

Understanding the Usage of Online Media for Parenting from Infancy to Preschool At Scale

Yujia Gao
Georgia Institute of Technology
yujia71@gatech.edu
Atlanta, GA, USA

Jinu Jang
Georgia Institute of Technology
jinu.jang@gatech.edu
Atlanta, GA, USA

Diyi Yang
Georgia Institute of Technology
dyang888@gatech.edu
Atlanta, GA, USA

ABSTRACT

New parents, defined as parents of children between the infant and preschooler stages, are increasingly turning to online media to exchange support and information to help with their life-changing transition. Understanding parents' discussion online is crucial to the design and development of technologies that can better support their media interaction. This work studies how new parents use online media using a large-scale parenting corpus. To do so, we first employed a card-sorting methodology to identify a set of parenting topics, with which we trained BERT classifiers to automatically identify the topics of Reddit posts. We then investigate at scale what parenting topics were talked about most by new parents, how topics change over the course of their participation, and how interactions with different topics affect members' engagement in the community. We conclude with implications of our research in designing future research and online parenting communities.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing.**

KEYWORDS

Social Support; Parenting; Online Media; Online Health Communities; Natural Language Processing; Data Augmentation

ACM Reference Format:

Yujia Gao, Jinu Jang, and Diyi Yang. 2021. Understanding the Usage of Online Media for Parenting from Infancy to Preschool At Scale. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3411764.3445203>

1 INTRODUCTION

As one of the most difficult and important jobs, parenting is a complex societal skill. Many new parents often find themselves confused about how to do well in their new roles, especially from the infancy to the preschooler stage, where most questions and difficult choices about parenting occur. [10] New parents go through crucial transitions to become mothers and fathers, and various

challenges from different aspects can be demanding to adapt to. Prior studies concluded that new parents are more prone to feeling loneliness and social isolation from being withdrawn from their social circles, and depletion from daily mundane tasks of taking care of the baby, especially when the child is totally dependent on the parents after born [9, 32]. Edhborg et al. [22] confirmed both new mothers and fathers are subject to heightened risk of depression. According to a meta-analysis report by the Centers for Disease Control and Prevention, 13 percent of new mothers result in postpartum depressions [39].

As a result, many new parents take advantage of technology and social networks to exchange information, resources, and social support, especially through online media platforms [19]. Online parenting support groups provide a supportive and nonjudgmental place for new parents to talk about their challenges, emotions, and experiences. One recent report showed that 75% of parents who use online media use it for parenting-related information and social support [19]. Prior research shows that such online parenting support groups help parents develop parenting skills and knowledge, as well as improve their ability to deal with childhood disability or mental health issues [2, 31]. Thus, understanding parents' discussion online is crucial to the design and development of technologies that provide better support to new parents.

Motivated by the above, our research further studies how parents use online media for parenting, especially towards an understudied subgroup of parents: parents of infants and preschoolers - a stage of parenting that is not only relatively challenging (e.g., "Terrible Twos"), but also central to the development of behaviors in young children [27]. Similar lines of research exist, with a non-exhaustive list being cross-cultural families [52], dads [4, 6], moms [6, 28, 34, 42], parents of special needs children [3], and LGBT parents [11]. Building upon these prior research, we take a closer look at a more general parenting group specific to a key stage in child development that all parents have to pass through [38, 47]. Compared to previous studies on generic parenting topics [4, 42], our research offers a finer-grained modeling of parenting discussion related to parenting skills and experience that people frequently discuss online. This work aims at addressing the following three research questions:

- (1) What do parents of infants and preschoolers talk about in online parenting discussion groups?
- (2) How do group members participate in these parenting discussions of various topics?
- (3) What factors affect members' interaction with different parenting topics and their engagement in the community?

To answer these questions, we collect and use the largest dataset in this field thus far: a total of over 600K posts and 10 million comments associated with these posts, originated from 46 subreddits

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8096-6/21/05...\$15.00
<https://doi.org/10.1145/3411764.3445203>

associated with parents of infants to preschoolers on the social aggregation site Reddit.com, written by 600K individual users. Then to capture what topics parents discuss online, we employ a qualitative approach of card sorting [45] to categorize these parenting discussions into a set of human-decided topics, which improves interpretability and coherence of the resulting topics compared to ones inferred from unsupervised topic modeling methods. In order to process millions of textual messages on parenting discussion, we use a human-annotated corpus to build state-of-the-art machine learning classifiers for predicting whether a post is about parenting and, if so, which topics it contains. We then apply it to automatically label the remaining posts with their respective topics. The remainder of the paper takes a deep dive into our Research Questions (RQ): We further investigate RQ1 by visualizing which topics were talked about among parents of infants and preschoolers in different subreddits, examine RQ2 by comparing the discussion topics at different stages of the parents' participation and how community members interact in these parenting discussions, and answer RQ3 by exploring what factors affect group members' engagement towards these parenting related discussion. In the end, we demonstrate the implications of our comprehensive study of how parents of infants and preschoolers use online media for parenting.

2 RELATED WORK

2.1 Parental Social Support and Online Media

Parents, especially new parents, have been reported to be notably prone to feelings of loneliness, stress, and depression [9, 22, 32, 39]. Social support was key to relieving this parental stress related to transitioning to parenthood, according to Gage and Christensen [24]. Numerous studies have documented the relationship between parental social support and parental perceptions of efficacy, parental stress, quality of family well being, etc. [7, 35]. For instance, Armstrong et al. [7] developed a conceptual framework for understanding the relationship between parental social support, family well being, parenting capacity and child resilience, in which parental social support serves as the mediator between environmental and social stressors for parents and family. Gage and Christensen [24] concluded that socialization is conducive to relieving parental stress related to transitioning to parenthood.

With online media becoming vital channels for people's daily interactions [25, 33], parents more frequently turn to online media for parenting related information and support [19, 37, 38, 47]. Previous research showed that parents use online technology, especially social media platforms, for the following three purposes: specific information acquisition, emotional support exchange, or connection-building with other parents [18]. With 64.5% of parents using parenting websites as a source for parenting information and 45.1% using social media [8], along with 75% of parents responding that they use social media to seek social support [19], online use for parenting is becoming increasingly important. Jang and Dworkin [28] also concluded that a mother's frequent use of social media is positively related to her bonding social capital. Through surveys and interviews, Madge and O'Connor [34] studied how young moms were turning to the worldwide web for "esteem support" and "knowledge support". The authors also found that the users of their research website felt more empowered and in control,

as they did not have to wait for appointments when going online, nor did they feel that their question was being judged.

Research also concluded that support for partial or complete anonymity on online parenting communities like YouBeMom.com or Reddit.com allow parents to self-disclose personal experiences without the fear of being judged [29]. Ammari et al. [5] looked at the effects of using a throwaway account on Reddit.com, enabling parents to feel more comfortable discussing stigmatized topics. Using randomized trial experiments, they also concluded that content posted by throwaway accounts were more likely to receive responses and social support. Based on these previous results, we chose to collect our data from Reddit.com, a site that offers anonymity and a space in which users would be more likely to be open and honest in their discussions.

