-the-art literature for understanding the factors that regulate event popularity in an EBSN (Xu et al. 2018; Zhang and Lv 2018); we have provided a brief survey in Sec. 2. Although existing works have extracted and exploited event specific semantic features such as sentiment, novelty and parts of speech tags from its description, they have proved to be ineffective in predicting the corresponding event's popularity (Zhang and Lv 2019). We believe that the lack of proper topical characterization of the events is the key factor behind it. Most recent endeavours have overlooked the necessity of efficient event characterization, beyond the vanilla *Meetup Category*, which may limit their understanding of the event popularity dynamics.

In this paper, we introduce the notion of *micro-categories* for efficient characterization of *Meetup* events. *Micro-categories* are essentially collections of related keywords, broadly describing sub-topics of a *Meetup* category. For instance, the *Meetup* category 'Tech' depicts the *Meetup* groups that are, in general, interested in hosting events such as technical conferences/talks and tutorials, whereas the *micro-category* {e-learning course, professional development course} may characterize the 'Tech' events specifically related to online technical courses and tutorials. Essentially, each *micro-category*, within a broad *Meetup* category, provides a fine grained characterization of the hosted *Meetup* events. All events characterized by a specific *micro-category* may exhibit a certain popularity profile, which may provide an in-depth understanding regarding the topical choice of the attending participants. This may further facilitate the event hosts to judiciously choose the suitable *micro-categories*, as the crisp topic of upcoming *Meetup* events.

Exploring the possibilities, *Meetup* event descriptions appear as an obvious source of information for mining *micro-categories*. One may apply various *Information Retrieval* techniques such as text summarization (Hovy and Lin 1998; Maybury 1999) and keyword extraction (Zhang 2008; Rose et al. 2010) on the event descriptions to characterize the individual *Meetup* events. However, vanilla text summarization & keyword extraction techniques concentrate only on characterizing the individual events and overlook clustering of events into sub-topics available under the *Meetup* category. Hence, attempts need to be taken for suitably constructing the *micro-categories* which can characterize a collection of closely related events, from the outputs of the aforementioned techniques. One may leverage on applying the state-of-the-art clustering algorithms on the extracted keywords, such as hierarchical clustering (Day and Edelsbrunner 1984), k-means clustering (Likas, Vlassis, and Verbeek 2003) etc. to identify the related events, which may be an important step towards constructing *micro-categories*.

In this paper, we propose a methodology to discover *micro-categories*, which is equipped to characterize the popularity profile of *Meetup* events. First, we highlight various challenges of characterizing the popularity of *Meetup* events and demonstrate the role of event topics as a proxy of such characterization (Sec. 4). Since keywords are the building blocks of *micro-categories*, first, we extract the keywords from the event descriptions and build a corpus of keywords for all the events hosted by the *Meetup* groups in same category. Next, we employ BERT to learn the keyword representations, such that two keywords that are semantically similar will be placed close to each other in the embedding space. Finally, we apply hierarchical agglomerative clustering on the generated keyword embeddings to obtain *micro-categories* as detected clusters (Sec. 5). Close inspection of the discovered *micro-categories* reveals that each *micro-category* may characterize multiple *Meetup* events with similar popularity profile and has a significant causal influence on the popularity of its associated events. This facilitates us to develop a simple *micro-category* based model to classify the *popular* & *unpopular Meetup* groups (Sec. 6). Finally, in Sec. 7, we demonstrate the utility of the detected *micro-categories* by developing a recommendation model that recommends a ranked list of *micro-categories* to the *Meetup* groups for hosting upcoming events. We show that these recommended *micro-categories* are indeed beneficial for the *Meetup* group organizers and may play a significant role to improve the popularity of their upcoming events (Sec. 8).

2 Related Work

We present a brief survey on EBSN event popularity prediction, followed by the EBSN based recommendations. Next, we explore various Information Retrieval techniques for text summarization and keyword extraction, which may facilitate the discovery of *micro-categories*.

Event Popularity in EBSN

Majority of attempts made in state-of-the-art literature have focused on understanding extrinsic factors that regulate the popularity of events in an EBSN (Zhang, Zhao, and Cao 2015; Zhang and Lv 2018, 2017). For instance, (Zhang and Lv 2017) have studied the role of spatial, temporal, semantic

City	Groups	Members	Events	Total Headcount
Chicago	7718	427613	458087	5465266
New York	23270	1192431	1008317	10198310
San Francisco	17647	848032	713967	5340869

Table 1: Dataset statistics

and group specific features in estimating the overall event popularity. On the other hand, (Xu et al. 2018) leverages the dynamic mutual influence & social influence among the users for improving event participation prediction. Further, Zhang et. al. (Zhang and Lv 2018) have proposed a personalized random walk based approach on a hybrid EBSN graph for predicting event participation of users. However, due to lack of proper topical characterization of the events, most of these works proved to be ineffective in predicting the corresponding event's popularity.

Event Recommendation in EBSN

Event recommendation to the *Meetup* members may regulate the popularity of events. State-of-the-art literature have focused on recommending relevant events to EBSN users by utilizing multiple signals affecting event popularity (Li et al. 2017a). In addition, few methods have employed Bayesian probabilistic models (Qiao et al. 2014) for cold-start event recommendation while event embedding based recommendation models proposed recently (Wang and Tang 2019) encode the observed relationships among different entities in EBSN. However, majority of recommendation attempts serve the interest of event participants, while ignoring the requirements of the event hosts to decide on the topics for the upcoming events.

Text Summarization Techniques

Classical *Information Retrieval* literature provides a wide gamut of tools to extract topics from EBSN event descriptions. For instance, text summarization (Maybury 1999) and keyword extraction (Rose et al. 2010) techniques may be used to characterize the topic of a textual description. Rich literature exists to extract the most salient words (or phrases) from a given text, such as using statistical approaches (Cohen 1995), linguistic approaches (Hulth 2003) and machine learning based approaches (Minh et al. 2005). However, vanilla text summarization & keyword extraction techniques may concentrate only on individual events and overlook the broad cluster of sub-topics available under each *Meetup* category.

3 Meetup Terminologies

In this section, we introduce notations and basic terminologies for representing different entities in *Meetup* and subsequently define the popularity of *Meetup* events & groups.

