Baby blues: Analyzing Facebook and health forums on Pregnancy

Jiazhao Li, V.G.Vinod Vydiswaran
University of Michigan
Ann Arbor, Michigan 48109
{jiazhaol, vgvinodv}@umich.edu

ABSTRACT

Patients and laypersons have multiple avenues to seek information online. On one hand, health forums and community groups allow members to ask questions that are answered by their peers or medical professionals, while Facebook and other social networking sites enable members to seek inputs from their friends. Despite many studies on health forums and social media on pregnant women, little research has explored the cross-platform comparative study on intent and behavior of pregnant women. In the study, we systematically analyze questions posted on pregnancy forums by young mothers and contrast it to a unique dataset of Facebook posts by expecting adolescent women. Distribution of themes across these platforms show significant differences on roles played by the two platforms. We conclude that Facebook is chosen as a self-expression place to seek emotional support while health forum serves as professional information provider.

ACM Reference Format:
and Jiazhao Li, V.G.Vinod Vydiswaran, University of Michigan, Ann Arbor, Michigan 48109, {jiazhaol, vgvinodv}@umich.edu . 2019. Baby blues: Analyzing Facebook and health forums on Pregnancy. In *Proceedings of* . ACM, New York, NY, USA, 5 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Internet has already become the most frequent self-identified source of information for pregnant women to retrieve and share healthcare related information and experience beyond books and clinic pamphlets/ brochures [4]. Most pregnant women search for information at least once a month, most often during early stages of pregnancy [6]. [9] found nearly half of the respondents reported dissatisfaction with information given by health professionals (48.6%) and lack of time to ask health professionals questions (46.5%) as key factors influencing them to access the Internet. Various venues are used by pregnant women include search engines, social media, health forums and etc. However, the roles that these platforms serve as for pregnancy community are not explored.

In most online health forums, there are a list of predefined and user-created user communities, often organized around certain medical conditions, ailments or treatment procedures. Users can join various groups, post questions, respond to others' questions and share their own experience in these communities [15]. In this study, to understand the desire and intents of pregnant women, only question of threads from pregnant groups in MedHelp forum were collected. As for online social media (Facebook), it is a pervasive and central way for youth to orient themselves to world and others and provides a window into their unfiltered values and beliefs [10]. Regardless of race or socioeconomic status, nearly all youth in the US (80%) use online social media, with 71% using Facebook [12]. In our study, 4424 Facebook posts are mined from young, pregnant women (age 16 to 24) during their pregnancy.

Despite of many researches related respectively to user distribution analysis of health forums and social media, little research is about content comparison on pregnancy community cross two different types of platforms. In this study, we propose a text-mining based approach to systematically analyze questions posted on pregnancy forums by young mothers and contrast it to a unique dataset of Facebook posts by expecting adolescent women. Based on medical definition, pregnancy is broken into three time periods, called trimesters. Most women feel differently, both physically and mentally, during each trimester [14]. Instead of analysis on the whole pregnancy period, we compare the difference of contents in trimester-level within and between two platforms. Keyword and pregnancy-related themes are retrieved independently from two datasets using Natural Language Processing(NLP) and evaluated with the similarity based on two similarity coefficient metrics: Kendall's τ coefficient and Jaccard coefficient.

2 DATA DESCRIPTION

The current research explores comparison within and between Facebook and MedHelp. Two independent datasets are used in this study, MedHelp forum dataset and Facebook posts dataset.

2.1 Medhelp

Established in 1994, MedHelp is one of the earliest and most well-known online forums dedicated to supporting user-driven discussions on health or healthcare related topics [15]. Over 10.8 million health discussions have already happened when this work was conducted and 6.9 million of them are about pregnancy.

