On Participation in Group Chats on Twitter

Ceren Budak*
University of California
Santa Barbara, CA
cbudak@cs.ucsb.edu

Rakesh Agrawal
Search Labs
Microsoft Research
rakesha@microsoft.com

ABSTRACT

The success of a group depends on continued participation of its members through time. We study the factors that affect continued user participation in the context of educational Twitter chats. To predict whether a user that attended her first session in a particular Twitter chat group will return to the group, we build *5F Model* that captures five different factors: *individual initiative, group characteristics, perceived receptivity, linguistic affinity,* and *geographical proximity*. Through statistical data analysis of thirty Twitter chats over a two year period as well as a survey study, our work provides many insights about group dynamics in Twitter chats. We show similarities between Twitter chats and traditional groups such as the importance of social inclusion and linguistic similarity while also identifying important distinctions such as the insignificance of geographical proximity. We also show that *informational support* is more important than *emotional support* in educational Twitter chats, but this does not reduce the sense of community as suggested in earlier studies.

Categories and Subject Descriptors

H.0 [**Information Systems**]: General

Keywords

group dynamics; online communities; Twitter chats; speech codes theory; ostracism; information overload; mixed methods

1. INTRODUCTION

Human beings are social animals that are scarcely able to lead a solitary life (Baruch Spinoza, Ethics, IV, proposition 35:note). They interact and form relationships with others that endure from one encounter to another. A central theme in the study of human behavior is what makes a person become a member of a social group [16].

With the emergence of Web, many of our group interactions are moving on-line. One example of this move is the curious emergent phenomenon called Twitter chats. Twitter chats are public conversations, regularly held on Twitter on specific topics at designated times. For instance #engchat is a chat about English education held at 7-8pm EST on every Monday. During a chat session, the participants continuously interact on the designated topic by tweeting their opinions and marking their tweets with the hashtag of the particular chat group. While weekly groups like #engchat are the most common ones, there are others such as #mathchat that meet twice a week, #collegechat that meet bi-weekly or #edchat that are week-long conversations. Most of the chat groups also have dedicated blogs that provide various resources such as transcripts of past sessions and schedule of upcoming discussions. See [14] for a crowd-sourced list of Twitter chat groups.

The research question we investigate is what factors ensure continued individual participation in a Twitter chat. More specifically, our goal is to develop a model that can predict whether a person attending his/her first chat session in a particular Twitter chat group will return to the group. Figure 1 pictorially depicts the *5F Model* we employ. Abstracting from the years of literature on group interactions in the physical off-line world [16, 23, 30, 34, 37, 40, 42, 45, 51], we identify five major factors that affect the participation of a person in a group: *individual initiative, group characteristics, perceived receptivity, linguistic affinity*, and *geographic proximity.* The Twitter specific measures corresponding to each of these factors are shown in boxes along the edges. For example, the number of tweets, the number of URLS in the tweets, the number of mentions and retweets contributed by the person during her first session provide indication of her individual initiative. Using data from thirty education-

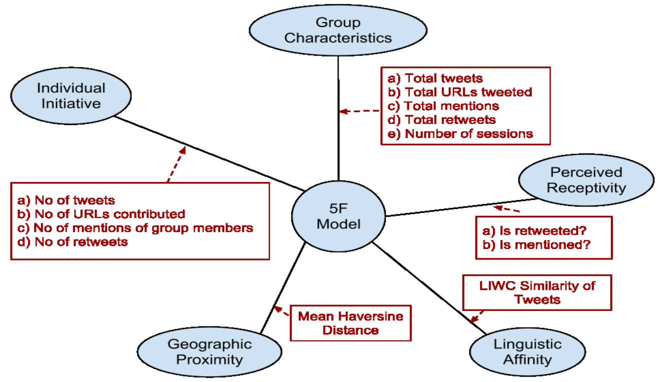


Figure 1: Overview of the *5F Model*

*Work done while the author was an intern at Search Labs.

WWW 2013, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2035-1/13/05.

related chat groups, we study the predictive power of these factors individually as well as collectively.

The paper proceeds as follows. We begin with a discussion of related in work in §2. We then present the *5F Model* in §3. The data sets used in the study are described in §4. We present the results of our statistical analysis in §5. This analysis delineates effectiveness of various Twitter measures in predicting the continued participation. We also carried out a survey to complement the statistical analysis. The findings from this survey are presented in §6. We conclude with a summary and directions for future work in §7.

2. RELATED WORK

Researchers in social sciences have long been studying group dynamics, focusing on different aspects such as formation, structure, cohesion or leadership [16]. In this paper, we focus on a key question relating to group dynamics: the ability of groups to sustain member relations. The social science literature on this topic is rich. Individual characteristics [30], group characteristics [16, 23, 40, 42, 45, 51], use of language [37] and geographical factors [34] are among the notions that have been studied and claimed to affect the individual-group interactions. Much of the classical social sciences research has been carried out for off-line groups.

With the advent of the Internet, groups and communities arose in the virtual world as well. Early research on online groups focused on structural characteristics of Usenet [7], listserv [35] and email groups [15]. With the arrival of next generation of social networks, the researchers had the opportunity to study group dynamics in online forums such as Yahoo groups [5], Google Groups [20], LiveJournal [3] and SecondLife [18]. Research questions addressed in the literature are as varied as the type of social networks studied and include topics such as structural properties [7], identification of sense of community [48], modeling individual behavior [5], use of language [20], identification of conversational themes [11], and evolution of the online communities [3].

The focus of our paper, i.e. the factors necessary for continued participation of online community members, have been studied in [1, 6, 8, 10]. In [1], authors study Usenet newsgroups and identify characteristics that increase the likelihood of a particular post receiving a reply and a poster coming back to newsgroups. They find that posters are less likely to get a reply if they were newcomers. Posting on topic, asking questions, and using less complex language are also identified as being important. In [8], authors analyze 200 responses to an open question regarding community loyalty in five Norwegian online groups and identify nine main reasons for decreasing participation over time: lack of interesting attendees, low quality content, low usability, harassment, time sink, low trust, over-commercialization, dissatisfaction with moderators and boring content. Chan et al. [10] identify three different forms of perceived recognition in a virtual community, namely identity, expertise and tangible recognition through an interpretive case study. Through surveys, they claim that members share their expertise because it makes them feel self-efficacious. In determining what makes an individual return to a particular community, Bateman et al. [6] study *commitment theory* originally introduced to study motivations behind voluntary work in online communities. Their survey based study reveals that affect-based, norm-based, and cost-based bonds collectively drive participation behavior in online communities.

In summary, related research on online communities mostly focused on a single factor in determining future participation while we introduce a framework that captures various factors. Also, we employ a mixed-method approach while the related work employs either a qualitative [8, 10] or quantitative [1, 6] solution but not both. With our approach, we can study group dynamics at scale as well as in detail. Finally, related work in online communities focused on asynchronous groups while we study Twitter chats that provide a much better proxy to synchronous face-to-face interactions.