Notations and Meetup Dataset

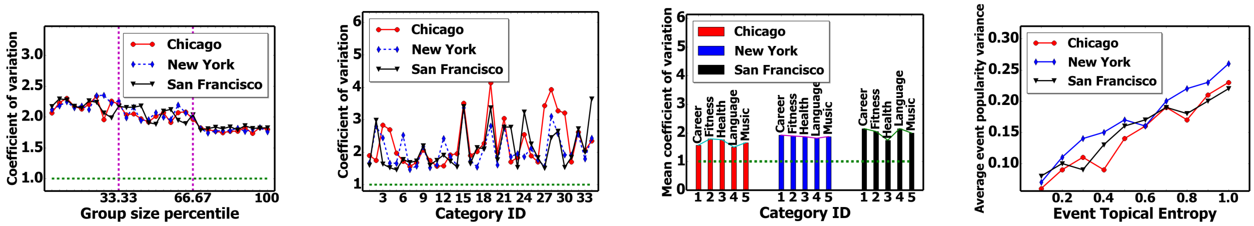
Let $\mathcal{U}$, $\mathcal{E}$ & $\mathcal{G}$ denote the set of all members, events and groups respectively in *Meetup*. Each *Meetup* group $g \in \mathcal{G}$ consists of a set of members that belong to $\mathcal{U}$. *Meetup* members physically attend *Meetup* events organized by group g at different locations, based on their preferences. At the time of group formation, *Meetup* group organizers need to specify the *Category* field, indicating the broad interest of the *Meetup* group. For instance, groups from 'Tech' category mostly indulge in hosting technical events, whereas groups from 'Fine Arts/Culture' category prefer hosting classes and workshops related to a specific art. During formation, each group is assigned to one of the 33 *official* categories (Pramanik et al. 2016) defined in *Meetup* (say, 'Career/Business', 'Tech', 'Health/Wellbeing', 'Socializing' etc.). Additionally, whenever a *Meetup* group is formed, organizers are asked to select a set of interest tags best describing the group. We utilize the *Meetup* API to crawl detailed information about *Meetup* groups and hosted events across three cities, namely, Chicago, New York and San Francisco for the period from August 2015 to December 2018. For every *Meetup* event, we collect its textual description describing the event as well as the corresponding RSVP responses from event participants. We collect the event *Headcount*, which provides an accurate approximation of the total number of participants who attended the event. We present a brief summary of the crawled dataset[1] in Table 1.

Meetup Event & Group Popularity

Popularity of an event e is defined as the actual number of *Meetup* members attending the event, obtained from the *Headcount* (H_C). In order to normalize H_C for an event to compute its popularity, we divide all the *Meetup* groups in same category into 10 bins based on their group sizes, and for each bin of groups, we normalize H_C of the events hosted by a group within that bin, considering the distribution of H_C of all events in that bin. For example, for an event e hosted by a *Meetup* group g in category c belonging to a particular bin, the popularity of e is computed as $P_e = \frac{A^e - A^c_{min}}{A^c_{max} - A^c_{min}}$ where A^c_{max} and A^c_{min} are the maximum and minimum number of members attending an event hosted by all the groups in that bin, while A^e is the number of members attending event e. This normalization policy fairly treats the popularity of events hosted by small as well as large sized *Meetup* groups.

In our dataset, popularity P_e of all events hosted by groups of category c follow the normal distribution (with mean μ and standard deviation σ). Following the principle of outlier detection, we designate an event e as *popular* if the corresponding P_e is above $\mu + 2\sigma$ or *unpopular* if P_e is below $\mu - 2\sigma$ (Grubbs 1969) while remaining events are left unlabeled. Notably, the popularity labeling has to be performed separately for individual group categories as the range of event attendance may widely vary across different group categories. A *Meetup* group is labeled as *popular* or *unpopular* based on the popularity of its hosted events. If a majority (more than 66.67%) of events hosted by a group g is labeled as *popular* (*unpopular*), the corresponding group

[1] Our anonymized dataset is available at https://bit.ly/32lFkd8



(a) (b) (c) (d)

Figure 3: Distribution of events with varying levels of popularity - (a) Event popularity over different group sizes, (b) Event popularity over different group categories, (c) Event popularity over different event clusters with similar location & time, (d) Correlation between event topical entropy and variance of event popularity

g is labeled as *popular* (*unpopular*). All other groups remain unlabeled.

4 Characterization of Event Popularity: Challenges and Opportunity

In this section, first we highlight the challenges in characterizing popularity of *Meetup* events. Next, we uncover the potential of event topics as a solution and introduce the notion of *micro-categories* to formally conceptualize event topics.

Challenges

The popularity of a *Meetup* event can be influenced by multiple factors closely associated with the hosted event, such as (a) the hosting group & its category (b) location of the hosting venue, (c) hosting time. Our pilot study demonstrates the limitation of the aforementioned factors in characterizing the popularity of *Meetup* events.

Limitation of hosting group In order to show the impact of hosting groups on the event popularity, we estimate how widely the popularity of events vary for each hosting group. We calculate the *Coefficient of Variation* (C_v) $= \frac{\sigma}{\mu}$ (Bendel et al. 1989) for each such group where μ and σ are respectively the mean and standard deviation of the popularity of the events hosted by the respective group. We obtain the mean C_v across all the groups as 1.98, which is higher than 1 and thus indicates huge variability in event popularity (Bhattacharjee and Sengupta 1996). In other words, just the identity of the hosting *Meetup* group fails to infer the popularity of its hosted events.

One step further, we examine if the popularity of a *Meetup* event specifically correlates with the size of the respective hosting group. For this purpose, we divide the *Meetup* groups in three classes, based on the distribution of group sizes for all groups in a specific city - (a) small (size less than 33^{rd} percentile of the group size distribution), (b) large (size greater than 67^{th} percentile of the group size distribution) and (c) medium (groups with sizes between 33^{rd} & 67^{th} percentile of the group size distribution). This ensures that there exists a uniform volume of *Meetup* groups in the small, medium and large size class in a city. As shown in Fig. 3(a), the C_v indices for each set of events hosted by groups of a particular size range are also found to be indicating widespread variation. Notably, we observe that the C_v indices are pretty high (between $2-3$) for *small groups* compared to *large groups*, which signifies that the variability in event popularity relatively reduces once a group is able to attract significant population as members. Nevertheless, even for the *large groups*, we cannot determine the event popularity from the size of the group.

Limitation of group category Next, we consolidate the events hosted by all groups from a specific *Meetup* category in a particular city and plot their corresponding C_v indices in Fig. 3(b). We observe high C_v, ranging between $1.5-4$, indicating that events hosted by the groups of a specific *Meetup* category in a city exhibit disperse behaviour in terms of the event popularity.

Limitation of event venue and time We construct clusters of events from each *Meetup* category in a city, that are similar in terms of (i) hosted venue (specified by latitude, longitude) and (ii) hosting time (specified by hour of day, day of week and month of year). In Fig. 3(c), we plot the mean C_v index over all such event clusters, obtained for Top-5 *Meetup* categories in each city. The high range of C_v (between $1.4-2.5$) across Top-5 categories in all the three cities points to the fact that events with overlapping group category, hosting venue and time also exhibit widespread variation in event attendance. A similar observation holds for the remaining *Meetup* categories as well across the cities. In a nutshell, our fine-grained analysis reveals that the factors commonly used for describing a *Meetup* event, such as *Meetup* group, category, event location and hosting time are intrinsically insufficient to accurately characterize the event popularity, observing its wide variation across these factors.

Opportunity: Role of Event Topics

Finally, we turn our attention to the topic of the hosted *Meetup* events. We consider *keywords*, a meaningful phrase or a small sequence of words, extracted from the description of the events (using keyword extraction tool *RAKE* (Bhat et al. 2018)) as a proxy of the event topic. For each group $g \in \mathcal{G}$, we obtain a distribution of all the keywords used for describing its events ($\mathcal{K}_g$) and calculate the entropy of this distribution as $E_g = \sum_{w \in \mathcal{K}_g} -p_w \log p_w$, where p_w is the fraction of times keyword w appears in the description of any event hosted by g. Side by side, we also calculate the variance of event popularity (σ_g^2) hosted by group g. Plotting these two measures (E_g & σ_g^2) for all the *Meetup* groups in each city in Fig. 3(d), we clearly observe that the popularity variance for events in a group increases with increase in their topical entropy (with correlation coefficient 0.52 for all groups in the three cities). This in turn indicates that events that are hosted on close & related topics (low entropy) are highly likely to exhibit similar event popularity. Hence, topically characterizing the *Meetup* events may pave the way for proper characterization of their popularity profile.