2.2 Parental Use of Online Media

To gain a better understanding of how parents make use of and benefit from online parenting media, many prior works investigate specific parenting groups' use of online parenting communities. Morris [38] looked at how mothers of young children (aged three and under) use social networking sites, particularly Facebook and Twitter, including mothers' perceptions of which social networks are appropriate for sharing information about their children. Ammari and Schoenebeck [4] revealed that fathers also have motive to visit online communities for information, skills, and support: many would read through other dad's parenting philosophies to aggregate and define their own. Recently, a growing number of parenting research studies has also been conducted to understand how parents from diverse groups such as cross-cultural parents, LGBT parents, and parents with special needs kids use technology to find information and support related to parenting [3, 11, 43]. Through interviews with 18 cross-cultural families, Yarosh et al. [52] concluded that most cross-cultural families tended not to post personal experiences publicly for fear of over-sharing, conflicts, or bullies online. Conversely, many cross-cultural families would use the Internet to find and connect with hybrid in-person/online sources of social support. As minority groups also use social media to fight for their rights, Blackwell et al. [11] found that LGBT parents became "incidental advocates", where even posts not intended for advocacy were perceived as such by the public. While these works contribute a variety of perspectives in understanding parents use of online media, little work focuses on generic understanding of online parenting in general. In order to achieve this goal, our work looks at parents of infants and preschoolers, which still remains under-investigated compared to extensive studies on either generic parenting research or online health communities.

2.3 Parenting Topics

One main focus of our work is to investigate which topics, related to parenting skills and experiences, parents discuss online. There has been a similar rise in studies that, rather than relying solely on traditional qualitative analyses conducted via interviews or surveys, investigate parents' discussions on online media using quantitative analyses. For example, via most frequent word usage analyses on an anonymous message board for parents called YouBeMom.com, Schoenebeck [42] investigated which topics moms talk about when

they are not constrained by the norms and expectations of face-to-face culture. Ammari and Schoenebeck [4] used Latent Dirichlet Allocation (LDA) to compare topic and platform usage differences between different gender identities and parenting roles. They concluded mothers and fathers share certain topics in their discussion, while differences remain, one of which is mothers tend to focus more on early childcare discussions.

We build off from these prior studies that create gendered representation of parenting topics and provide a more comprehensive and generic taxonomy on what people talk about on parenting platforms. We incorporate qualitative and quantitative methodologies to produce coherent and human-understandable results that can be applied to online parenting communities in general. Moreover, we also make use of the taxonomy and our large-scale corpus to visualize how parents and community members from those online parenting support groups engage in parenting related discussions and what factors facilitate such discussion.

3 RESEARCH SITE ON PARENTING

Reddit.com is a social news aggregation and discussion website where users can submit content into user created forums of their choice, called “subreddits”. In order to retrieve parenting related discussions on Reddit, we first used an external Reddit API¹ to collect all parenting related subreddits by querying it with topic keywords “parenting”, “parent”, and “new parent”. Based on the returned subreddit list, we then manually examined each subreddit to filter out subreddits that are irrelevant to parenting before preschool. We finally obtained 46 parenting subreddits related to our target group after the manual annotation, and the complete list of all scraped subreddits are provided here:

r/Adoption, r/AskParents, r/AtheistParents
 r/BabyBumps, r/BabyExchange
 r/AttachmentParenting, r/beyondthebump
 r/birthcontrol, r/boobsandbottles
 r/breastfeeding, r/BreastPumps, r/Buyingforbaby
 r/ClothDiaps, r/CuteKids, r/daddit
 r/downsyndrome, r/FamilyDinner, r/FormulaFeeders
 r/homeschool, r/Midwives, r/Mommit,
 r/NaturalPregnancy, r/NewParents, r/NewDads
 r/NewMomStuff, r/parent, r/Parents
 r/parentdeals, r/Parenting, r/ParentingLite
 r/ParentsOfMultiples, r/playgroup, r/PreDaddit,
 r/pottytraining, r/radicalparenting
 r/Preterm r/raisingkids, r/SAHP, r/SingleParents
 r/SpecialNeedsChildren, r/stepparents
 r/teachingchildren, r/ThingsMyKidSaid, r/toddlers
 r/TryingForABaby, r/UKParenting, r/VBAC

We then scraped the thread-starting posts and their comments from pushshift.io² for these 46 parenting related subreddits between 2015 - 2020, which then served as the research corpus for all of the following analyses. During this period, 602,078 posts were collected from 202,988 distinct users, along with 10,794,023 comments from 425,059 users related to these posts. The mean length of the posts was 185.3 words, while the median was 128. The mean length of the

comments was 56.3 words, while the median was 36.0. Amongst the 202k Redditors, users started a mean of 2.96 and a median of 1.0 post threads and wrote a mean of 24.7 and a median of 3.0 comments. These are the data for which we base our following analysis on.

4 UNDERSTANDING PARENTING TOPICS

To understand what parents discuss when participating in online parenting platforms, we conducted a thorough examination of common parenting topics via card sorting [45]. Card sorting is a technique in user experience design for discovering how a group of subject experts or users categorize information or generate a taxonomy. We first randomly sampled 1000 thread-starting posts from the aforementioned parenting related subreddits, then asked annotators from Amazon Mechanical Turk (MTurk)³, an online crowdsourcing platform that allows people to post jobs and for workers to choose jobs to perform, to provide a set of keywords to describe the topics or themes in each post. We provided the turkers with detailed description and examples of this task. Annotators were asked to rate whether a post is about parenting from infancy to preschool; if yes, they then provided a few keywords (a single word or a phrase) to summarize the topics in an open-ended manner. To ensure annotation quality, we required turkers to have a 98% approval rate and at least 5000 approved tasks, or “HITS”, for their previous work on MTurk. Furthermore, turkers received training materials and were required to pass the free-form keyword qualification tests after the training in order to participate in the task. Each post was annotated by three turkers and we paid \$0.08 for rating each message. Once we collected those rated posts with their candidate topical keywords, we then organized two card-sorting sessions where we worked closely with three graduate research assistants majoring in HCI (not authors of this work) to design the parenting topic taxonomy. In the first session, each annotator was asked to go through his/her received list of keywords independently and come up with a candidate taxonomy to categorize those keywords into clusters. Annotators would also provide a name for each cluster as the candidate parenting topic. In the second session, based on the three candidate taxonomies from three annotators, the authors of this paper identified overlapped topics and resolved possible disagreement among annotators about the categories. This led to a finalized list of 30 parenting topics in Table 1, which annotators agreed could serve as a representative taxonomy of topics in parenting discussion.