3. 5F Model

We now present our *5F Model* for studying continued individual participation in a Twitter chat group. The schematic we use is as follows. For each factor in the model, we first discuss the prior literature from which the factor has been abstracted. In some cases, we also identify subfactors that comprise the corresponding factor. We then discuss the Twitter specific measures for their computation.

3.1 Individual Initiative

3.1.1 Background

Certain people are more likely than others to seek out membership in groups due to differences in personal characteristics or motivations [16]. The big five theory [30] claims that people differ in five dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness. These differences are claimed to affect the likelihood of participating in groups. Of the five dimensions, *extraversion* is found to be a particularly influential determinant [16]. The importance of individual-level characteristics such as high-engagement and longevity in group interactions have also been observed in online communities such as Yahoo Groups [5].

3.1.2 Twitter Measures

In order to capture the importance of user-specific characteristics in group interactions, we consider the actions taken by the user at the first chat session that she attends. The independent variables that relate to individual initiative are:

1. *usertweetcount* denotes the number of tweets the user contributes to the session. This measure captures the *extraversion* dimension of the big five theory.
2. *userurl* denotes the number of urls the user contributes to the chat session. This variable acts as a proxy for informational contribution by the user.
3. *usermentions* is the total number of times the user mentions another (by using @). This measure captures how much the user engages in conversations.
4. *userretweets* is the number of retweets by the newcomer user and captures the amount of information she found to be worth sharing with her followers.

3.2 Group Characteristics

3.2.1 Background

The experience that a user has in attending her first chat session depends on the context which is shaped by the actions of the others in the group. Therefore, these group-level characteristics can affect the decision of the newcomer to return to the group. For instance, [1] shows that group-level factors can be predictive of future participation in online blogs. We next identify various group-level factors and study their significance. Four main themes studied are:

Amount of Information: *Information overload* refers to a person's state in which the overwhelming amount of information leads to communication inputs not being processed and utilized [40]. Information overload is especially important in the online context. For instance, in a recent work on Usenet newsgroups it has been shown that users are more likely to end active participation as the overloading of mass interaction increases [23].

Conformance *Informational influence* refers to group members using responses of others as reference points and informational resources [16]. This concept relates to *conformity* and has been observed in experimental studies in offline groups [42]. Conformity can result in coherent groups, eliminating controversies. Yet, too much conformity also throttles diverse perspective with reduced value for new users.

Inter-member relations *Intermember relations* play a critical role in determining whether a newcomer will come back to the group. Too much or too little dyadic interaction between group members can affect how a newcomer views the group [16]. The significance of inter-member relations has also been studied in [4] which focuses on individual arrival patterns in group discussions.

Group Maturity Various group formation theories (e.g. *stages of group development* [45]) assert that groups become cohesive over time where uncertainty about goals and roles are resolved. At the same time, such groups can become closed [51] to new members.

3.2.2 Twitter Measures

1. *sessiontweetcount* denotes the number of tweets in the chat session and captures the *amount of information*.
2. *sessionurl* is the number of urls shared in a chat session. This measure also captures the *amount of information*. We study *sessionurl* as a separate factor (in addition to *sessiontweetcount*) since tweets with URLs tend to be more informational than ordinary tweets.
3. *groupretweets* is the *number of retweets* in the chat session and captures conformity in the group.
4. *groupmentions* denotes the *number of mentions* in the chat session and quantifies *intermember relations*.
5. *groupmaturity* is the age of a group at a date D, and is computed as the *number of sessions* held until D.

For all these measures, we discount the tweets shared by the newcomer user for whom prediction is being made since the goal is to capture the context that the user *interacts* with rather than the context she *creates*.

3.3 Perceived Receptivity

3.3.1 Background

Research in social sciences has established the importance of social inclusion in one's desire to affiliate with both traditional [31, 39, 44] and online [1, 24] groups. The importance of social inclusion is generally studied under *ostracism* which refers to the act of individuals or groups excluding or ignoring others [28, 49]. While some studies found that the excluded members respond by leaving the group [31, 39], others found increased desire to belong as a response [44].

3.3.2 Twitter Measures

We capture perceived receptivity through two variables:

1. *ismentioned* denotes whether the user is mentioned by at least one person in the chat session.
2. *isretweeted* indicates whether the user is retweeted.

3.4 Linguistic Affinity

3.4.1 Background

Language is a key element in social interactions [2, 12, 21]. *Speech Codes Theory* [37] encapsulates this notion and states that "whereever there is a distinctive culture, there is to be found a distinctive speech code". This theory affirms individuals as either within or outside of the social structure by how in sync their language is with the language of the group. It has been shown that linguistic similarity has positive correlation with lasting *dyadic* relationships [21]. This effect has also been observed for group interactions in a recent study carried out in parallel to our study [13]. Another related work shows that the power differentials between individuals can be revealed by how much one echoes the linguistic style of the person they are responding to [12].

3.4.2 Twitter Measures

We make use of *Linguistic Inquiry and Word Count (LIWC)* to compare linguistic markers between a user and a group. *LIWC* is a text analysis software that calculates the degree to which people use different categories of words across a wide array of texts [36]. *LIWC* uses positive or negative emotions, self-references, causal words, and 83 other dimensions and has been used in various studies [21].

We consider the set of tweets a user u_i shares in her first session as a text document and compute the value of each linguistic marker to obtain her *LIWC-vector* for that particular session. Similarly, we aggregate all the tweets from users other than u_i and compute the *LIWC vector* of the group. To identify the similarity, we use Pearson correlation measure that provides the degree of linear relationship between two vectors. This measure ranges from +1 to -1, with a large positive value indicating similar linguistic usage. We refer to this measure as *liwccors* in our model.

3.5 Geographic Proximity

3.5.1 Background

Proximity Principle [34], a theory introduced in offline groups, states that people tend to join close by groups. Unlike the offline context, recent research suggests that online social networks can overcome this barrier [18, 27]. However, [38, 46] show that geography plays an important role in our use of language or choice of friends even in the online world. Hence, the role of geography is context-dependent and cannot be dismissed without investigation.

3.5.2 Twitter Measures

To study the influence of geographical proximity, we calculate the mean distance of the user to everyone else in the group. The location for each user is determined based on the location field of the user profile. We mine for the following patterns: latitude-longitude, {city, region, country}, {city,region}, {city,country}, {region,country} and country. For this purpose, we make use of data available from [29], which contains complete hierarchical information and coordinates for nearly 50,000 cities from all over the world. We convert all mined patterns to latitude-longitude pairs. For cities, we rely on the latitude-longitude points provided by the Gazetteer [29] and for regions and countries, we use the latitude-longitude of their most populated city. Given this data, the distance d (in meters) between two users u_i and u_j with locations (lat_i, lon_i) and (lat_j, lon_j) can be computed through Haversine formula [43]:

$$a = (\sin(dlat/2))^2 + \cos(lat_i) * \cos(lat_j) * (\sin(dlon/2))^2$$
$$c = 2 * \arcsin(\min(1, \sqrt(a)))$$
$$d = R * c$$

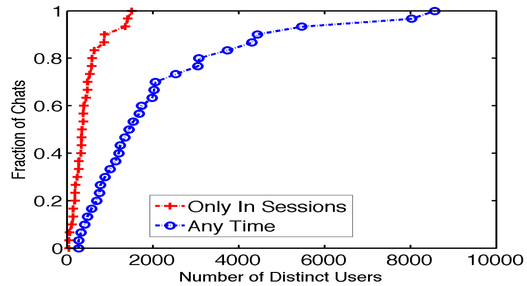
where $dlon = lon_j - lon_i$, $dlat = lat_j - lat_i$ and R is the radius of the earth. The proximity of a user u_i to group g_x at a given session is computed as the mean distance between u_i and all other users u_j who attended that chat session. This measure is referred to as *distance* in our regression tasks.