Notion of Micro-categories

In this paper, our key contribution is to introduce the concept of micro-categories for proper characterization of the *Meetup* events. We define *micro-category* as a cluster of semantically similar keywords representing the topical inclinations of *Meetup* events. Events hosted by a specific *Meetup* group are mostly similar; nevertheless, they exhibit minor but subtle unique characteristics, which group category fails to capture as they all broadly come under the single category of the hosted group. *Micro-categories* should enable us to discriminate the *Meetup* events with different topical inclinations hosted by a single *Meetup* group. The discovery of *micro-categories* becomes particularly useful in classifying the *popular* and *unpopular* events hosted by the same *Meetup* group, as well as recommending suitable *micro-categories* to (especially unpopular) groups for hosting *popular* events in future. In the following, we propose a principled methodology to detect the *micro-categories*, present under a single *Meetup* category.

5 Methodology to Detect Micro-categories

Our proposed method[2] for the detection of *micro-categories* from a *Meetup* category consists of the following four broad steps - (i) Extracting relevant keywords from descriptions of events hosted by the *Meetup* groups of same category, (ii) Filtering relevant keywords, (iii) Generating keyword embeddings and (iv) Applying hierarchical agglomerative clustering on the generated keyword embeddings to obtain *micro-categories* as detected clusters. The procedure is repeated across all *Meetup* categories to detect corresponding *micro-categories*.

Keyword Extraction

Here we extract the set of keywords (denoted by K_e^c) for each *Meetup* event e, hosted by a group in category c. We first perform tokenization, punctuation & stopword removal on the raw unstructured event descriptions (using *NLTK* (Bird and Loper 2004)) followed by applying *RAKE* (Bhat et al. 2018), a domain independent keyword extraction algorithm, to obtain K_e^c. K_e^c represents the inherent semantics of the textual description of e and hence, can be used to characterize the topic of event e. For each keyword $w \in K_e^c$, *RAKE* also returns a salience score s_w^e which indicates the relevance of w in the description of event e.

[2]Our codes are publicly available at: https://bit.ly/35uMCx7

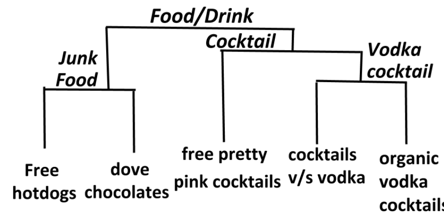


Figure 4: A toy example of a dendrogram for a set of 5 keywords in 'Food/Drink' category

Keyword Filtering

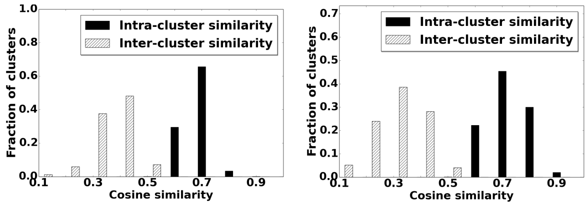
We observe that many of the keywords returned by *RAKE* have a low salience score. Since retaining such keywords would lead to poor quality of detected *micro-categories*, we propose two filtering methods to eliminate such irrelevant keywords in this step. The choice of a specific filtering method depends on the available dataset.

Mean salience score based filtering We first obtain the global set of keywords $\mathcal{K}_c$ of category c as the union of keyword sets K_e^c for all the events hosted by the *Meetup* groups of category c. We then compute a mean salience score for each keyword $w \in \mathcal{K}_c$, inspired from the notion of *TF-IDF score* (Wu et al. 2008). Here the term frequency of a keyword w is weighted by its salience score s_w^e (returned by *RAKE*), whereas the inverse document frequency for w is computed as $\frac{1}{|\mathcal{E}_w|}$ where $\mathcal{E}_w$ is the set of events whose textual description contains the keyword w. Now we compute the mean salience score for w as:

$$s_w^c = \frac{1}{|\mathcal{E}_w|} \times \sum_{e \in \mathcal{E}_w} s_w^e \tag{1}$$

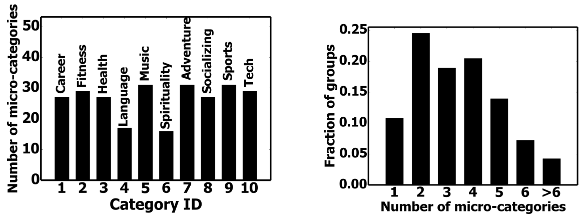
We may choose different thresholds from various central tendency statistics (say, mean, median, mode etc.) computed on the distribution of mean salience scores of the keywords in $\mathcal{K}_c$ and filter out all such keywords having mean salience scores below the chosen threshold. In our implementation, we have chosen median as the threshold.

Stochastic filtering This approach is motivated from the *False Discovery Rate (FDR)* (Strimmer 2008) in null hypothesis testing. We aim to stochastically eliminate the highly frequent generic keywords present in most of the events. Here we first create a global multiset of keywords $\mathcal{M}_c$, considering all events hosted by the *Meetup* groups in category c. Notably, a given keyword $w \in \mathcal{K}_c$ can have multiple occurrences in $\mathcal{M}_c$ if w appears as a keyword in multiple events. Now for each event e, if $n_e^c = |K_e^c|$ denotes the number of keywords describing e, we randomly sample n_e^c number of keywords from the multiset $\mathcal{M}_c$. We repeat this random selection 1000 times for event e and compute the fraction of times a keyword in $\mathcal{M}_c$ is randomly selected. We eliminate all those keywords from K_e^c for which this fraction is greater than 0.5.



(a) (b)

Figure 5: Distribution of inter-cluster and intra-cluster similarity of *micro-category* clusters for - (a) Mean salience score based filtering, (b) Stochastic filtering



(a) (b)

Figure 6: Statistics of the extracted *micro-categories* - (a) Number of extracted *micro-categories* in Top-10 *Meetup* categories, (b) Distribution of number of *micro-categories* across the *Meetup* groups

From both the approaches, we finally obtain a filtered set of salient keywords K_e^{c*} for each event e and obtain $\mathcal{K}_c^*$ as a union of all these keyword sets K_e^{c*} for category c.

Learning Embedding Vectors of Relevant Keywords

Given the obtained corpus of salient keywords $\mathcal{K}_c^*$, we aim to learn the representation of each keyword $w \in \mathcal{K}_c^*$. We employ *BERT* (Devlin et al. 2018) which learns text representations by using masked language models that enables to obtain pre-trained deep bidirectional representations. Since each keyword w denotes a sequence of English words, we implement a variation of the *BERT* model as a sentence encoding service[3] that maps each variable length keyword w to a fixed length embedding vector $\mathbf{y}_w$ of dimension d. The embeddings obtained in this manner ensure that two keywords that are semantically similar will be placed close to each other in embedding space.

Detection of Micro-categories

In this step, our objective is to cluster the semantically similar keywords in $\mathcal{K}_c^*$ based on their learned embeddings. For this clustering, we rely on constructing a dendrogram using the hierarchical agglomerative clustering algorithm (Forina, Armanino, and Raggio 2002).