Some of our finalized 30 topics covering what parents discuss online are consistent with parenting taxonomies and interview excerpts from prior research [6, 38, 47]. However, the difference is that our topic taxonomy captures more nuanced parenting topics, such as Moms’ Health, Language Training, and Walk Training. These topics are subtly nuanced, such as Child Behavior Issue versus Child Health, which are often hard to be inferred from a simple keyword perspective, nor be directly reflected by LDA based techniques that often require both manual interpretation of latent topics and extensive choices of topic numbers to determine the optimal one. We attribute these benefits to our qualitative card sorting methodology for constructing the parenting topic taxonomy. Compared to topic modeling or other automatic inferences methods,

¹<https://snoopsnoo.com/>

²<https://pushshift.io>

³<https://www.mturk.com/mturk/welcome>

Topic	Definition	ICC	#
Baby food	Soft, easily consumed food for babies, other than breast milk or formula.	0.469	114
Birth	Talk about childbirth, birth choices or the delivery of a new baby.	0.438	127
Breastfeeding	Breastfeeding or nursing babies or young children.	0.743	502
Child Behavior Issue	Discuss child behavioral problems and solutions.	0.240	142
Child Health	Kids' health issues, such as symptoms, vaccination.	0.608	265
Child Appearances	Talk about kids'/babies' appearances, weights, etc.	0.733	60
Child Milestones	Describe kids' milestones, such as walk, crawl, move.	0.251	31
Child Product	Products that are designed for babies and kids such as toys and clothes.	0.766	556
Childcare	Looking after children by a day-care center, nannies, babysitter, teachers or other providers.	0.044	90
Dad Parenting	Dad's involvement and issues about parenting children.	0.561	159
Discipline	Setting rules and disciplines for children	0.167	83
Economy	Economical factors or aspects of parenting	0.452	71
Emotion Coaching	Help children recognize and understand their feelings.	0.498	40
Language Training	Teach children to learn to speak and talk.	0.347	18
Moms' Health	Health, mental health issues related to moms	0.490	93
Multiple Children	Having multiple children	0.414	136
Non-biological Parents	Issues with step parents, LGBT parents, adoption, etc.	0.562	74
Parent Education	Educating and training parents for better parenting.	0.142	69
Parent Support	Support and social activities for parents.	0.281	244
Parent-child Interaction	Interaction and relationship (e.g., attachment, bond, separation) with kids.	0.497	145
Play	Children's play, toys and activities.	0.188	149
Potty Training	Toilet or potty training.	0.725	151
Pregnancy	Talking about pregnancy and pre-birth issues.	0.510	328
School	Choosing preschool, kindergarten, education, training.	0.469	107
Sell/Exchange	Discussion of selling, exchanging or swapping things.	0.647	156
Single Parents	Single parent's or divorced parents' situation and parenting strategies.	0.749	148
Sleeping	Sleeping issues that parents and babies might have, such as baby naps, nighttime training.	0.752	392
Walk Training	Training babies to learn to walk.	0.508	29
Working Parents	Parents' work-life balance and relevant issues.	0.577	83
Other	Parenting related topics that are not listed above.	0.348	118
Parenting		0.624	3939

Table 1: A taxonomy of 30 parenting topics, their definitions, the Intra-class Correlation (ICC) for measuring annotation agreement, and the number of posts in each parenting topic (#).

our process of qualitative keyword sorting encodes how annotators sort these collected keywords and categorized these unstructured information into comprehensive and exclusive parenting topics.

5 PREDICTING PARENTING TOPICS

Given that users can discuss various topics even within specific subreddits, it is not reasonable to assume all posts in a subreddit belong to that subreddit subject. For instance, parents typically join “r/breastfeeding” to exchange information on breastfeeding related issues, but may also post funny stories that involve their pets and kids. Thus, identifying a high quality set of posts where users engage in parenting discussion is essential for understanding what topics parents are talking about, and how community members react to those discussions. To this end, we built machine learning models to automatically predict different parenting topics that parents discuss. First, human coders rated a set of messages about whether a post is about parenting and its involved topics based on our parenting topic taxonomy in Table 1. These annotations were then served as the ground truth in developing the automated measure of parenting topics. Second, we represented each message as a

set of language features as input for the machine learning models and constructed statistical models based on the hand-coded data and evaluated its performance. Finally, we applied the validated model to all messages in all parenting related subreddits.

5.1 Dataset Annotation

In order to construct an annotated “ground-truth” dataset for training machine learning models, we randomly selected a sample of 4,800 Reddit posts for manual labeling on MTurk. We asked turkers to first judge whether a post was about parenting between the infancy and preschooler stages via a simple yes/no question; if yes, they were then asked to select any number of topics they thought classified that post from the list of 30 topics in Table 1. We provided definitions and examples to each parenting topic, as well as a set of test questions as qualifications. We again required annotators to have a 98% approval rate and at least 5000 approved HITs for their previous work as well as passing the parenting topic qualification tests but also required them to be located in the US. We paid \$0.10 for rating each message, and each post was rated by 3 different annotators. The average annotation time per post is 219 seconds.

To validate the reliability of the turkers’ annotations, we used Intra-class Correlation (ICC) to calculate the annotation agreement, as shown in Table 1. We found that, in terms of whether a post is about parenting from infancy to preschool, we obtained an ICC score of 0.624, indicating moderate agreement among three raters. Most parenting topics, such as Breastfeeding, Child Health, and Potty Training, had fair ICC scores (greater than 0.4) [16]. We also observed that some parenting topics, including Child Behavior Problem, Childcare, Discipline, Parent Education, Parent Support, Play, and Other, had an ICC score less than 0.4, indicating that these topics did not have consistent annotations. We then discarded those 7 categories from our follow-up analyses, and focused on the remaining topics with higher quality annotations. For all annotated posts, we finalized a label for a post if at least 2 out of the 3 annotators agreed on that label via a majority voting. After this process, there were 3,939 Reddit posts about parenting, and the detailed breakdown of their specific topics is in Table 1.

5.2 Predicting Parenting Topics

To automatically predict the topics of each post, we converted this parenting topic classification task into several binary classification tasks: one for whether a specific post is talking about parenting from infancy to preschool, and one each for whether it is about a certain topic. To build robust and reliable machine learning classifiers, we removed topics that have less than 50 data points or have an ICC annotation agreement score less than 0.4. This led to a total of 19 classification tasks, together with the prediction of Parenting discussion. We experimented with two machine learning models that are widely used in text classification tasks to automatically predict these parenting topics:

- (1) Logistic Regression Classifier: we built generic logistic regression models with bag-of-words representation, Word2Vec representation [36], and a combination of TF-IDF and Word2Vec.
- (2) Bidirectional Encoder Representations from Transformers (BERT) [17]: BERT is a pre-trained language representation model, and has achieved state-of-the-art results on eleven natural language processing tasks, outperforming baseline methods. We used the Pytorch implementation of BERT⁴.

The maximum length of the token sequence was 512, so we truncated sentences (around 2%) that exceeded this limit. Although our annotated corpus had approximately 4000 posts with their corresponding parenting topics, it was still not enough, especially for some of the less common topics. To mitigate this problem, we applied a widely-used data augmentation method called back-translation to create a set of “augmented data and labels” [14, 15, 23, 44]. It works by translating the posts from one language (e.g. English) to another (e.g. German), and back again to English. We used the Fairseq⁵ translation model and effectively doubled the number of posts for each topic via data augmentation.