4. DATA SET

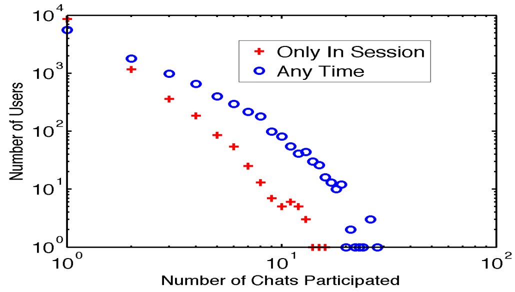
4.1 Group Chats Studied

We examined over 100 education related chats appearing in a crowd-sourced list on the Web [14]. We chose to focus on education-related groups because of potential synergies with the current research in MOOCs [9]. In addition, relevant research underlines the informational power of Twitter for educators [17] and this power could potentially be enhanced through an improved community experience. Given the list in [14], we identify all tweets in each chat by capturing tweets with the corresponding hashtag. For this purpose, we analyze *all* Twitter updates from June 2010-July 2012. Next we filter the list to eliminate chats with the first identified tweet before September 2010 or after September 2011. This eliminates chats that can potentially pre-date our data collection or do not have a sufficient number of sessions.

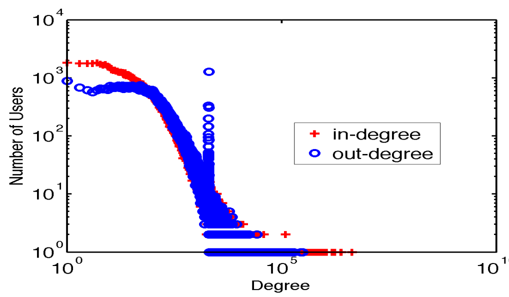
Hashtags that represent these chats are also used outside the scheduled chat sessions. For instance, a user sharing a tweet that relates to teaching can use the hashtag #teachchat even if the tweet is not shared during an actual session. In order to filter out such chatter, we rely on the bursty activity that is a general characteristic of hashtags that are associated with scheduled chat meet-ups. By identifying the hours of high activity (and including the preceding and following hours in order to capture the anticipation and wrap-up tweets), we capture the sessions for each chat. The hours of high activity are defined as those that have a five-fold or more increase w.r.t. the earlier time frame. Our manual inspection has revealed that this simple technique accurately detects chat sessions. Next, we further filter the list of chats to only those with ≥ 10 sessions, which reduces the list of chats to 30. We also considered relying on advertised chat hours to capture chat sessions. However, this resulted in incorrect characterization as chat schedules get changed and various chat groups skip certain weeks, especially during summer.



(a) CDF of Number of users in per chat



(b) Log-log scale plot of the number of chats per user



(c) Log-log scale plot of degree distribution

Figure 2: Characterization of Twitter Chats

Table 1 gives an overview of the chats studied. The columns of this table are: chat name, total number of tweets, total number of distinct users, number of sessions, and popular locations. Popular locations are identified by counting tweets from a given location and selecting the top 5 locations.

4.2 Salient Statistics

We next provide high level characterization of the chats studied as well as the users that participate in them.

Distribution of the number of users in and outside chat sessions: The distribution of the number of distinct users in each chat is provided as a CDF graph in Figure 2(a). The X-axis denotes the number of distinct users while the Y-axis denotes the fraction of the chats studied that have at most that many users. Here, we distinguish between user activity within and outside of chat sessions. The curve with + markers provides the distribution of users that participate in actual chat sessions while the other curve gives the distribution of users that tweeted same the hashtag at least once at some point of time (irrespective of whether the tweet was during a chat session). The difference between the two curves shows that a large number of users do not partici-

chat name	discussion topic	# tweets	# users	# sessions	most popular locations
#eltchat	English language teaching	90445	3515	95	Athens, Oxford, North Yorkshire, Stuttgart
#sschat	Social Studies	79455	6351	86	Illinois, Ogden, Berkeley, Chicago, Plymouth
#kinderchat	Early childhood education	40851	2436	80	Princeton, Ontario, North Canton, Kansas
#engchat	English teachers	51894	6757	65	Pennsylvania, Chicago, New Jersey, Iowa, Michigan
#langchat	Language teaching	26621	2029	60	Louisville, Napa, Michigan, Evansville, Newton
#edchatie	Irish educators/education	24167	1575	59	Ireland, Dublin, Clonmel, Nenagh, Galway
#libchat	Librarian discussions	11120	954	58	Tallahassee, Ohio, Carrollton, Indianapolis, USA
#4thchat	4^{th} grade teaching	18712	1663	57	New Orleans, Massachusetts, Colorado, Michigan, Ontario
#phdchat	Current, former or aspiring PhD researchers	53717	4524	57	UK, Melbourne, Sussex, London, New Zealand
#asechat	Science education	14254	1106	52	UK, Cardiff, London, York North Yorkshire, Bristol
#5thchat	5^{th} grade teaching	13685	1240	48	Ontario, Georgia, USA, Dublin, San Antonio
#isedchat	Independent school educators	18261	1661	46	USA, Florida, Connecticut, Portland, Boston
#1stchat	1^{st} grade teaching	11625	961	44	Hershey, Woodstock, Vancouver, Rochester, Montana
#addcym	Welsch education system	9639	583	44	Cupertino, Cardiff, Swansea, UK, London
#fycchat	First year composition	5857	467	42	Dallas, Alabama, Minneapolis, Kansas City, Spartanburg
#gtie	Gifted and talented network Ireland	7135	341	38	Dublin, Wicklow, Ireland, United Kingdom, New Zealand
#spedchat	Learning issues	23993	3578	37	Maryland, New York, USA, Wichita, Ohio
#pblchat	Project-based learning	16570	2365	32	Napa, Portland, Tacoma, Round Rock, Dallas
#teachchat	All about teaching	7273	693	30	Florida, Fort Worth, Lake Forest, California, USA
#atplc	Professional Learning Communities	8065	1196	28	Bloomington, Iowa, Chicago, San Diego, Mankato
#titletalk	How to promote reading	14069	1182	24	Bedford, Texas, Michigan, Ohio, Los Angeles
#k12media	K-12 Education	2346	236	23	Toronto, Canada, Chicago, Ontario, Illinois
#jedchat	Jewish educations	9196	585	22	Israel, San Francisco, New York, Boston, USA
#flipclass	Flipped classroom	19313	2847	21	Lake Forest, Evansville, Kelowna, Texas, New Jersey
#digcit	Digital Citizenship	4194	919	15	Birmingham, USA, Texas, Natick, Indianapolis
#satchat	School leadership	4543	702	15	New Jersey, Jericho, Virginia, Nebraska, Philadelphia
#tichat	Tech Integration	4231	745	15	Sachse, Pittsburgh, Texas, Ohio, Burlington
#ageduchat	Agricultural education	2387	284	14	Michigan, Raleigh, Iowa, Indianapolis, Wisconsin
#globalclassroom	Global classroom project	6614	642	11	New Jersey, New Zealand, Melbourne, Bandung, Fort Worth
#slpchat	Speech language pathologists	4053	397	11	Sydney, Barbados, Maryland, Indiana, North Dakota

Table 1: Education Chats Studied

pate in regular chat sessions but share at least one tweet with the corresponding hashtag. Such users are excluded from our analysis in identifying members of chat groups.