Since agglomerative clustering is a bottom-up approach, each keyword $w \in \mathcal{K}_c^*$ is initially placed in its own cluster. Subsequently, pairs of similar keywords (clusters) are merged as we move up the hierarchy, based on *cosine similarity* between their respective embeddings, using the average linkage policy to compute similarity between cluster pairs. The topmost level of the resulting dendrogram consists of a single cluster containing all the keywords in $\mathcal{K}_c^*$ while the bottom level consists of $|\mathcal{K}_c^*|$ clusters, each containing a single keyword (see the toy dendrogram in Fig. 4). At level i, we obtain n_i clusters $\mathcal{C}_c^i = \bigcup_{j=1}^{n_i} \mathcal{C}_{cj}^i$, where each cluster $\mathcal{C}_{cj}^i \subseteq \mathcal{K}_c^*(1 \leq j \leq n_i)$. In order to determine the optimal set of keyword clusters for the *Meetup* category c, we compute the *Silhouette coefficient* (Gat-Viks, Sharan, and Shamir 2003) for each keyword at a level i and obtain the mean *Silhouette coefficient* over all keywords considering the set of clusters at this level. We then choose the level i^* with the highest mean *Silhouette coefficient* as the optimal level in dendrogram to determine the optimal set of clusters. Each cluster $\mathcal{C}_{cj}^{i^*}(1 \leq j \leq n_{i^*})$ then represents an obtained *micro-category*, henceforth designated as $\mathcal{M}_j^c$ corresponding to the *Meetup* category c.

Importance of Keyword Filtering

We start with an anecdotal example to highlight the importance of keyword filtering (see Sec. 5). We apply the *micro-category* detection algorithm on the events hosted by the *Meetup* groups of 'Food/Drink' category, using the Stochastic keyword filtering approach. From Fig. 4, we can visibly observe that Stochastic filtering yields meaningful clusters for a set of five keywords in the bottom level of this dendrogram, denoting the identification of high quality and relevant micro-categories 'Junk Food' and 'Cocktail'.

This observation becomes more pronounced when we measure the quality of obtained clusters in terms of *intra-cluster* and *inter-cluster* similarity indices. We measure these similarity indices by calculating the average cosine similarity between every pair of keywords *within clusters* and *across clusters* respectively, such that higher values denote greater similarity. In Figs. 5(a) and 5(b), we plot the distributions of inter-cluster and intra-cluster similarity for the *micro-category* clusters obtained over all the *Meetup* categories for both Mean salience score based and Stochastic keyword filtering approaches. We observe that the distribution of intra-cluster similarity and inter-cluster similarity are highly distinctive in case of Stochastic filtering compared to the Mean salience score based filtering.

In a nutshell, Stochastic filtering approach yields superior and more stable *micro-category* clusters compared to the Mean salience score based approach in our dataset. Moreover, the Mean salience score based filtering is highly sensitive to the choice of various central tendency based thresholds. Henceforth, we only employ the *micro-category* clusters obtained using Stochastic filtering for the interest of space and readability.

6 Properties of Micro-categories

First glimpse of the discovered *micro-categories* reveals that the number of obtained *micro-categories* widely varies across the *Meetup* categories. Precisely, in Fig. 6, we observe that (minimum) 17 *micro-categories* are present

[3]https://github.com/hanxiao/bert-as-service

in ‘Spirituality’ category, while (maximum) 34 *micro-categories* are present in Adventure’ category. Next, we dissect the *micro-categories* to uncover their various properties connected to *Meetup* events, groups and their popularity.

Every Meetup Event is Associated with One Micro-category

We aim to identify the *micro-category* $\mathcal{M}_e^c$ associated to each *Meetup* event e hosted by groups of category c. Considering the set of salient keywords K_e^{c*} for event e (outcome of Sec. 5), we compute the fraction of such keywords present in each *micro-category* $\mathcal{M}_j^c (1 \leq j \leq n^c)$ as identified in Sec. 5 where n^c is the number of *micro-categories* identified under category c. We designate the *micro-category* having highest overlap with the keyword set K_e^{c*} as the *micro-category* $\mathcal{M}_e^c$ associated to event e. Since the *micro-categories* $\mathcal{M}_j^c\ \forall j$ are mutually disjoint, each event e gets uniquely associated to a single *micro-category*.

In Fig. 7(a), we plot the distribution of fraction of overlapping keywords for an event e in its associated *micro-category* $\mathcal{M}_e^c$. We observe that a majority of events (82%, 85% and 80% for Chicago, New York and San Francisco respectively) exhibit a very high overlap of its keywords (greater than 70%) with its associated *micro-category*. Hence, we conclude that the associated *micro-category* $\mathcal{M}_e^c$ correctly characterizes the respective *Meetup* event e. Next, in Fig. 6(b), we show the distribution of *micro-categories* associated to events hosted by a single *Meetup* group. We observe that most of the *Meetup* groups (78%) prefer to host events in only a few sub-topics (reflected by $2-5$ *micro-categories*), whereas events hosted by 10% of groups are indeed diverse (in > 6 *micro-categories*).

Meetup Events Characterized by Same Micro-category Exhibit Homogeneity

Each *micro-category* $\mathcal{M}_j^c$ may get associated with multiple *Meetup* events hosted by the groups of category c. Here we analyze the homogeneity (in terms of popularity) of all the *Meetup* events, characterized by a single *micro-category*. Fig. 7(b) demonstrates the *coefficient of variation (V)* values for events associated with each *micro-category* belonging to a single *Meetup* category ‘Movies’ across all cities. The V values corresponding to individual *micro-categories* are found to be much lower than 1.0 whereas the overall V for the same category ($= 1.91$) comfortably surpasses the 1.0 threshold. This indicates that the homogeneity of popularity of events associated to each *micro-category* $\mathcal{M}_j^{'Movies'}$ is much higher compared to the homogeneity of all the events belonging to category ‘Movies’. The overall trend of V values (mean and variance) for *micro-categories*, as shown in Fig. 7(c), clearly validates the same observation across all the *Meetup* categories.

We now consider comparing the homogeneity of events associated with each *micro-category* with the homogeneity of events associated with groups of a particular size considering all cities (see Sec. 4). We find the *micro-categories* characterizing all events hosted by groups of a particular size range and compute the mean and variance of the V values over all such *micro-categories*. Fig. 7(d) depicts that the mean V values corresponding to events associated with *micro-categories* are much less (all below 1.0) compared to all the events hosted by groups of specific sizes (all above 1.0). In other words, *micro-categories* can characterize the popularity of *Meetup* events in a much finer-grained manner compared to raw group sizes.

Factor	All city	Chicago	New York	San Francisco
Micro-category	**0.807**	**0.789**	**0.803**	**0.811**
Category	0.670^*	0.689^*	0.669^*	0.690^*
Group size	0.633^*	0.686^*	0.537^*	0.637^*
Event venue	0.212^*	0.220^*	0.207^*	0.200^*
Event time	0.585^*	0.516^*	0.570^*	0.637^*

Table 2: ATE values for different causality factors. "*" denotes statistical significance of ATE value of micro-category (treatment) over the respective confounding factor with p-value ≤ 0.01 for a one-sided t-test.