To construct the machine learning models, the annotated dataset was randomly partitioned into training (80%), validation (10%), and testing (10%) sets. The training set was used to build the models, while the validation set was used as performance evaluation for different hyper-parameter configurations. We evaluated the final

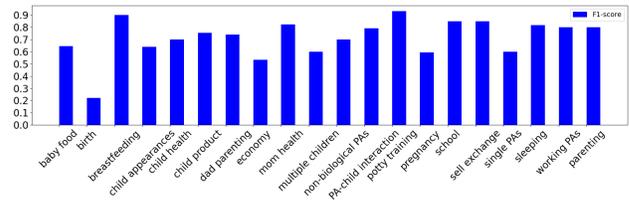


Figure 1: F1-Score from BERT Classifier + Augmented Data for predicting different parenting topics

performance using the testing set once the model achieved good performance on the validation set. Note that we only performed the data augmentation on the training set and did not use it for testing nor validation.

5.3 Result Discussion

We trained the binary classifier for all 19 remaining parenting topics and calculated their average performance with each combination of the features and models in Table 2. We found that, compared to traditional Logistic Regression models, BERT based models achieved higher performances in terms of both Accuracy and F1 scores. Additionally, adding augmented data to the BERT Classifier significantly improved F1 scores from 0.612 to 0.710. We thus chose this BERT Classifier with augmented data as our final model, and reported the detailed results on each parenting topic in Figure 1. We found that our machine learning classifier obtained reasonable accuracy for predicting different parenting topics consistently. Given the adequate validity of these models, we then applied them to the remaining 602K unlabeled Reddit posts. We first apply the parenting relatedness classifier to identify posts that talked about parenting from the infancy to the preschoolers, and then perform our parenting topic classifiers on these posts identified as parenting to further predict a list of parenting topics that they belong to.

6 VISUALIZING PARENTING TOPICS

This section utilizes a large-scale corpus to study RQ1 and RQ2 by visualizing which parenting topics related to infants and preschoolers were discussed in different subreddits at scale and investigating how members in those online parenting support groups participate in parenting discussions of various topics.

6.1 Parenting Topic Discussion Differences

Using the corpus labeled by our machine learning models, we aggregated topics from all posts in a subreddit to obtain the most frequently discussed parenting topics of each subreddit. When analyzing these most discussed parenting topics for each subreddit, we found that the same parenting topics tended to show up in multiple subreddits. All topics in our list of 19 were represented as one of the top 3 topics for a subreddit in at least a single subreddit. The Chi-Squared goodness of fit test was conducted to investigate whether the parenting topics per subreddit followed a uniform distribution. We found that all subreddits showed significant deviation from the null hypothesis, for instance, for “r/ParentingLite”, which showed the least deviation, the chi-squared statistic is $\chi^2(18) = 93, p <$

⁴<https://github.com/huggingface/transformers>

⁵<https://github.com/pytorch/fairseq>

Model	Accuracy	Recall	Precision	F1
Logistic Regression + Word2Vec Features	0.970	0.534	0.534	0.633
Logistic Regression + (Word2Vec & TF IDF Features)	0.966	0.393	0.906	0.518
Logistic Regression + Bag of Words Features	0.949	0.179	0.474	0.238
BERT Classifier	0.972	0.565	0.692	0.612
BERT Classifier + Augmented Data	0.977	0.650	0.843	0.710

Table 2: Models' average performances for predicting different parenting topics

Topics	Top 3 Subreddits	Topics	Top 3 Subreddits
baby food	r/boobsandbottles (5.86%) r/FormulaFeeders (5.86%) r/beyondthebump (1.45%)	non-biological parents	r/BabyExchange (17.2%) r/Adoption (14.0%), r/cutekids (6.5%)
birth	r/Preterms (4.1%), r/downsyndrome (4.0%), r/SpecialNeedsChildren (3.9%)	parent-child interaction	r/thingsmykidsaid (22.2%), r/cutekids (20.4%), r/FamilyDinner (16.1%)
breastfeeding	r/BreastPumps (31.5%), r/boobsandbottles (23.5%), r/breastfeeding (23.1%)	potty training	r/pottytraining (25.9%), r/raisingkids (16.6%), r/toddlers (2.36%)
child appearances	r/AttachmentParenting (23.5%) r/NewParents (9.6%) r/beyondthebump (9.6%)	pregnancy	r/NaturalPregnancy (23.9%) r/NewDads (21.4%)
child health	r/UKParenting (8.6%) r/SpecialNeedsChildren (7.2%), r/Parenting (4.0%)	school	r/homeschool (46.6%) r/parentdeals (25.0%), r/playgroup (16.7%)
child product	r/Breastpumps (46.7%) r/breastfeeding (43.9%), r/clothdiaps (36.3%)	sell exchange	r/playgroup (66.7%), r/raisingkids (14.1%) r/BabyExchanges (52.3%)
dad parenting	r/FamilyDinner (22.2%) r/Buyingforbaby (20.4%), r/clothdiaps (16.1%)	single parents	r/pottytraining (34.5%) r/SingleParents (11.8%) r/stepparents (5.5%)
economy	r/SAHP (3.4%), r/NewDads (2.6%) r/SingleParents (2.5%)	sleeping	r/vbac (50.9%), r/Midwives (41.2%) r/NaturalPregnancy (47.8%)
mom health	r/NaturalPregnancy (2.1%) r/BabyBumps (0.47%), r/birthcontrol (0.34%)	working parents	r/SingleParents (4.37%) r/SAHP (1.61%), r/NewDads (1.34%)
multiple children	r/birthcontrol (33.7%) r/parentsofmultiples (16.7%) r/vbac (10.9%)		

Table 3: Top 3 subreddits discussing each parenting topic, together with the percentage of the subreddit posts on a given topic.

0.001. This suggests that subreddits do not discuss a wide variety of topics, but instead, focus on a few topics.

Table 3 presents the three subreddits where each topic is most likely to be discussed. We observed that, in general, the subreddit name accurately describes the corresponding topics. For instance, the topic breastfeeding aggregates in “r/BreastPumps” and “r/breastfeeding”, and the top subreddit for potty training is its same-name subreddit. However, we also found some exceptions. Unexpected topics showing up in unexpected subreddits could be attributed to that the topic composition of different subreddits are restrained to our target groups, i.e. new parents of infancy to preschool, and does not comprehensively represent the entire parenting communities on Reddit. For instance, Dad parenting shows up most frequently in “r/FamilyDinner”, “r/Buyingforbaby”, and “r/clothdiaps”. These topic subreddit lists may or may not suggest that dads are more focused on the materialistic, but it is certain

that the reason why r/daddit is not among the most frequent subreddits is because only a small part of the subreddit is about new parents. Thus, the results need to be interpreted in the context of new parents. In another example, the topic sleeping is frequently found in Midwives, Natural Pregnancy, and VBAC, which might communicate to us that new moms going through pregnancy stage frequently experience sleeping difficulties. An example post from r/VBAC (Vaginal Birth After C-section) is “I experienced contractions very soon. I felt a lot of pain in the evening and we thought I was going to go into labour properly. Around 1am the contractions faded away. The Foley bulb had probably fallen out but it was still in my vagina. I slept for a while but woke up several times...”

To sum up, each subreddit seems to focus on a few selected topics, and the subreddit name in general reflects the topics well, suggesting that parents can follow specific subreddits for targeted parenting topic discussion.