Distribution of the number of distinct chats users participate in: Even though the specific goal of chat groups studied in this paper vary, they are all related to education. We found that users participate in more than one of them. In Figure 2(b), we plot the distribution of the number of unique chat groups users participate. The plot shows a skewed distribution. While a large fraction of the population participates in only one chat group, there are a small number of users that participate in a large number of groups.

Degree distribution of education chat users: In Figure 2(c), we show the the in- and out-degree distributions of Twitter users that participated in at least one education chat. The plot shows a skewed distribution. Most users follow and are followed by a small number of users while there are a small number of users that follow (or are followed by) a large number of users. The skewed distribution (as well as the spike in the number of friends at ≈ 2000) is in agreement with related work in general Twitter population [26].

Geographical distribution of education chat users: We could geo-tag over 50% of the users using the mining procedure described in §3.5.2. This number is significantly larger compared to the 21% for the general Twitter population [50]. Table 1 provides an overview of geographical distribution of each chat by listing the top-5 locations. It shows that education chats are distributed all over the world. Figure 3 plots users of 3 interesting education chats, namely #globalclassroom, #5thchat and #edchatie. The three chats show distinct characteristics. While #globalclassroom is distributed all over the world, #5thchat is mostly in U.S., and #edchatie is mostly in Ireland. We note that #edchatie is a chat formed by Irish educators to discuss education in Ireland.

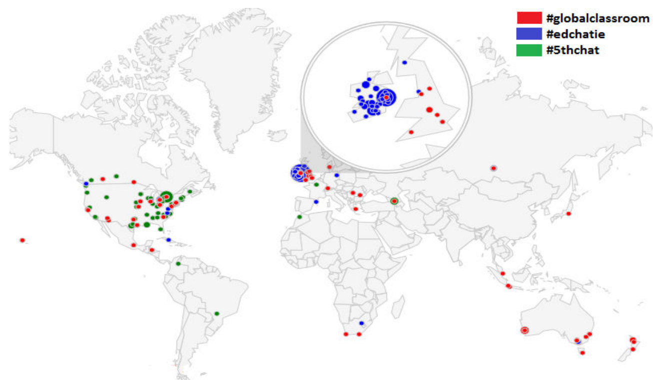


Figure 3: Geographical Distribution of Three Chats

5. STATISTICAL ANALYSIS

To predict whether a first-time visitor to a group will return for at least one more session in the future, we build individual models for the five factors (*individual-initiative, group characteristics, perceived receptivity, linguistic affinity,* and *geographical proximity*) as well as the unified *5F Model*. We use logistic regression [19] for statistical analysis and a *Pseudo-R* measure to compare the models. In particular, we employ the commonly used *Nagelkerke* R^2 *Index* [33] which can be computed as: $\frac{1-(L(M_{intercept})/(L(M_{full}))^{2/N}}{1-L(M_{intercept})^{2/N}}$, Here N is the size of the data set, $L(M_{full})$ is the likelihood of the model given the data set, and $L(M_{intercept})$ is the likelihood of the null model. Larger Nagelkerke values indicate a better fit.[1]

We performed this analysis at the individual chat group level as well as by combining tweets from all the chat groups. We present the results only for the combined case. The results for individual chat groups were similar, except for the geographic proximity model. We discuss those differences along with the discussion of the geographic proximity model.

5.1 Results

The regression results are summarized in Table 2. This table has four columns. The first column is the name of the

[1] We also calculated *Akaike Information Criterion (AIC)* value for each model [19]. *AIC* results were consistent with the *Pseudo-R* results and are not included.

model and corresponds to one of the five factors. The second column lists the Twitter specific variables used for each of the corresponding factors. The third column consists of two subcolumns. The first subcolumn shows the cofficients of the corresponding explanatory variables in the individual-level models, whereas the second subcolumn gives the coefficients for the unfied model. The third column gives the pseudo-R measure for the individual models. The pseudo-R value for the unfied model is 0.14 and is shown at the bottom of the table. The statistically significant variables are marked with * for p-value < 0.05, ** for p-value < 0.01 and *** for p-value < 0.001.

Individual Initiative Model: The results show that all the variables except for *usermentions* are statistically significant. The number of tweets are positively correlated with returning to the chat group, emphasizing the predictive power of early interest exhibited by the user. The variable *userurl* is negatively correlated with returning to the group. One possible explanation for this result can be given as follows: For users that share a large number of urls, i.e. users that already acquire a certain level of knowledge, the added informational gain from chat sessions can be smaller, resulting in less incentive to attend future sessions.

The negative correlation for *userretweets* indicates that retweeting behavior can be used to distinguish *real* participants of chat groups from those that are merely retweeting the tweets of their friends who are attending a chat session. Consider the following illustrative scenario. Assume that *user1* attending *#1stchat* shares a tweet "Check out article bit.ly/342dfser #1stchat". This tweet is seen not only by the attendees of *#1stchat* but also the followers of *user1*. One such follower, say *user2*, can find the tweet interesting and retweet it. Here, *user2* who *appears* to be attending his first *#1stchat* session may not return to this group.

Group Characteristics Model: Statistically significant variables are *groupretweets, sessiontweetcount, sessionurl* and *groupmaturity*. Capturing the significance of information overload, *sessionurl* and *sessiontweetcount* have negative correlation. The variable *groupmaturity* has negative correlation with the odds of come back, i.e. users that attempt to join more mature groups are less likely to return to the group. The results also indicate the significance of *informational influence* as demonstrated by the statistical significance and positive correlation of *groupretweets*. However we observe that the correlations of these factors are relatively mild. For instance, an increase of 1 retweet in group discussion decreases the log odds of come back by 0.0014. *Pseudo-R*(=0.03) values for this model are worse when compared to those of *individual initiative model*, showing that *individual initiative* factors are relatively better indicators of future participation.

Perceived Receptivity Model: Both *ismentioned* and *isretweeted* are statistically significant and positively correlated with returning to a group. Correlation measure is strong emphasizing the value of social inclusion. This result is in agreement with relevant research in other online communities [1, 24]. The *Pseudo-R*(=0.08) value for this model is the third best among the individual models.