Causality Analysis of Event Popularity

In this section, we investigate the causal relationship between *micro-category* $\mathcal{M}_e^c$ associated with an event e (hosted by a group of category c) and its corresponding popularity value P_e. For this purpose, we apply standard causal inference techniques (Mehrotra et al. 2017) where we define the event *micro-category* $\mathcal{M}_e^c$ as the treatment (variable) and event popularity P_e as the response (variable) to the treatment. In order to claim that $\mathcal{M}_e^c$ influences the popularity P_e of an event e, the following two conditions have to be satisfied : (1) Treatment variable $\mathcal{M}_e^c$ should always precede the response variable P_e, and (2) There should not be any other variable X_e (denoted as confounding variable) that could be a potential factor influencing the event popularity P_e. The *micro-category* $\mathcal{M}_e^c$ of an event can be obtained from the textual description of the event e which is available from the point of time when the event e is announced on *Meetup*. On the other hand, the popularity P_e computed from event attendance can only be made available after e has physically occurred. Hence, we can safely say that $\mathcal{M}_e^c$ always precedes P_e i.e. the condition (1) is satisfied.

We assume that there is a causal relationship between the treatment (event *micro-category*) and the response (event popularity) if a small change in the treatment variable causes a significant change in the response variable whereas no significant change in the response variable is observed on changing a confounding variable X_e. This expected change in the response variable with respect to change in the treatment or a confounding variable is measured as the Average Treatment Effect (ATE value) (Rubin 1974). In our experiments, we use the four factors (identified in Sec. 4), namely, size of the hosting group, category of the hosting group, location of hosting venue and time of hosting, as candidate confounding variables. To compute ATE, we identify matching event pairs, e^1 and e^2, which are hosted by groups of same category and similar sizes as well as have similar venue locations and hosting times. However, respective *micro-categories* associated with e^1 and e^2 are highly diverse. We now compute the average treatment effect as:

$$ATE = E(\frac{P_{e^1} - P_{e^2}}{f(X_{e^1}, X_{e^2})})_{(e^1,e^2)\in Y} \quad (2)$$

where $E(.)$ is the expectation, e^1 and e^2 is a matching pair in Y and $X_.$ is a causal factor (treatment variable or a confounding factor). Here function $f(.)$ either denotes cosine similarity between corresponding vectors[4] of e^1 and e^2 when $X_.$ is treatment variable or confounding factor other than size of hosting group while $f(.)$ denotes difference when size is the factor. A high value of *ATE* denotes that the corresponding factor $X_.$ influences the popularity value $P_.$. Table 2 shows the corresponding *ATE* values for the treatment variable i.e. event micro-category as well as the four confounding factors, for all events as well as city specific events. We can conclude from the high *ATE* values for event micro-category compared to all the confounding factors that event micro-category is the sole factor having a significant causal influence on event popularity, overshadowing the confounding factors.

Popularity Classification of Meetup Groups Using Micro-categories

We demonstrate the effectiveness of *micro-categories* for classifying *Meetup* groups based on popularity using standard machine learning classifiers. Given a *Meetup* group g belonging to category c, we form a *micro-category* feature vector $\mathbf{f}_g$ of length n where the j^{th} element ($1 \leq j \leq n$) of $\mathbf{f}_g$ is computed as the fraction of events hosted by g that is associated with *micro-category* $\mathcal{M}_j^c$. We obtain similar feature vectors for all *Meetup* groups of category c.

Evaluation setup We leverage on the *micro-category* vectors $\mathbf{f}_g$ as simple features to develop a supervised learning framework to classify a *Meetup* group g as *popular* or *unpopular* (see Sec. 3 for ground truth labels). We implement a *Logistic Regression* model to train a binary classifier using $10-$fold cross validation over all groups of category c and evaluate the classification performance on the test set in each iteration. We compute the average *Accuracy*, *Precision*, *Recall* and *F1-scores* to measure the classification performance. We repeat this procedure for all the *Meetup* group categories and report the mean indices.

Baseline algorithms We implement the following baselines for classifying *Meetup* groups.

(1) Tag vector based: For a *Meetup* group g, we form a binary tag vector $\mathbf{t}_g$ of length m, where m is the number of possible group tags considering all groups. We assign j^{th} element ($1 \leq j \leq m$) of $\mathbf{t}_g$ to 1, if the corresponding tag j is present in the description of group g; otherwise, it equals 0. The resulting tag vector $\mathbf{t}_g$ serves as the feature for popularity prediction of groups.

(2) DeepRSVP: We adapt the method proposed in (Li et al. 2017b) that implements a deep neural network for predicting RSVP counts (see Sec. 3) of future events to estimate the event popularity. Subsequently, we predict the popularity of each *Meetup* group as *popular* or *unpopular* from the estimated RSVP counts.

(3) *LDA* based: We implement a variation of our classification model based on *micro-category* feature vector, where we replace *RAKE* by *LDA* (Latent Dirichlet Allocation) (Blei, Ng, and Jordan 2002) to obtain the set of keywords (K_e^c) from the description of event e (see Sec. 5). Notably, each keyword $w \in K_e^c$ obtained using *LDA* is essentially a single topical word. Here the salience score s_w^e denotes the probability of event e to belong to the topic w.

Results Table. 3 shows that our classification model based on *micro-category* feature outperforms all the baseline methods, achieving an *Accuracy* and *F1-score* of 0.838 and 0.866 respectively across all three cities. Essentially, *popular Meetup* groups primarily host *popular* events, which get associated to distinct *micro-categories*, different from the *micro-categories* associated to *unpopular* events (see Sec. 6); this observation depicts the role of *micro-categories* as popularity signature. Evidently, the *LDA* based baseline shows a much inferior performance. Close inspection reveals that *micro-categories* obtained from *LDA* based method represents a cluster of generic topical words, which can be associated with both popular and unpopular events, leading to their failure in properly classifying the group popularity. Among the other baselines, *DeepRSVP* performs the best with comparable performance to our model; however, it underperforms due to noise involved in estimating the total RSVP counts of future *Meetup* events. The tag vector based baseline exhibits poor performance, since it concentrates only on broad interest of groups and overlooks event popularity.

In a nutshell, we conclude that *micro-categories* exhibit popularity signatures of *Meetup* events and groups, as evident from data study, causality analysis as well as popularity classification model.

7 Recommending Micro-categories for Upcoming Events

In this section, we demonstrate the utility of *micro-categories* by proposing a model to recommend a ranked list of *micro-categories* $\mathcal{R}_g$ to an *unpopular* group $g \in \mathcal{G}$ for hosting its upcoming events. The expectation is that the upcoming events, hosted on the recommended *micro-categories*, would be *popular*.

Recommendation Model

We propose a two-step model to obtain a ranked list of recommended *micro-categories* $\mathcal{R}_g$ for an *unpopular* group g:

Step 1: Identifying a candidate set of *micro-categories* We obtain the candidate set of *micro-categories* $\mathcal{M}_g$ to be recommended for target *unpopular* group g, by first identifying a set of *Meetup* groups $\mathcal{G}_g$, which belong to the (i) same category as group g and (ii) are labeled as *popular*. The candidate set of *micro-categories* $\mathcal{M}_g$ would be the collection of all *micro-categories*, that are associated to all the *popular events* $\mathcal{E}^{\mathcal{G}_g}$ hosted by the groups in $\mathcal{G}_g$.