6.2 Parenting Topics and Gender Differences

Next, we examine gender roles in parenting related discussion, especially in terms of what topics parents discuss and how they vary across mother-focused vs. father-focused parenting subreddits. The comprehensive nature of our parenting discussion corpus across 46 subreddits enables us to study fine-grained differences and similarities in users' parenting topics of interest at scale.

We analyzed father-focused subreddits against mother-focused subreddits on their respective topics of interest by selecting three father-focused subreddits: "r/daddit", "r/NewDads", and "r/preDaddit", and two mother-focused subreddits: "r/Mommit" and "r/NewMomStuff". Threads from these subreddits were then used to compare the topic distribution among father- and mother-focused subreddits. The independent z-test of proportions for each topic revealed that six topics, birth, mom health, school, sell exchange, single parents, and working parents, are discussed with similar frequency across both groups. Topics including baby food, breastfeeding, child appearance, child health, child product, multiple children, potty training and sleeping are discussed significantly more frequently in mother-focused groups compared to dad-focused channels, with associated p-value less than 0.001. This aligns with Ammari et al's conclusion that mothers tend to discuss more on breastfeeding, preparing food for younger babies, training kids' with sleeping or using toilets [6]. Additionally, one interesting discovery that we have made is that the topic working parents are discussed by the same frequency from mothers and fathers. This results is backed up by how mothers with children of any age have been showing a consistent increase in the participating in the workforce ever since 1975, with the curve flattening out at around the year 2000, where 40% of American households had a mother who was the main breadwinner for the family [40]. As of 2019, 76.8% of mothers with own children between the ages of 6 to 17 were in the workforce, and for mothers with children under 3, 63.5% were in the workforce[12], suggesting that nowadays new mothers will not be limited by the role of "mother" and are actively going back to the workforce and taking more responsibility outside the household while pursuing their personal career.

In comparison, aside from the obvious dad parenting, topics including economy, non-biological parents, parent-child interaction, and pregnancy are discussed more frequently in father-focused parenting groups ($p < 0.001$). One surprising finding from above is that fathers discuss pregnancy more than mothers. Examining pregnancy-related threads from both genders, we found that both fathers and mothers announce the big news about having a baby and express concerns or anxieties towards the pregnancy. While mothers tend to talk about their physical distress and pains from expecting and delivering, or their worry about not failing other's expectations to be the "perfect mother", fathers are more likely to discuss apprehension regarding the need to take more responsibilities, or the plight of financially supporting the family after a new member arrives, which echoes the result that topic economy is discussed more by fathers. These results were also were closely in-line with previous research done on the topics new fathers talk about on Reddit done using a different analysis method [46].

To conclude, certain topics are shared across father- and mother-focused parenting discussion groups. Similar discussion frequency

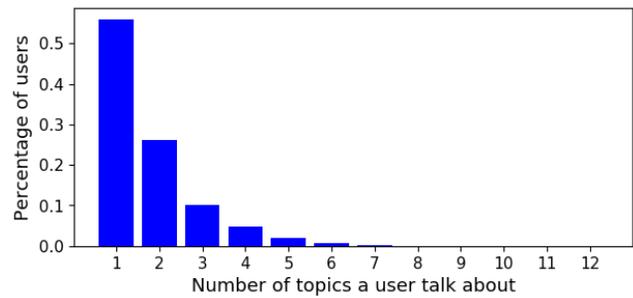


Figure 2: The distribution of the number of topics discussed by each user.

on the topic working parents suggest mothers are increasingly return to their careers as well as fathers, to share responsibilities in support their family as well as fulfilling personal pursuits rather than just being a "mother". In the meantime, mother discuss specific ways to take care of the baby like potty training or baby food, while fathers discuss more on economical or financial situations associated with the baby, possibly due to the social stereotypes which automatically impose more financial burden on the fathers [30]. Our research confirmed other prior works on the topics of interests that new mothers have, while for new dads, our research would be a valuable data point as studies focused on new dads are not as common as ones on new mothers.

6.3 User-level Dynamics on How Parenting Discussion Changes Over Time

We further visualized the number of topics discussed per person. While Section 6.1 explained whether subreddits focus on particular topics, this analysis enables us to understand whether our target parents use online media to discuss only a few selected topics, a variety of topics, or if it differs case by case [48]. We used the Gini coefficient, a percentage ranging from 0 to 1 representing degree of inequality in a distribution, to measure inequality in the parenting topic count distribution for each user. If the Gini coefficient is closer to 1, it indicates that users' topic discussion is not uniformly distributed but rather pertains to certain behaviors. Upon calculation, we obtained a Gini coefficient of 0.76, suggesting that parents' discussed topics can be grouped to certain clusters, further validated by Figure 2. 56% of the users discussed only 1 topic, 91% of the users only focused on 3 or less topics, while the maximum topic number is 12, by exactly 1 user. Interestingly, out of the 22,692 users who discuss two topics on Reddit, one thirds (32%) of them talked about the combination of sleeping and pregnancy, which suggests sleeping problems often co-occur with pregnancy for new parents. Another common pair of topics is single parents and economy, 5% of total two-topic users.

Next, we visualized and compared the topic distributions of new parents over their different participation stages. We hypothesized that parents may join online parenting support groups during a specific milestone or key development stage of child-raising, such as during pregnancy or after the birth of the child. In this case, the user's interest towards these initial topics may change as they

progress through parenting stages, and there might be an increase of interest in other topics. Specifically, we looked at four separate participation stages: up to the first month since the user initially joined the parenting support groups, one to six months, six months to one year, and after one year. The visualizations of the frequency change can be found in Figure 3. We found that, sleeping and pregnancy, two most frequently discussed topics up to the first month of initial participation experience consecutive decrease throughout the four participation stages, dropping by 14.1 and 7.4 percent respectively. As mentioned previously, sleeping is relevant to pregnancy as it consists of discussions on sleeping difficulties experienced during pregnancy stage. This supports our hypothesis that parents tend to join during important turning part, suggesting many parents would join the community during pregnancy stage to learn about parenthood. The trends for other topics also back up this idea: mom health topic frequency also drop steadily, whereas topic frequencies for baby food, child health and child appearances consistently increase over time. This shows a gradual shifting of attention from the mother to the baby after child birth. Additionally, the topic working parents starts with a lower discussion frequency, but increases and reaches peak during the six months to one year since initial participation, which might be due to parents' tendency to discuss this topic and seek balances between work and family after the end of parental leave.

Based on the above observations, we found that many parents join the parents communities during pregnancy stages; new parents started by discussions on sharing news and stories on pregnancy and address on parent-focused issues like physical disturbance brought by pregnancy. Later, the focus of discussion gradually shifts towards topics that relate to taking care of the new-born and working parents situations.

7 ENGAGEMENT IN PARENTING

This section investigates what factors affect members' interaction with different parenting topics and their engagement. Specifically, we use survival analysis to assess the influence of members' participation in different parenting topics and Reddit user activities on user commitment to their fellow communities.

7.1 Methodology

Survival analysis is a statistical technique that estimates the influences of time on an event of interest by defining the dependent variable as the duration of participation. We define user commitment as the duration of participation in parenting related subreddits in units of participant-month [50, 51]. The control variables are the number of comments and posts the user writes, the average number of responses to their posts, the average length of each response. The independent variables are binary representations of the participation of each of these parenting topics. Continuous variables were standardized with a mean of zero and standard deviation of one, while binary variables were left in their original scale so that zero indicated an absence of some characteristic and one meant its presence. We use the statistical software package Stata to execute the analyses and assumed Weibull distribution [49].