Linguistic Affinity Model: Linguistic similarity is statistically significant and highly correlated with returning to a chat group. This finding supports our hypothesis and is in line with research in social sciences, particularly the *speech codes theory*. The highest *Pseudo-R* value for this model shows that the linguistic characteristics are the best indicators of future participation. In addition to evaluating linguistic similarity in an aggregate level, we also performed regression to identify the statistical significance of similarity along individual LIWC dimensions. Our results show that the use of conjunctions (p-value = 0.002, examples: but, and), discrepancy words (p-value=0.003, examples: should, would), causal words (p-value=0.03), WPS (words per sentence, p-value < 0.0001), punctuation letters (p-value=0.02) were among the most significant dimensions.

Factors	Variables	Coefficients		Pseudo-R
		Individual Model	Unified 5F Model	
Individual Initiative	usermentions	-0.016	-0.007	0.09
	userretweets	-0.13***	-0.077***	
	userurl	-0.16***	-0.092***	
	usertweetcount	0.147***	0.05***	
Group Characteristics	groupmentions	-0.0001	-0.0004	0.03
	groupretweets	0.0014*	0.002***	
	sessionurl	-0.003***	-0.002*	
	sessiontweetcount	-0.0005	-0.0008*	
	groupmaturity	-0.01***	-0.007***	
Perceived Receptivity	ismentioned	1***	0.445***	0.08
	isretweeted	0.69***	0.24	
Linguistic Affinity	liwccors	2.159***	1.215***	0.1
Geographical Proximity	distance	-0.00005***	-	0.01

Pseudo-R for the unified *5F Model* = 0.14
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 2: Results of Statistical Analysis

Geographical Proximity Model: We see that returning to a group is only mildly correlated with geographical proximity. An increased distance of 1km reduces the log odds of returning to the group by only 0.00005. Regression tasks performed *per-chat group* showed that geographical proximity is statistically significant for only seven educational Twitter chats. Two chats had positive correlation and five had negative correlation. For instance, #globalclassroom has positive correlation with the variable *distance*, indicating the positive effect of diverse locations in returning to the group. Such behavior is to be expected given the global goal of this particular group. Yet groups like #jedchat have negative correlation with increased distance. This group is on Jewish education and is mostly popular in Israel. Overall, the *Pseudo-R* value for this model is the worst among all models, showing that geographical characteristics are generally poor indicators of future participation.

The Unified 5F Model: In this model, we consider all the explanatory variables in conjunction, except geographic proximity (*distance*). The reason for omitting the latter is that we could determine the location of only a subset of users and this factor anyway turned out to have limited fit. As expected, this model has the largest *Pseudo-R* value. Each independent variable has similar explanatory trend as we observed with individual models.

5.2 Discussion

Our results show various parallels between Twitter chats and other offline and online groups. For instance, the importance of social inclusion that has been identified in both offline and online groups [28, 31, 39, 44, 49] is also significant in Twitter chats; newcomers that are mentioned or retweeted by others in a chat session are much more likely to come back to the chat group. Also, the use of language plays a critical role; newcomers that have a similar language to a given group are likely to continue participation. While the similar use of words related to education might be anticipated to effect participation in educational Twitter chats, our results go beyond this finding. We show that similarity in the language along dimensions such as sentence complexity (words per sentence), causal words, or even punctuation letters reveal much about group interactions. In other words, language defines individuals as either *insiders* or *outsiders* of groups in Twitter chats as suggested by speech codes theory [37].

We also found that users are less likely to return to established groups, i.e. groups that have held a large number of chat sessions, indicating that groups become closed to new members over time [45]. The total number of tweets and urls shared in the chat session also decrease the likelihood of returning to a group. While the idea of more information driving away participants can sound counter-intuitive at first, it can in fact be explained by the effects of information overload [23, 40]. The large scale and synchronous nature of discussions held in Twitter chats likely magnify the effects of information overload that is already prevalent in asynchronous online communities [23]. Our survey study also supports this finding as we will demonstrate in §6.

One major distinction of Twitter chats from traditional groups is the insignificance of geographical proximity. Unlike offline groups [34], we show that there is no correlation between geography and group affiliation for most Twitter chats, the exceptions being chats organized around local topics (e.g. #jedchat on Jewish education in Israel). In fact, geographical diversity is considered as the most important *advantage* of Twitter chats as demonstrated in §6.

Turning to the *relative* importance of the factors, the best individual models to capture continued user participation in order are: *linguistic affinity, individual initiative, perceived receptivity, group characteristics* and *geographical proximity.* This ordering reveals distinctions of Twitter chats from other groups studied in the literature. For instance, related research claims that individuals are initially drawn to online communities by a desire to interact with like-minded others but whether they return is significantly influenced by the content of the community [6]. Yet, our analysis shows that individual characteristics are better indicators of future participation than group characteristics. In addition, past research emphasized the value of perceived receptivity as the dominant factor [10] while we find it to be the third statistically determining factor.

The impact of linguistic similarity in group participation is largely unexplored, yet we found this factor to be the *best* predictor of future participation. We note that the 140 character limit that Twitter enforces for tweets likely introduces added challenges to the newcomers as this limit drives participants to use shorthand descriptions for various notions which can be confusing for the newcomers. This finding is also supported by our survey study.

6. USER SURVEY

We complemented the results from the statistical data analysis with a user survey to directly understand from users involved in Twitter chats their attitudes towards these chats. We circulated an online survey of 26 questions (through Twitter) to users that participated in education chats. Table 3 lists the questions asked in our user survey. The options of multiple choice questions with one possible answer are marked as (a), (b), and so on. When more than one option can apply, they are marked as (i), (ii), and so on. The survey was publicized in educational Twitter chats through the hashtag of each chat group studied. Respondents of the survey were encouraged to share the survey with their Twitter followers. In all, sixty users responded to our survey.

Characteristic	No of survey respondents
The sense of belonging	26
Emotional Support (Receiving encouragement, being listened to or sharing feelings)	17
Informational Support (Advice, guidance, or links to new useful tools shared in group discussions)	57
Instrumental Support (tangible resources shared by the members such as assisting with work or providing favors)	36
Networking with friends/colleagues	46
Making new friendship/professional connections	41

Table 4: Uses of Twitter Education Chats

Fifty of the survey participants identified themselves as educators. One identified himself as a student. Three participants stated that they were both an educator and a student. Three stated that they were both an educator and a parent of a student. Fourteen, nineteen, twenty three and four of the survey participants stated that they participate in one, two, three-to-five and more than five education chats respectively. The number of distinct Twitter education chats participated by our survey respondents is sixty six. [2]

6.1 Findings

The survey had three main parts, addressing questions related to: (1) usage, advantages and disadvantages, (2) sense of community and responsibility, and (3) evolution of participation. We discuss findings for these segments next.

Usage, Advantages and Disadvantages: Table 4 shows the number of survey participants who identified a particular characteristic of Twitter chats. The results show that most users value the informational support provided in Twitter chats. The ability of Twitter to connect people with others is another important theme. The results show an interesting distinction between educational Twitter chats and other online groups [47, 48] for which emotional support is found to be more significant. In fact, in past research, sense of community is found to be negatively correlated with informational support [48]. Yet, as we will discuss later, sense of community is strong in Twitter chats despite the fact that informational support is more dominant than emotional support.