[4]We learn a vector representation corresponding to the set of keywords within the micro-category associated with the event using *BERT*

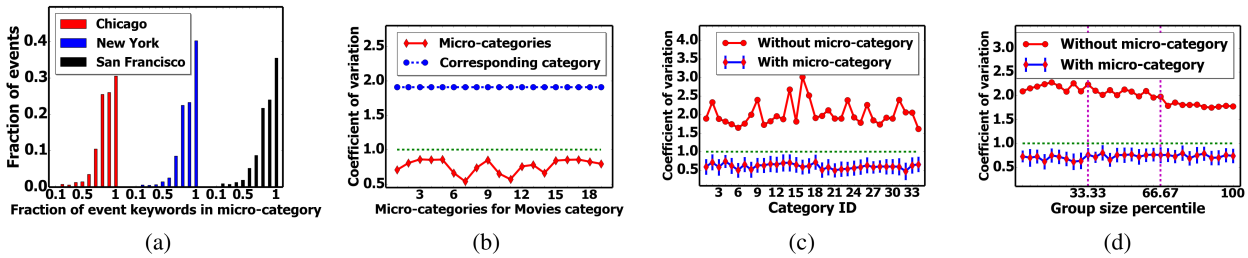


Figure 7: Properties of *micro-categories* - (a) Fraction of event keywords in associated *micro-category*, (b) *Meetup* events described by same *micro-category* within 'Movies' category exhibit homogeneity in popularity, (c) Events belonging to same *Meetup* category are associated with *micro-categories* having low variation in event popularity, (d) Events belonging to groups of similar sizes are associated with *micro-categories* having low variation in event popularity.

Step 2: Ranking the *micro-categories* in $\mathcal{M}_g$ First we compute a relevance score ∇_j^g for each *micro-category* $\mathcal{M}_j \in \mathcal{M}_g$. Next, we rank the *micro-categories* in $\mathcal{M}_g$ based on the relevance scores to recommend *micro-categories* ($\mathcal{R}_g$) to a target group g for hosting the upcoming events. Intuition behind the relevant score is the following. For each event $e \in \mathcal{E}^{\mathcal{G}_g}$, we compute its contribution to the relevance score of $\mathcal{M}_j$ as the average salience score s_j^e of the keywords extracted from the event e, which are overlapping with *micro-category* $\mathcal{M}_j$. We suitably scale up this contribution if the event e is hosted by a group g_e with (a) high tag overlap T_{g,g_e} with the target group g and (b) high event attendee overlap U_{g,g_e} with the attendees of the events hosted by the target group g. Accordingly, we estimate the relevance score of *micro-category* $\mathcal{M}_j \in \mathcal{M}_g$ as:

$$\nabla_j^g = \frac{1}{|\mathcal{E}^{\mathcal{G}_g}|} \cdot \sum_{e \in \mathcal{E}^{\mathcal{G}_g}} s_j^e \cdot (w_1 T_{g,g_e} + w_2 U_{g,g_e}) \qquad (3)$$

We train the model to learn the model parameter vector $\theta_w = [w_1, w_2]$ as follows.

Model Training

Consider the set of events $\mathcal{E}_g$, sorted chronologically, which are hosted by the target group g. We split $\mathcal{E}_g$, denoting the first 70% of events as *past events* and remaining 30% as *future events* of group g. We repeat this procedure for all groups $g \in \mathcal{G}$. Next, we divide the set of all *unpopular* groups into 70% training set ($\mathcal{G}^{train}$) and 30% test set ($\mathcal{G}^{test}$). For training the recommendation model, we obtain a ground truth relevance score ∇_j^{g*} for each *micro-category* $\mathcal{M}_j \in \mathcal{M}_g$ from the *future events* hosted by group $g \in \mathcal{G}^{train}$. If $\mathcal{M}_j \in \mathcal{M}_g$ is a *micro-category* associated to any future *popular* event of g, we set the ground truth relevance score $\nabla_j^{g*} = 1$; otherwise, we set $\nabla_j^{g*} = 0$. Side by side, we estimate the relevance score ∇_j^g for each *micro-category* $\mathcal{M}_j$ from the *past events* hosted by groups in $\mathcal{G}_g$ for group $g \in \mathcal{G}^{train}$. We learn the model parameters θ_w by optimizing the loss in correctly estimating relevance score for groups in $\mathcal{G}^{train}$.

Loss optimization The objective function to learn the optimal parameter set θ_w for a *micro-category* $\mathcal{M}_j$ to be recommended to the group $g \in \mathcal{G}^{train}$ becomes

$$\min \frac{1}{2|\mathcal{M}_g|} \sum_{\mathcal{M}_j \in \mathcal{M}_g} (\nabla_j^g - \nabla_j^{g*})^2 + \frac{\lambda}{2|\mathcal{M}_g|} \cdot {\theta_w}^2 \qquad (4)$$

where λ is the regularization parameter. Since we learn the optimal parameter set θ_w over all groups g in the training set $\mathcal{G}^{train}$, the combined objective function can be written as:

$$\min \sum_{g \in \mathcal{G}^{train}} \frac{1}{2|\mathcal{M}_g|} \Big(\sum_{\mathcal{M}_j \in \mathcal{M}_g} (\nabla_j^g - \nabla_j^{g*})^2 + \lambda \cdot {\theta_w}^2 \Big) \quad (5)$$

Once we learn the optimal model parameters θ_w in the training phase, we estimate the relevance score ∇_j^g of every candidate *micro-category* $\mathcal{M}_j \in \mathcal{M}_g$ from the *past events* of groups in $\mathcal{G}_g$ for the target group $g \in \mathcal{G}^{test}$. Subsequently, we rank the *micro-categories* in $\mathcal{M}_g$ accordingly and recommend the ranked list $\mathcal{R}_g$ to the target *unpopular* group $g \in \mathcal{G}^{test}$.

8 Evaluation of Recommendation Model

In this section, we demonstrate the performance of the proposed recommendation model compared to the baseline methods. Additionally, we perform a set of experiments to reveal key insights pertaining to the proposed model.

Experimental Setup

First, we describe the procedure to conduct the evaluation of the recommendation model. Subsequently, we introduce the evaluation metrics as well as the baseline algorithms.

Evaluation procedure In order to evaluate the quality of the recommended *micro-categories* $\mathcal{R}_g$ for hosting *future events* of a target *unpopular* group $g \in \mathcal{G}^{test}$, we use the corresponding binary ground truth relevance labels ∇^{g*} as discussed in Sec. 7. We define the ground truth set of relevant *micro-categories* for each group $g \in \mathcal{G}^{test}$ as $\mathcal{M}_g^{rel} = \{\mathcal{M}_{j'}$ such that $\nabla_{j'}^{g*} = 1\}$ where $\mathcal{M}_{j'}$ can be any *micro-category* from the same category as group g. High overlap between $\mathcal{R}_g$ and $\mathcal{M}_g^{rel}$ indicates that if an *unpopular* group g hosts future events associated to a recommended *micro-category* in $\mathcal{R}_g$, such events are highly likely to become *popular*.

Method	Accuracy	Precision	Recall	F1-score
Micro-category	**0.838**	**0.846**	**0.887**	**0.866**
LDA	0.738	0.781	0.744	0.763
Tag vector	0.752	0.752	0.765	0.759
DeepRSVP	0.813	0.824	0.859	0.841

Table 3: Overall classification performance for predicting group popularity in all cities

Evaluation metrics We introduce the following standard metrics to compare the recommended *micro-categories* $\mathcal{R}_g$ with the ground truth set of relevant *micro-categories* $\mathcal{M}_g^{rel}$.