We look at 105,497 parents of infants to preschoolers who started at least one parenting related thread on Reddit. The timestamp of

their first activity (e.g., starting a thread, or making a comment on others' thread) is viewed as their starting date for participating in parenting communities on Reddit, and the date of the last activity as the end of the participation. If the last activity was made in the month of the data collection, we consider the end of the participation to be censored, as we do not know the exact date of dropping out of the community. A user is defined to be dropping out of the community if they have not written any posts or comments within three months of their last post/comment. Under this criterion, users can leave the community and re-join multiple times.

7.2 Results

We describe the results of the survival analysis in Table 4. Model 1 predicts whether survival rates are influenced by users' activities on the platform, defined by how long they have been staying in Reddit (tenure), how many posts or comments were sent out, and how were the responses to their posts. Since the continuous variables are standardized, the Hazard Ratio (HR) represents the predicted change in probability of leaving the community per one standard deviation increase in the predictor. An HR greater than one means the variable is associated with a higher than average likelihood of dropping out, while a hazard ration less than one means a lower than average likelihood of dropping out. The HR for tenure indicates that members who stayed in the Reddit one standard deviation longer than average were 27% more likely to stay in the community; similarly, initializing one standard deviation more threads than average was associated with 21% more likely to engage with the community. The HR for members' comments suggest that survival rates were a staggering 50% higher for those who wrote one standard deviation more comments. In contrast, the number of comments members received was associated with a 12% lower survival rate, and the average length of these responses has little effect in survival rates. This suggests that giving support or interacting with others through the means of commenting positively correlated with members' commitment to their fellow communities, consistent with prior research that showed casual relationships between support provision and self-esteem [41], self-worth [26], and increased happiness [1, 20, 21]. This might communicate to us that parents may come to receive advice, but stay to give advice and reap the mental health benefits of doing so.

Model 2 summarizes the influences of different parenting topics on the survival rates. We found that, topics such as breastfeeding, mom health, child product, baby food, and sleeping are associated with higher survival rates. In contrast, economy, school, non-biological parents, and single parents significantly increased the user's probability of dropping out. The increased dropout probabilities associated with these parenting topics might be closely related to what prior work on "throwaway accounts" on Reddit suggested [5], as members in parenting communities are more likely use these throwaway accounts to discuss sensitive topics. Certain topics revealed by prior work [5] such as Financial problems, Work-parenting demands, which echo with our topic of economy, and Gender & Parenting expectations, which involves being a step-parent, or a gay step-parent, and largely overlap with our topic of non-biological Parents. It might be that a significant amount of users on the topics economy, non-biological parents, and single parents discuss sensitive topics and utilize throwaway

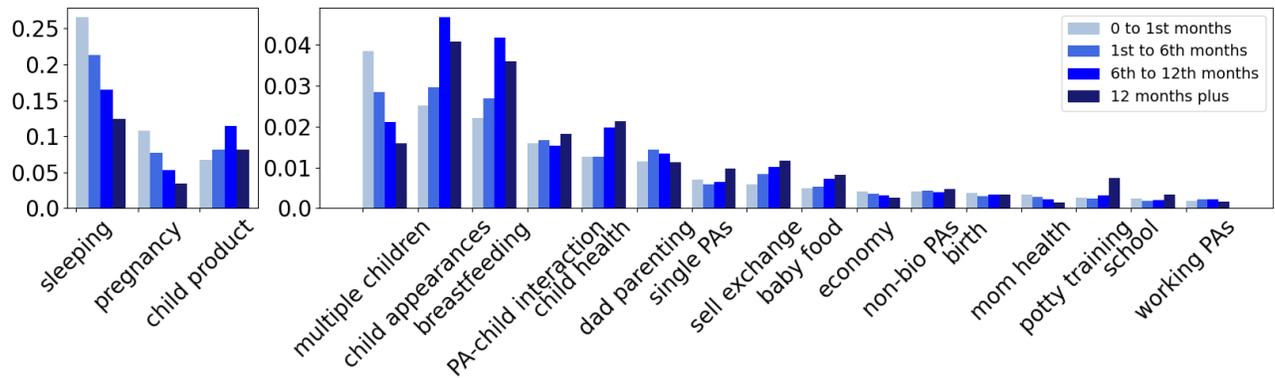


Figure 3: Changes in each parenting topic frequency (%) over time. Each bar represents the frequency 1) during 1st month since participating in parenting-related subreddits, 2) during 1st - 6th months, 3) during 6th month to 1 year, 4) after 1 year

Predictors	Model 1	Model 2	Predictors	Model 1	Model 2
<i>tenure</i>	0.737***	0.742***	economy		1.253***
<i>number of threads initialized</i>	0.790***	0.836***	mom health		0.815
<i>number of comments sent</i>	0.497***	0.501***	multiple children		1.112***
<i>mean received comment length</i>	0.991*	0.992	non-biological parents		1.193***
<i>mean received comment number</i>	1.116***	1.111***	parent-child interaction		0.917**
baby food		0.818**	potty training		0.884
birth		1.045	pregnancy		1.080**
breastfeeding		0.739***	school		1.439***
child appearances		0.879***	sell / exchange		0.926
child health		0.992	single parents		1.195***
child product		0.768***	sleeping		0.833***
dad parenting		0.902*	working parents		1.012

Table 4: Survival analysis predicting how long members continue to participate in parenting community. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$. The number of observations is 524,688. The italic variables are standardized count variables; the rest are binary variables.

accounts to do so; we urge future research to further investigate this. The reason behind the topic school might be that, the topic was meant to focus not on the entire education up to 12th grade, but education only up until preschool. Thus it is expected that this period will be rather short so users might not need to be on the topic for an extended period of time.

8 CONCLUSION AND DISCUSSION

This research studies how parents of infants to preschoolers use online media for parenting, regarding what topics people talk about, how their topical interests change over time, and how discussion of these parenting topics influence their commitment to the community. Compared to previous work, our research looks at a large scale corpus of parenting related discussion with over 600K posts and 10 million comments from 46 subreddits. Utilizing our fine-grained parenting topic taxonomy and large-scale corpus, our quantitative analyses reach the following findings that contributes to theoretical understanding of parenting support groups:

- There were 30 topics that parents frequently talked about online. Moms and dads of infants and preschoolers share a

set of parenting topics, but they focus on different aspects of the same topics and also have different topics of interest.

- Each parenting-related subreddit focuses on a select few parenting topics. The main topics discussed by each subreddit can generally be inferred from the subreddit name, with a few exceptions. The majority of users also showed a narrow set of parenting interests. Users show a shifting of interests over time, in tune with the development stages of their children. Interest and focus of topics transitions from parents to children as time progresses.
- Parents who initialized more threads and interacted with others via comments are more likely to stay in the parenting community. Parents who discuss sensitive topics such as Economy, non-biological parents or single parents tend to stay for a relatively shorter time period.