When asked to identify the advantages of Twitter chats compared to face-to-face meetings and online groups, the survey respondents gravitated towards similar high level observations even though they were not provided with a set of predefined options. The results are presented in Table 5. The most common advantage identified is the diversity in backgrounds and geographical locations of chat participants.

[2]Since the population in our survey does not constitute a statistically sound representative sample, the reader should view the findings as anecdotal. We also note the possibility of positive bias towards Twitter groups since those who left them are less likely to have responded to the survey. The results are instructive, nonetheless.

Introduction
1) What is your twitter username? (Twitter username can be found on your profile page and starts with '@')
2) Are you... (a) An educator (b) A student (c)A parent of a student (d)Other: [specify]
3) How many different twitter chats do you participate in? (a) 0 (b) 1 (c) 2 (d) 3-5 (e) more than 5
4) How many of those chats are related to education? (a) 0 (b) 1 (c) 2 (d) 3-5 (e) more than 5
5) Please provide a comma-separated list of the names of these twitter chats (The name of the chat is the hashtag that is used to organize is.)
Uses, Advantages, and Disadvantages of Twitter chats
6) What are some of the most important characteristics of twitter chats for you?
(i) The sense of belonging (ii.) Emotional Support (for instance receiving encouragement, being listened to or sharing feelings)
(iii.) Informational Support: Advice, guidance, or links to new useful tools shared in group discussions
(iv.) Instrumental Support: Tangible resources shared by the members such as assisting with work or providing favors
(v.) Networking with friends/colleagues (vi.) Making new friendship/professional connections
(vii.) None of the above. Please list other important characteristics that are not listed above [specify]
7) What do you think is the most important advantage of twitter chats over other chat forms (like face-to-face meet ups or blog chats)?
8) What do you think is the most important disadvantage of twitter chats compared to other chats (like face-to-face meet ups or blog chats)?
9) Please give one or two examples of something you learned the last time you participated in a chat.
10) Have you been able to convince others that you work with to join Twitter chats? (a) Yes (b) No If so, how many? [specify]
Sense of Community and Responsibility
11) Do you communicate with other participants (in education chats) outside of the chat session hours? If so, please select the options that apply
(i) Over twitter (follow, mention or retweet) (ii) Other online means such as emailing or blogging
(iii) Off-line (examples: face-to-face meet-ups, phone calls) (iv.) Other: [specify]
12) Do you feel a sense of community in twitter chats? (a) Yes (b) No Please elaborate.
13) Do you feel a responsibility to the community to participate in chat sessions? (a) Yes (b) No (c) Other: [specify] Why? (or why not?)
14) Please check any of the following actions that you have performed for the chat group
(i) Moderating (ii) Recommending novel ideas for discussions, approaches, solutions (iii) Providing data/facts/tools useful for making decisions
(iv) Giving your opinion on topics (v) Refocusing or stimulating discussions that flag (vi.) Taking notes or providing the archives for the chat
(vi.) Verbally evaluating the quality of discussion in chat sessions as well as the results of discussions (vii.) Engaging others in discussion (for instance through @mention) (viii.) Publicizing the chat (ix.) A task that is not listed here (x.) I do not perform any task
Any other task you can think of that is not included in this list? [specify]
15) Do you feel the need/urge to contribute to group by carrying out specific tasks? (a) Yes (b) No (c) Other: [specify]
16) If your answer to the previous question was yes, can you elaborate more? Do you consistently carry out this task? Is it self-assigned or assigned by the community? How long have you been holding this task?
Evolution
17) How did you first hear about the chats you participate in? In case you participate in more than 1 such chat, please mark all that apply
(i) Through another twitter chat (ii) Through general twitter usage (iii) Web search (iv) Education related forum/blog (v) Facebook
(vi) Email (vii) Offline connections (through a friend, colleague etc.) (viii) I founded/co-founded the chat (ix) Other: [specify]
18) Please think back to the first time you participated in a education-related twitter chat. What were your original goals in participation?
(i) Out of curiosity, to explore (ii) To learn new information/tools/methods (iii) To make new connections (iv) To communicate with the friends and colleagues you already knew (v) For the sense of community, belonging (vi) For emotional support: to share feelings/frustrations
(vii) For receiving help on tasks (viii) None of the above. Any other reasons not listed above? [specify]
19) Thinking back to your very first chat session, what was your first impression? What were the difficulties, the positive and the negative surprises? What made you come back to attend another session (or not)?
20) Thinking back to your initial impression of the chats and comparing it with your current view, what has changed?
21) Did your view of the chat and what it entails change over time? (a) Yes (b) No. If so, how?
22) Did your your position or responsibilities in the group change? (a) Yes (b) No. If so, how?
23) Have you made new friends, personal/professional connections? (a) Yes (b) No. If so, how?
24) Overall, what is the most significant change twitter chat created for you?
25) Any way in which the quality of discussion can be improved?
26) Can you think of question(s) that I should have asked but did not?

Table 3: Survey Questions

Advantage	No of survey respondents
Diversity in backgrounds and geography	26
Convenience	25
Ease of sharing information	10
Ability to archive and search older chats	9
Public form and equality	3

Table 5: Advantages of Twitter Chats

Disadvantage	No of survey respondents
Pace and Amount of Information Flow	9
Twitter syntax	6
Lack of face-to-face interactions	5

Table 6: Disadvantages of Twitter Chats

This finding is in line with the result of our data analysis that showed that geography is not a limiting factor for most chat groups. One interesting advantage identified by the survey respondents notes a significant difference between face-to-face groups and Twitter chats: the ability to record group discussions.

Twenty six survey participants provided disadvantages of Twitter chats. Three most dominant themes are listed in Table 6. Interestingly, lack of face-to-face interactions was only at the third place while it has been found to be the dominant factor in the literature [22]. Instead, the most important disadvantage is the pace and amount of information. This result is a direct implication of the unique characteristic of Twitter chats since unlike other online groups, they introduce the added challenge of interacting with a large crowd in a synchronous manner. We note that three people indicated that there were *no* disadvantages of Twitter chats. The rest (five survey participants) touched on various topics such as the public characteristic of conversations limiting negative discussions or the timing of the chats.

Sense of community and responsibility: Fifty three of the survey respondents stated that they felt a sense of community in Twitter education chats while three stated that they did not. The ties built during the course of chat sessions extend beyond the group meetings. Fifty six respondents stated that they interact with other chat participants over Twitter (follow, mention, retweet) outside the sessions. Forty five interact with others through other online means such as emailing and blogging. In fact, the interactions go beyond the virtual world. Twenty seven respondents stated that they communicate with others in their chat groups through off-line means as well (examples: face-to-face meet-ups, phone calls). Video-based online tools such as Skype, Google Hangout are also popular (eight respondents). The fraction of chat participants that communicate outside chat sessions is much larger compared to other virtual communities discussed in the literature [41]. We give below a few example quotes from our survey that demonstrate the strength of ties:

> I created #1stchat when I was a first grade teacher. It is an amazing group of teachers and I consider them great friends.

> We are more than a group, it's almost family like. People know of vacations, major events, and other things in the lives of those who chat.