(a) Normalized Discounted Cumulative Gain (NDCG): NDCG (Yilmaz, Kanoulas, and Aslam 2008) is the normalized version of the Discounted Cumulative Gain (DCG) scores computed as the sum of the relevance scores of the Top-k recommended *micro-categories* after discounting them for their corresponding positions in the ranked list (dividing by log of the corresponding ranking position). Here we use the average popularity of future events associated with *micro-categories* as ground truth relevance scores.

(b) F1@k: It is simply computed as the harmonic mean of Precision@k (fraction of *micro-categories* in the Top-k recommended list that belongs to $\mathcal{M}_g^{rel}$) and Recall@k (fraction of *micro-categories* from $\mathcal{M}_g^{rel}$ that are in the Top-k recommended list).

(c) Mean Average Precision (MAP): It is defined as the average of Precision@i ($1 \leq i \leq k$) values for the Top-k recommendations.

Baseline algorithms In the absence of any direct state-of-the-art *micro-category* recommendation model, we adapt the existing algorithms for implementing baselines.

(1) Event recommendation centric baseline: We adapt the method proposed in (Müngen and Kaya 2017), which recommends the Top-k events to a *Meetup* user to attend. In order to recommend *micro-categories* to a target *unpopular* group g, first we collect the *micro-categories* associated to the events, that appear in Top-k recommendations of users who participated in *past events* hosted by group g, as the candidate set of *micro-categories*. Subsequently, we rank this candidate set giving higher preference to *micro-categories* associated with events frequently ranked higher in the Top-k recommendations to users and finally obtain a ranked list of *micro-categories* for g.

(2) Model variant baseline: This is a variation of the proposed recommendation model, where a target *unpopular* group g is recommended *micro-categories* associated to the past *popular* events, hosted by the *popular* groups of same category as group g. Higher preference is given to *micro-categories* that have higher overlap in keywords with its associated *popular* events hosted in the past.

(3) *LDA* based baseline: We implement a variation of the proposed recommendation model, where the candidate set of *micro-categories* are obtained by replacing *RAKE* by *LDA* (Blei, Ng, and Jordan 2002) (in Sec. 5) to extract the set of keywords (K_e^c) from the description of event e.

Model	NDCG@k	F1@k	MAP
Our method	**0.754**	**0.696**	**0.762**
LDA based baseline	0.747	0.669	0.755
Model variant baseline	0.726	0.647	0.738
Event recommendation centric baseline	0.718	0.641	0.721

Table 4: Overall recommendation performance in terms of relevance (each cell value indicates the mean metric value for Top-k recommendations where $k = 5$)

Overall Performance

In Table. 4, we summarize the performance of our recommendation model with respect to baseline algorithms in light of the evaluation metrics. We observe that our proposed recommendation model outperforms all the baselines with high values of *NDCG* and *MAP* (0.754 and 0.762 respectively). Among the baselines, the *LDA* based method is observed to attain a comparable performance. Here the *micro-categories* obtained from the *LDA* based method are mostly composed of generic topical keywords, which are likely to exhibit a decent overlap with descriptions of many popular events from the same category. As a result, many Top-k recommended *micro-categories* get associated with future popular events, hence becoming relevant. Nevertheless, since our *RAKE* based methodology yields *micro-categories* very specific to events with a similar popularity profile, it gives superior performance albeit by a smaller margin. On the other hand, the model variant baseline exhibits slightly inferior performance implying that *micro-categories* of *popular* events hosted by *popular* groups of same *Meetup* category can still provide reasonable recommendation for the *future events* to be hosted. However, it fails to provide higher preference to the relevant *micro-categories*, as the past event popularity does not necessarily repeat in the future. Moreover, this model overlooks the importance of the similar groups (in terms of tag and event attendees) hosting *popular* events, which plays a role in gaining popularity for the *future events* hosted by an *unpopular* group. The event recommendation centric baseline shows even worse performance since it fails to distinguish between different *micro-categories* associated with events hosted by the same group, while computing their relevance.

Drilling Down the Recommendation Model

(1) Role of *micro-category* detection method on the utility of the recommendation model We study the impact of extracting the *micro-categories* from different dendrogram levels (see Sec. 5) on the utility of the recommended *micro-categories*. We characterize the utility of a recommended *micro-category* M_j for an *unpopular* group g (of category c) by *F1-score*[5], which measures the keyword overlap between the hosted *popular* future event e (K_e^c) and the recommended *micro-category* M_j. In Fig. 8(a) we compare the

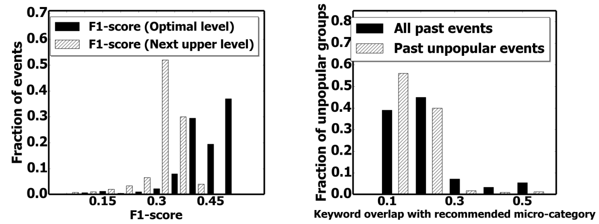
[5]We compute *Precision* as the fraction of keywords in a Top-k *micro-category* $\mathcal{M}_j$ that also belong to K_e^c while *Recall* is computed as the fraction of keywords in K_e^c that belong to $\mathcal{M}_j$. Subsequently, *F1-score* is obtained as the harmonic mean of *Precision* and *Recall*.

distribution of the *F1-score* values corresponding to the set of *micro-categories* obtained at the (i) optimal dendrogram level i^* and (ii) the next upper level $i^* + 1$. We observe that *F1-scores* are much higher (mean 0.41, with mean precision 0.32) for *micro-categories* at optimal level i^*, compared to *micro-categories* at level $i^* + 1$ (mean 0.29, with mean precision 0.22). This observation confirms that the *micro-categories* obtained at the optimal level of the dendrogram are more relevant than their parent clusters for the proposed recommendation model. In the hindsight, we note that the *Recall* suffers at the optimal level (mean 0.43) as *Recall* favors larger cluster sizes obtained at the next upper level (mean 0.49).

(2) Recommended *micro-categories* are distinct from the topics of past hosted events Top-k *micro-categories* that are recommended to an *unpopular* group g are denoted as 'distinct' if past hosted events of that group have low keyword overlap with these recommended *micro-categories*. In Fig. 8(b), we observe that for 83% of *unpopular* groups, the mean keyword overlap with the Top-5-recommended *micro-categories* over all the past hosted events is less than 20%. This confirms that recommendations made by our model consist of new and distinct *micro-categories* and hence, if followed, are likely to benefit the group in enhancing the popularity of its upcoming events. Fig. 8(b) also shows that the mean keyword overlap with the Top-5 recommended *micro-categories* is low (less than 20%) for the past *unpopular* events, hosted by the *unpopular* groups.

(3) Recommended *micro-categories* are aligned to the interest of the group We investigate the relevance of the recommended *micro-categories* with respect to the interest of the target *Meetup* groups. We apply *BERT* to learn a joint representation of each target unpopular *Meetup* group's interest, combining its (i) official category, (ii) group interest tags, and (iii) interest tags of members. Similarly, we apply *BERT* to obtain a combined representation of the keywords associated with the Top-3 *micro-categories* recommended for that target *Meetup* group. We compute the average cosine similarity between these two learned representations, and compare that with the null model, where we replace the recommended *micro-categories* by random *micro-categories* from the same target group category. We observe that the average cosine similarity between the target *Meetup* group representation and recommended *micro-category* representation is pretty high (0.41) compared to that of the null model (0.25), considering all groups across all three cities. Hence, we conclude that the recommended *micro-categories* are indeed aligned with the interests of the respective target unpopular groups.