8.1 Ethical Consideration

This study involves the use of social media data of a sensitive group—parents of infants and preschoolers. Thus, many ethical measures have been taken to effectively ensure the privacy of the user information we have collected. The research study has been

approved by the Institutional Review Board (IRB) at the researchers' institution. The data collection and annotations are anonymous, with pseudo-username in place of actual usernames and other user-related information removed to secure user's personal information. We also paraphrased the posts or comments we quoted in this work to prevent finding original content via web search [13].

8.2 Implications

Our work contributes to the theoretical understanding of parents' use of online media for information and support. Our research not only indicates that parents use online media to discuss different aspects of parenting issues, and we also provide the first human-annotated taxonomy on topics that parents of infants and preschoolers discuss online, which offers a comprehensive and more nuanced overview of online parental discussions. Our taxonomy is generic and domain-agnostic, meaning that it can be referenced to study parenting conversations from online parenting communities in general, not just Reddit discussions. It should be noted that the set of different topics being discussed across different contexts might differ. For example, public platforms like Facebook are widely used by parents to share updates of their children [38], which corresponds to the topic Child Milestones. On the other hand, anonymous or secret parenting groups tend to discuss more inhibited topics. Thus, the application of the taxonomy should take into consideration of the defining factors of the online parenting group, like publicity, anonymity, demographics, etc. We also introduce a set of machine learning models for predicting parenting related topics with reasonable performances, which suggests that it is feasible to utilize computer programs to automatically analyze online parenting related discussions in a large scale with a small amount of ground truth data. The parenting topic taxonomy and identification process described here are reproducible broadly within the HCI community. Our parenting discussion corpus, annotated parenting topics, and code will be made public to facilitate future research. Furthermore, our analyses highlight novel conclusions about parents' use of online media that add to the theoretical foundations of online parenting and open up opportunities for future examination in the domain.

Our findings also shed lights on the design and usability aspects for parenting sites. Based on our findings in Section 6, while user-created forums create diversifying online environments for new parents, it might cause confusion to parents new to the community who wanted to look for specific information or have their questions answered. To provide better support for new parents, our work makes several design recommendations. For example, Reddit offers community moderators the option to pre-define a set of tags for posts, called a "post flair". These "flairs" allow users to quickly search through all posts with the same tag. Building on our research, moderators of parenting subreddits could provide better tags for their communities and users. Another potential possibility for the parenting sites is incorporating a real-time tagging feature when the user is editing a post. Our machine learning model suggests the possibility to build high accuracy models with small amount of ground-truth data to automatically analyze online parenting discussions and generate labels for posts. With a pre-trained model deployed to the site back-end, it can optionally predict parenting topics for the post instantly based on the user input. Incorporating

with our findings in Section 6.1, the system can make recommendations on most relevant subreddits for the topics to post on. As a result, users can locate the most relevant niches and connect with others with similar experiences accurately and efficiently.

Our work revealed that certain topics were associated with limited user engagement, such as *single parents*, *working parents*, *school*, and *non-biological parents* partially due to the sensitive topics, throw-away accounts, and smaller user base, compared to other more general parenting topics like *breastfeeding*. The identification of these topics suggests a direction for the parenting research community to take a closer look at these topics towards better parenting support groups design.

8.3 Limitations and Future Work

One limitation to note for our work is that the quality of parenting topics annotations is not as desirable as expected. While the annotators went through training processes and qualification tests to be selected for the tasks, we acknowledge that potential bias may be introduced in the labeling process. Additionally, due to the limited amount of ground truth labels we are working with, some topics with infrequent occurrence ended up with lower ICC scores and had to be discarded in this work, to ensure the quality and reliability of the analysis. However, we acknowledge that many of these topics, such as Child Behavior Problem and Childcare, remain relevant to our target groups and should be taken into consideration and further examined as future work. Another limitation of our results comes from its correlational nature. Despite the reasonable correlations from our statistical methodology, we cannot explain the causation without natural experiments. This parenting research mainly utilized a quantitative approach to study how parents use online media for parenting, and although we believed such a data-oriented approach would reveal aspects of parent media interaction as a whole that the individual users may not realize, it did not allow for the incorporation of interpretations from users. Future research could employ surveys or interviews to better understand what users were personally looking for, and investigate the level of satisfaction users experience when using Reddit posts as a means for parenting related information, resource, and support. Since we used a large scale dataset from Reddit to understand parenting, many of our findings are specific to Reddit. Our findings should therefore be used with caution when applied to other domains.

ACKNOWLEDGEMENT

The authors would like to thank Robert Kraut, the members of SALT group and the reviewers for their feedback. This work was supported in part by Georgia Tech GVU/IPaT/GTRI Seed Grant. Diyi Yang was supported by a grant from Google.

REFERENCES

- [1] Lara B Aknin, Elizabeth W Dunn, Ashley V Whillans, Adam M Grant, and Michael I Norton. 2013. Making a difference matters: Impact unlocks the emotional benefits of prosocial spending. *Journal of Economic Behavior & Organization* 88 (2013), 90–95.
- [2] Elin K Alfredsson and Anders G Broberg. 2016. Universal parent support groups for parents of adolescents: Which parents participate and why? *Scandinavian journal of psychology* 57, 2 (2016), 177–184.
- [3] Tawfiq Ammari and Sarita Schoenebeck. 2015. Networked empowerment on Facebook groups for parents of children with special needs. In *Proceedings of the*