> We call participants in Satchat the #satchat family.

On the downside, as a group gets more mature and connected, it can also become closed to new members as demonstrated by the responses of two survey participants:

> ... tends to become very cliquey and the key players *over time* use more and more "insider" references or hold more and more side discussions during the chat.

> I feel as though a *hierarchy has developed* and there are times where people within that hierarchy at times will dismiss other ideas.

These write-ins support the results presented in §5 that show the negative correlation of group maturity with returning to a group.

Thirty nine survey participants stated that they feel a responsibility to the community to participate in chat sessions while seventeen stated that they did not. Drawing on *commitment theory* [32] that was initially introduced to reason about volunteer behavior, we considered the responses of the survey participants to identify behavior indicative of three types of commitment: *affect-based* which refers to individual's emotional attachment to a group, *norm-based* which captures individual's felt sense of obligation and *cost-based* which refers to individual's awareness of the costs associated with leaving an organization. We found *norm-based* and *cost-based* bonds to be prevalent. Examples of *cost-based* bonds can be seen in users that see Twitter chats as a valuable utility and are driven to participate due to its benefits:

> People "lurk and lear" all the time. You go if you want - it's YOUR PD, when and how YOU want it.

Although the open form of Twitter chats allows for "lurking" behavior, there is still a large number of dedicated members that are driven by *norm-based* bonds. These participants view reciprocity as an important notion and feel *obligated* to participate as demonstrated by the following quotes:

> I take others' ideas, so it's only right I respond in kind.

> I know that I appreciate learning from others and sharing ideas, so I think it's my duty to reciprocate.

Unlike related work [6], we have not observed *affect-based* bonds to be prevalent in Twitter chats. This outcome is consistent with the dominance of *informational support* over *emotional support*.

Origins and Evolution: Fifteen respondents discovered their primary education chat group through another Twitter chat. Seventeen became aware of the existence of such groups through general Twitter usage. Nine stated that they created/co-created their chat groups. While research in online social networks focuses on the effectiveness of social networks to spread information [25], our survey revealed that a notable fraction of Twitter users discovered their primary educational Twitter chat through exogenous channels (nine through education related forums/blogs, six through offline connections and three through emails).

The four main reasons listed by the survey respondents as the initial goals/reasons for participating in Twitter chats were: to explore, out of curiosity (27), to learn new information/tools/methods (28), to make new connections (17), and for the sense of belonging (9). These reasons are more information-based rather than social-connection-based unlike related work [41].

Thirty respondents stated that their view of the chat group changed over time; with most people stating that it became easier for them to follow conversations that caused them to change their view. Only ten respondents stated their position and responsibilities had changed over time. Two stated that they had become less active, with one of them ending active participation. Eight respondents had become more integral to groups over time taking on more responsibilities. The users with decreased participation mentioned the unwelcoming environment. One stated:

> I was welcomed and greeted warmly - I went back - it wasn't repeated - but the conversation was worth it, so I lurk and read archives.

This quote reinforces the importance of *perceived receptivity* which was found to be important in our statistical data analysis (§5).

Feedback: We also asked the survey respondents how the quality of conversations can be improved in Twitter chats. There were fourteen responses with suggestions. Most replies highlight problems that can be addressed through technological solutions. Some common themes were: "Once a tweet is retweeted in the chat, protocol should be that no one else retweets it. I find multiple retweets to very frustrating and clutters up the feed.", "All chats should be archived", "A central place to find other chats that are on topic would be helpful."

Some comments go to the nature of the groups: "Allow differing opinions to be more than fodder for side-chats.", "Sometimes chats can use a tone that is condescending to new people. Please don't use terms like "newbies!"", "More educators need to get on the bus and join in on twitter.","A few chats could be better-prepared by organisers: eg greater publicity (reminders in advance), more guiding questions and more resources","If people are looking for professional development on Twitter, they need to be willing to be challenged and respond, rather than run away with hurt feelings that leave them unchanged."

6.2 Discussion

Our survey study marks various distinctions between Twitter chats and other online groups and face-to-face discussions. For one, *informational* support has been found to be more important to Twitter chat users than *emotional support*. Although related work suggests that informational support is negatively correlated with the sense of community [48], the sense of community is very strong in Twitter chats. In fact, its members communicate with one another outside chat sessions much more than expected from the literature [41]. Disadvantages identified by the survey respondents also mark an interesting distinction between Twitter chats and other online groups. While for other online communities, the lack of face-to-face interactions is a main disadvantage, Twitter chat users focus on the content. More specifically, due to the synchronous and open nature of Twitter, the pace of information is the biggest challenge of Twitter chats.

We have also observed that the survey results reinforce most findings presented in Section 5. The importance of social inclusion is observed from the responses of two survey participants that reduced (one ending) their participation due to the lack of receptivity. Groups becoming closed to new members over time (as captured by *groupmaturity* in our model) is seen anecdotally in survey results. The geographical diversity listed as an advantage also indicates that geography is not a limiting factor for Twitter chats.

7. CONCLUSIONS AND FUTURE WORK

What makes a person become a member of a particular group? We addressed this question in the context of Twitter chats which are time-bound synchronous group interactions carried out in real time on a focused topic on Twitter. We developed *5F Model* that predicts whether a person attending her first chat session in a particular Twitter chat group will return to the group. This model considers five different classes of factors: *individual-initiative, group characteristics, perceived receptivity, linguistic affinity* and *geographical proximity*, building upon findings from prior research on asynchronous online groups and communities.

We performed statistical data analysis for thirty educational Twitter chats involving 71411 users and 730944 tweets over a period of two years. Analysis was performed to identify the significance of separate models for each of the five factors listed above as well as the unified *5F Model*. We also complemented the results of statistical analysis with a survey study.

Our results show that users are more likely to return to Twitter chats that mention or retweet them, stressing the importance of social inclusion. Unlike offline groups, we find that for most Twitter chat groups, geographical proximity is not a limiting factor for a user to affiliate with a group. In addition, we show that *informational support* is more important than *emotional support* in educational Twitter chats. Unlike what is suggested by research in other online communities, the sense of community in Twitter chats is very strong despite this finding. Given the synchronous nature and popularity of Twitter chats, we also observed that *information overload* was a significant challenge. Interestingly, our results indicate that the best predictor for future participation is *linguistic affinity*, as opposed to individual or group characteristics that the literature mostly focuses on.

To the best of our knowledge, this is the first work to consider group dynamics questions in Twitter chats. The findings of this study provide various insights for Twitter chat organizers. For instance, creating a welcoming environment and providing ways to alleviate the unpleasant effects of information overload are two paths to long-lasting user participation in Twitter chats.

As future work, we aim to extend our work on educational Twitter chats to chats held on other topics. This would allow us to identify the generalizability of the findings in educational chats and determine characteristics that are unique to them. While we focused on the first interactions of participants with Twitter chat groups, it is also important to identify how individual-group interactions evolve over time. For this purpose, we aim to study the evolution of Twitter chats, identify different types of chat participants and quantify their contributions to the success of Twitter groups over time.