(4) Effectiveness of recommended *micro-categories* Finally, we conduct two surveys to evaluate the relevance of our recommended *micro-categories* to the *Meetup* stakeholders. We recruit 30 volunteers from our university, who are mostly familiar with the *Meetup* platform, to conduct these surveys. We handpick five unpopular *Meetup* groups from various categories and show the volunteers the event hosting & event attending profiles of those groups. In the first survey, we concentrate on the *Meetup* event hosts and examine if our recommended *micro-categories* are relevant for them, whereas in the second survey, we focus on the *Meetup* event participants and examine if the *Meetup* members will be interested to attend the events hosted following our recommended *micro-categories*. The first survey outcome shows that more than 80% of volunteers feel that those *Meetup* groups would possibly be interested to host events following our recommended *micro-categories* whereas for the second survey, more than 72% of volunteers feel that those *Meetup* members will be interested to attend events hosted following our recommended *micro-categories*.



(a) (b)

Figure 8: Key insights of recommendation model - (a) Comparison of Top-5 recommended *micro-categories* at optimal level and next higher level of dendrogram, (b) Top-5 recommended *micro-categories* are highly distinct

9 Conclusion

In this paper, we have introduced the novel concept of *micro-categories* capturing the specific topical inclinations of individual events in *Meetup*. We have developed a principled method for detecting such *micro-categories* within a *Meetup* category as a collection of semantically related keywords extracted from textual description of events hosted by same category groups. We have employed hierarchical clustering on keyword representations learned from *BERT* to obtain *micro-categories* as clusters of similar keywords. Notably, every event is observed to be associated with only a single *micro-category* indicating its core topical affinity. Further, our data study has revealed that each *micro-category* characterizes a set of related events that show homogeneity in terms of popularity, thereby playing an important role in distinguishing *popular* and *unpopular* events that broad group categories as well as individual groups fail to do. We have validated the significance of these fine-grained *micro-categories* by demonstrating their causal impact on event popularity as well as successfully classifying the popularity of *Meetup* groups (with high *F1-score* of 87%) using *micro-categories* as features. Moreover, we have demonstrated the effectiveness of the *micro-category* information in a real-world application scenario by developing a recommendation model; this model recommends a ranked list of relevant *micro-categories* to a group for hosting its upcoming events with enhanced popularity. Our extensive experimental evaluation have not only shown the superior quality of our proposed recommendation model compared to competing baselines but also demonstrated its capability of recommending newer and distinct *micro-categories* (not used in past events) to individual *Meetup* groups.

References

Bendel, R.; Higgins, S.; Teberg, J.; and Pyke, D. 1989. Comparison of skewness coefficient, coefficient of variation, and Gini coefficient as inequality measures within populations. *Oecologia* 78(3): 394–400.

Bhat, A.; Satish, C.; D'Souza, N.; and Kashyap, N. 2018. Effect of Dynamic Stoplist on Keyword Prediction in RAKE .

Bhattacharjee, A.; and Sengupta, D. 1996. On the coefficient of variation of the L-and L-classes. *Statistics & probability letters* 27(2): 177–180.

Bird, S.; and Loper, E. 2004. NLTK: the natural language toolkit.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2002. Latent dirichlet allocation. In *NIPS*, 601–608.

Cohen, J. D. 1995. Highlights: Language-and domain-independent automatic indexing terms for abstracting. *ASIS Journal* 46(3): 162–174.

Day, W. H.; and Edelsbrunner, H. 1984. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of classification* 1(1): 7–24.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Forina, M.; Armanino, C.; and Raggio, V. 2002. Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta* 454(1): 13–19.

Gat-Viks, I.; Sharan, R.; and Shamir, R. 2003. Scoring clustering solutions by their biological relevance. *Bioinformatics* 19(18): 2381–2389.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11(1): 1–21.

Hovy, E.; and Lin, C.-Y. 1998. Automated text summarization and the SUMMARIST system. 197–214. ACL.

Hulth, A. 2003. Improved automatic keyword extraction given more linguistic knowledge. 216–223. EMNLP.

Li, C.; Bendersky, M.; Garg, V.; and Ravi, S. 2017a. Related event discovery. 355–364. WSDM.

Li, G.; Liu, Y.; Ribeiro, B.; and Ding, H. 2017b. On group popularity prediction in event-based social networks. 644–647. ICWSM.

Likas, A.; Vlassis, N.; and Verbeek, J. J. 2003. The global k-means clustering algorithm. *Pattern recognition* 36(2): 451–461.

Liu, X.; He, Q.; Tian, Y.; Lee, W.-C.; McPherson, J.; and Han, J. 2012. Event-based social networks: linking the online and offline social worlds. In *KDD*, 1032–1040.

Maybury, M. 1999. *Advances in automatic text summarization*. MIT press.

Mehrotra, A.; Tsapeli, F.; Hendley, R.; and Musolesi, M. 2017. MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. *IMWUT* 1(3): 1–21.

Minh, L. N.; Shimazu, A.; Xuan, H. P.; Tu, B. H.; and Horiguchi, S. 2005. Sentence extraction with support vector machine ensemble .

Müngen, A. A.; and Kaya, M. 2017. A novel method for event recommendation in meetup. 959–965. ASONAM.

Pramanik, S.; Gundapuneni, M.; Pathak, S.; and Mitra, B. 2016. Can i foresee the success of my meetup group? 366–373. ASONAM.

Qiao, Z.; Zhang, P.; Zhou, C.; Cao, Y.; Guo, L.; and Zhang, Y. 2014. Event recommendation in event-based social networks. In *AAAI*.

Rose, S.; Engel, D.; Cramer, N.; and Cowley, W. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1: 1–20.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* 66(5): 688.

Strimmer, K. 2008. A unified approach to false discovery rate estimation. *BMC bioinformatics* 9(1): 303.

Wang, Y.; and Tang, J. 2019. Event2Vec: Learning Event Representations Using Spatial-Temporal Information for Recommendation. 314–326. PAKDD.

Wang, Z.; He, P.; Shou, L.; Chen, K.; Wu, S.; and Chen, G. 2015. Toward the new item problem: context-enhanced event recommendation in event-based social networks. In *ECIR*, 333–338.

Wu, H. C.; Luk, R. W. P.; Wong, K. F.; and Kwok, K. L. 2008. Interpreting tf-idf term weights as making relevance decisions. *ACM TOIS* 26(3): 1–37.

Xu, T.; Zhu, H.; Zhong, H.; Liu, G.; Xiong, H.; and Chen, E. 2018. Exploiting the dynamic mutual influence for predicting social event participation 31(6): 1122–1135.

Yilmaz, E.; Kanoulas, E.; and Aslam, J. A. 2008. A simple and efficient sampling method for estimating AP and NDCG. 603–610. SIGIR.

Zhang, C. 2008. Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems* 4(3): 1169–1180.

Zhang, J. S.; and Lv, Q. 2019. Understanding event organization at scale in event-based social networks. *ACM TIST* 10(2): 1–23.

Zhang, S.; and Lv, Q. 2017. Event organization 101: Understanding latent factors of event popularity. In *ICWSM*.

Zhang, S.; and Lv, Q. 2018. Hybrid EGU-based group event participation prediction in event-based social networks. *Knowledge-Based Systems* 143: 19–29.

Zhang, X.; Zhao, J.; and Cao, G. 2015. Who Will Attend?– Predicting Event Attendance in Event-Based Social Network. volume 1, 74–83. MDM.