- 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 2805–2814.
- [4] Tawfiq Ammari and Sarita Schoenebeck. 2015. Understanding and Supporting Fathers and Fatherhood on Social Media Sites. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1905–1914.
 - [5] Tawfiq Ammari, Sarita Schoenebeck, and Daniel Romero. 2019. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
 - [6] Tawfiq Ammari, Sarita Schoenebeck, and Daniel M Romero. 2018. Pseudonymous parents: Comparing parenting roles and identities on the Mommit and Daddit subreddits. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, NY, USA, 1–13.
 - [7] M. I Armstrong, S. Birnie-Lefcovitch, and M. T Ungar. 2005. Pathways between social support, family well being, quality of parenting, and child resilience: What we know. *Journal of child and family studies* 14, 2 (2005), 269–281.
 - [8] Sabine Baker, Matthew R Sanders, and Alina Morawska. 2017. Who uses online parenting support? A cross-sectional survey exploring Australian parents' internet use for parenting. *Journal of Child and Family Studies* 26, 3 (2017), 916–927.
 - [9] L. Barclay, L. Everitt, F. Rogan, V. Schmied, and A. Wyllie. 1997. Becoming a mother—an analysis of women's experience of early motherhood. *J Adv Nurs Apr*; 25, 4 (1997), 719–28.
 - [10] J. Belsky. 1986. Transition to Parenthood. *Medical Aspects of Human Sexuality* 20, 9 (1986), 56–59.
 - [11] Lindsay Blackwell, Jean Hardy, Tawfiq Ammari, Tiffany Veinot, Cliff Lampe, and Sarita Schoenebeck. 2016. LGBT Parents and Social Media: Advocacy, Privacy, and Disclosure during Shifting Social Movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 610–622.
 - [12] Bureau of Labor Statistic BLS. 2019. Labor force participation rate. (2019). <https://data.bls.gov/cgi-bin/surveymost>
 - [13] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4 (2002), 217–231.
 - [14] Jiaao Chen, Yuwei Wu, and Diyi Yang. 2020. Semi-Supervised Models via Data Augmentation for Classifying Interactive Affective Responses. (2020).
 - [15] Jiaao Chen, Zichao Yang, and Diyi Yang. 2020. MixText: Linguistically-Informed Interpolation of Hidden Space for Semi-Supervised Text Classification. (2020).
 - [16] Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 6, 4 (1994), 284.
 - [17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2019).
 - [18] Patricia Drentea and Jennifer L. Moren-Cross. 2005. Social capital and social support on the web: the case of an internet mother site. *Sociology of Health and Illness* 27 (2005), 920–943. Issue 7.
 - [19] Maeve Duggan, Amanda Lenhart, Cliff Lampe, and Nicole B Ellison. 2015. Parents and social media. Pew Research Center. (2015).
 - [20] Elizabeth W Dunn, Lara B Aknin, and Michael I Norton. 2008. Spending money on others promotes happiness. *Science* 319, 5870 (2008), 1687–1688.
 - [21] Elizabeth W Dunn, Lara B Aknin, and Michael I Norton. 2014. Prosocial spending and happiness: Using money to benefit others pays off. *Current Directions in Psychological Science* 23, 1 (2014), 41–47.
 - [22] M. Edhborg, A.-S. Matthiesen, W. Lundh, and A.-M. Widström. 2005. Some early indicators for depressive symptoms and bonding 2 months postpartum – a study of new mothers and fathers. *Arch Womens Ment Health* 8 8 (2005), 221 – 231.
 - [23] Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding Back-Translation at Scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 489–500.
 - [24] M Geraldine Gage and Donna Hendrickson Christensen. 1991. Parental role socialization and the transition to parenthood. *Family Relations* 40, 3 (1991), 332–337.
 - [25] L Gibson and VL Hanson. 2013. Digital motherhood': How does technology support new mothers? CHI. (2013).
 - [26] Tara L Gruenewald, Diana H Liao, and Teresa E Seeman. 2012. Contributing to others, contributing to oneself: Perceptions of generativity and health in later life. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 67, 6 (2012), 660–665.
 - [27] Kylie D Hesketh, Trina Hinkley, and Karen J Campbell. 2012. Children's physical activity and screen time: qualitative comparison of views of parents of infants and preschool children. *International Journal of Behavioral Nutrition and Physical Activity* 9, 1 (2012), 152.
 - [28] Juyoung Jang and Jodi Dworkin. 2014. Does social network site use matter for mothers? Implications for bonding and bridging capital. *Computers in Human Behavior* 35 (2014), 489 – 495.
 - [29] Adam Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology* 31, 2 (2001), 177–192.
 - [30] Ralph LaRossa. 1997. *The modernization of fatherhood: A social and political history*. University of Chicago Press, Chicago.
 - [31] Mary Law, Susanne King, Debra Stewart, and Gillian King. 2002. The perceived effects of parent-led support groups for parents of children with disabilities. *Physical & Occupational Therapy in Pediatrics* 21, 2-3 (2002), 29–48.
 - [32] Katharine Lee, Konstantina Vasileiou, and Julie Barnett. 2019. 'Lonely within the mother': An exploratory study of first-time mothers' experiences of loneliness. *Journal of Health Psychology* 24, 10 (2019), 1334–1344.
 - [33] Claire Lerner and Marisa O Nightingale. 2016. Tuning In: Parents of Young Children Speak up about What They Think, Know, and Need. *Zero to Three* 37, 1 (2016), 44–49.
 - [34] Clare Madge and Henrietta O'Connor. 2006. Parenting gone wired: empowerment of new mothers on the internet? *Social & Cultural Geography* 7, 2 (2006), 199–220.
 - [35] Demaray Malecki, Christine Kerres and Michelle Kilpatrick. 1991. What Type of Support Do They Need? Investigating Student Adjustment as Related to Emotional, Informational, Appraisal, and Instrumental Support. *School Psychology* 18, 3 (1991), 231–252.
 - [36] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. (2013).
 - [37] Sue Miller and Kay Sambell. 2003. What do parents feel they need? Implications of parents' perspectives for the facilitation of parenting programmes. *Children & Society* 17, 1 (2003), 32–44.
 - [38] Meredith Morris. 2014. Social Networking Site Use by Mothers of Young Children. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, Vol. 10. Association for Computing Machinery, New York, NY, USA, 1272–1282.
 - [39] M. W. O'Hara and A. M. Swain. 1996. Rates and risk of postpartum depression-A meta-analysis. *International Review of Psychiatry* 8, 1 (1996), 37–54.
 - [40] Kim Parker and Juliana Menasce Horowitz. 2015. Parenting in America: Outlook, worries, aspirations are strongly linked to financial situation. Pew Research Center. (2015).
 - [41] Rachel L Piferi and Kathleen A Lawler. 2006. Social support and ambulatory blood pressure: An examination of both receiving and giving. *International Journal of Psychophysiology* 62, 2 (2006), 328–336.
 - [42] S.Y. Schoenebeck. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on Youbemom.Com. *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013* 5793 (01 2013), 555–562.
 - [43] Sarita Yardi Schoenebeck and Amy Bruckman. 2013. If we build it, will they come? Designing a community-based online site for parents. *iConference 2013 Proceedings* 2013 (2013), 54–64.
 - [44] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 86–96. <https://www.aclweb.org/anthology/P16-1009>
 - [45] Donna Spencer. 2009. *Card sorting: Designing usable categories*. Rosenfeld Media, New York, NY, USA.
 - [46] Samantha J Teague and Adrian BR Shatte. 2018. Exploring the transition to fatherhood: feasibility study using social media and machine learning. *JMIR pediatrics and parenting* 1, 2 (2018), e12371.
 - [47] A. L. Toombs, K. Morrissey, E. Simpson, C. M. Gray, J. Vines, and M. Balaam. 2018. Supporting the complex social lives of new parents. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* 18 (2018).
 - [48] Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy. 2016. Who did what: Editor role identification in Wikipedia. *10th International Conference on Web and Social Media (ICWSM-16)*, 10, 1 (2016), 446–455. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13077>
 - [49] D. Yang, R. Kraut, and Levine JM. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. *Proceedings of the 2017 CHI conference on human factors in computing systems* 2017 (2017), 6363–6375.
 - [50] Diyi Yang, Robert E Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* 3142 (2019), 1–14.
 - [51] Diyi Yang, Zheng Yao, and Robert Kraut. 2017. Self-disclosure and Channel Difference in Online Health Support Groups. *Proceedings of the ... International AAAI Conference on Weblogs and Social Media*. *International AAAI Conference on Weblogs and Social Media* 2017 (05 2017), 704–707.
 - [52] Svetlana Yarosh, Sarita Schoenebeck, Shreya Kothaneth, and Elizabeth Bales. 2016. "Best of Both Worlds": Opportunities for Technology in Cross-Cultural Parenting. In *CHI 2016 - Proceedings, 34th Annual CHI Conference on Human Factors in Computing Systems* 10, 1145 (05 2016), 635–647.