8. ACKNOWLEDGMENTS

We would like to thank James Cook, Krishnaram Kenthapadi, and Nina Mishra for bringing the Twitter chat phenomenon to our attention. We would also like to thank Nina Mishra and Jed Brubaker for the insightful discussions.

9. REFERENCES

[1] J. Arguello, B. S. Butler, E. Joyce, R. Kraut, K. S. Ling, C. Rosé, and X. Wang. Talk to me: foundations for successful individual-group interactions in online communities. In *CHI '06*, pages 959–968. ACM, 2006.

[2] Y. Artzi, P. Pantel, and M. Gamon. Predicting responses to microblog posts. In *HLT-NAACL*, pages 602–606, 2012.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD '06*, pages 44–54. ACM, 2006.

[4] L. Backstrom, J. Kleinberg, L. Lee, and C. Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: Expansion, focus, volume, re-entry. In *Proceedings of WSDM*, 2013.

[5] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *WSDM '08*, pages 117–128. ACM, 2008.

[6] P. J. Bateman, P. H. Gray, and B. S. Butler. Community commitment: How affect, obligation, and necessity drive online behaviors. In *ICIS*, page 63, 2006.

[7] C. Borgs, J. Chayes, M. Mahdian, and A. Saberi. Exploring the community structure of newsgroups. In *KDD '04*, pages 783–787. ACM, 2004.

[8] P. B. Brandtzæg and J. Heim. User loyalty and online communities: why members of online communities are not faithful. In *INTETAIN '08*, pages 11:1–11:10, 2007.

[9] S. Carson and J. P. Schmidt. The massive online professor. *Academic Matters: The Journal of Higher Education*, May 2012.

[10] C. M. L. Chan, M. Bhandar, L.-B. Oh, and H.-C. Chan. Recognition and participation in a virtual community. In *HICSS '04*, pages 70194.2–. IEEE Computer Society, 2004.

[11] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann. What makes conversations interesting? themes, participants and consequences of conversations in online social media. In *WWW'09*, pages 331–331, 2009.

[12] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg. Echoes of power: language effects and power differences in social interaction. In *WWW '12*, pages 699–708. ACM, 2012.

[13] C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, and C. Potts. No country for old members:user lifecycle and linguistic change in online communities. In *WWW'13*, 2013.

[14] Educational chats on twitter. http://www.cybraryman.com/edhashtags.html.

[15] T. Finholt and L. S. Sproull. Electronic groups at work. *Organizational Science*, pages 41–64, 1990.

[16] D. Forsyth. *Group Dynamics*. Belmont, CA: Wadsworth/Cengage, 5th edition, 2010.

[17] A. Forte, M. Humphreys, and T. Park. Grassroots professional development: How teachers use twitter. In *ICWSM*, 2012.

[18] S. Green-Hamann, K. C. Eichhorn, and J. C. Sherblom. An exploration of why people participate in *second life* social support groups. *J. Computer-Mediated Communication*, 16(4):465–491, 2011.

[19] D. Hosmer, S. Lemeshow, and S. May. *Applied survival analysis: regression modeling of time to event data*, volume 618. Wiley-Interscience, 2011.

[20] D. A. Huffaker. The impact of group attributes on communication activity and shared language in online communities. *First Monday*, 16(4), 2011.

[21] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker. Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1):39–44, Jan. 2011.

[22] C. Johnson. A survey of current research on online communities of practice. *The internet and higher education*, 4(1):45–60, 2001.

[23] Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Info. Sys. Research*, 15(2):194–210, 2004.

[24] E. Joyce and R. E. Kraut. Predicting continued participation in newsgroups. *J. Computer-Mediated Communication*, 11(3):723–747, 2006.

[25] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD'03*, pages 137–146, 2003.

[26] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW '10*, pages 591–600, 2010.

[27] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas. Homophily in the digital world: A LiveJournal case study. *IEEE Internet Computing*, 14(2):15–23, 2010.

[28] M. R. Leary. Responses to social exclusion: Social anxiety, jealousy, loneliness, depression, and low self-esteem. *Journal of Social and Clinical Psychology*, 9:221–229, 1990.

[29] Maxmind world cities with population. http://www.maxmind.com/app/worldcities.

[30] R. McCrae and P. Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52:81–90, 1987.

[31] D. R. Mettee, S. E. Taylor, and S. Fisher. The effect of being shunned upon the desire to affiliate. *Psychonomic Science*, 23:429–431, 1971.

[32] J. P. Meyer and N. J. Allen. A three-component conceptualization of organizational commitment. *Human Resource Management Review*, 1(1):61 – 89, 1991.

[33] N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.

[34] T. M. Newcomb. Varieties of interpersonal attraction. *Group dynamics: Research and theory*, pages 104–119, 1960.

[35] B. Nonnecke and J. Preece. Persistence and lurkers in discussion lists: A pilot study. In *HICSS '00*, pages 3031–. IEEE Computer Society, 2000.

[36] J. Pennebaker, R. Booth, and M. Francis. Linguistic inquiry and word count (liwc2007). http://www.liwc.net/.

[37] G. Philipsen. A theory of speech codes. *Gerry Philipsen and Terrance Albrecht, eds., Developing Communication Theories*, 1997.

[38] B. Poblete, R. Garcia, M. Mendoza, and A. Jaimes. Do all birds tweet the same?: characterizing twitter around the world. In *CIKM '11*, pages 1025–1030. ACM, 2011.

[39] S. Predmore and K. Williams. The effects of social ostracism on affiliation. *Midwestern Psychological Association*, 1983.

[40] E. M. Rogers and R. Agarwala-Rogers. Organizational communication. *Communication behaviour*, pages 218–239, 1975.

[41] D. R. Sedo. Readers in reading groups : An online survey of face-to-face and virtual book clubs. *Convergence: The International Journal of Research into New Media Technologies*, 9(1), 2003.

[42] M. Sherif. *The psychology of social norms.* Harper, 1936.

[43] R. W. Sinnott. Virtues of the haversine. *Sky and Telescope*, 68(2):159+, 1984.

[44] J. D. Snoek. Some effects of rejection upon attraction to a group. *Journal of Abnormal and Social Psychology*, 64:175–182, 1962.

[45] B. W. Tuckman. Developmental sequence in small groups. *Psychological bulletin*, 63(6):384–399, 1965.

[46] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the facebook social graph. *CoRR*, 2011.

[47] Y.-C. Wang, R. Kraut, and J. M. Levine. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *CSCW '12*, pages 833–842. ACM, 2012.

[48] J. L. Welbourne, A. L. Blanchard, and M. D. Boughton. Supportive communication, sense of virtual community and health outcomes in online infertility groups. In *CT '09*, pages 31–40. ACM, 2009.

[49] K. Williams, C. Cheung, and W. Choi. Cyberostracism: effects of being ignored over the internet. *J Pers Soc Psychol*, 79(5):748–62, 2000.

[50] B. Wing and J. Baldridge. Simple supervised document geolocation with geodesic grids. In *ACL*, pages 955–964, 2011.

[51] R. C. Ziller. Toward a theory of open and closed groups. *Psychol Bull*, 64:164–82, 1965.