Data used in our study is part of one completed dataset collected in February 2013 by V.G.Vinod Vydiswaran, more details in reference [15]. Completed set includes all threads (questions, comments reply to questions and comments reply to response) and all user profiles from MedHelp. There are five distinct types of user communities in MedHelp: medical support communities, ask-a-doctor forums, forums on pets, international forums, and user groups [15]. In this paper, we focus on user groups that are initiated by end users

, ,

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

(‘user-created groups ’). Based on descriptions of purpose, all user groups can be categorized into at least 10 categories, including ‘Specific conditions ’, ‘Family support’, ‘Public policy’, ‘Socializin’, ‘Pregnancy’and etc. There are 122 ‘User-created groups’(16.3% of 747) in Pregnancy communities, including attempting to conceive, conditions and complications during pregnancy and post-pregnancy care for the mothers and babies, such as ‘Trying to conceive after 40’, ‘Pregnancy after tubal ligation surgery’and ‘Pregnancy: Sep 2012 Babies Help’[15]. We extracted forums in format of last example (referred as pregnancy groups in this paper): ‘Pragnancy: Month Year Babies Help’which is used to indicate all users have the same expected month of confinement.

2.2 Facebook

As this study is conducted, there are over one billion daily active users on Facebook. Facebook use is primarily determined by two basic social needs, including need to belong and need for self-presentation [3]. To understand the behavior and intents of pregnancy communities, thirty-eight young pregnant women (ages 16-24) from two Midwest primary care clinics in low income areas consented for Facebook mining. Extraction was performed using the Facebook application program interface (API) at two separate times once upon recruitment (usually in first trimester) and later in the second or third trimester. At each extraction, participants granted access to their accounts by signing into Facebook. Access was immediately terminated after the participants logged out of their accounts. These posts were sorted by time stamp so that those outside of pregnancy were excluded from analysis. The gestational age of each post was calculated using the time stamp and last menstrual period (LMP) [10]. After deleting Spanish-language posts, only English-language posts were analyzed.

3 METHODS

In this section, we first display statistics overview on content of posts. Then, similarity analysis was performed through comparison of keywords based on two similarity metrics. Finally, theme analysis was conducted by comparing post themes distribution within and between two datasets in trimester level.

3.1 Statistics overview

The MedHelp forum dataset analyzed in our study includes total 4424 threads from 17 pregnancy forums from July 2012 to October 2013. Each thread consists of one question and responses. In order to analyze the main desire and behavior of pregnancy community, only questions were extracted. Based on date of questions posted and community they are in(expected month of confinement), all questions were separated into four trimesters (sometimes referred as phase in this study): first trimester (0 to 12 weeks), second trimester (13 to 24 weeks), third trimester (25 to 36 weeks) and trimester after baby birth. The distribution of questions is showed in Table 1.

The Facebook dataset analyzed in our study includes total 16,519 posts from thirty-eight pregnant women during pregnancy. With Facebook account authorization ending after baby birth , trimester after baby birth is replaced by trimester before pregnancy. The distribution of posts in four phases is showed in Table 2.

Phases	Questions n = 4424
1.First trimester(**1st-Tri**)	1453 (32.8%)
2.Second trimester(**2nd-Tri**)	1163(26.3%)
3.Third trimester(**3rd-Tri**)	1697(38.4%)
4.Trimester after baby birth(**Tri-AB**)	111 (2.5%)

Table 1: MedHelp threads distribution

Phases	Posts n = 16,519
1.Trimester before pregnancy(**Tri-BP**)	2469 (14.9%)
1.First trimester(**1st-Tri**)	2705 (16.3%)
2.Second trimester(**2nd-Tri**)	2779(16.7%)
3.Third trimester(**3rd-Tri**)	8566(51.8%)

Table 2: Facebook Threads distribution

In order to get preliminary overview and direct understanding on content of posts/questions in different phases of two datasets, we generated the word-cloud for each trimester respectively after using NLP model to process corpus. The results are showed in Figure 1, 2. Qualitatively the content are similar within single platform while different between two platforms. Further quantitative comparison will be performed next.

3.2 Analysis based on keywords

Next, we use two metrics Kendall's τ coefficient and Jaccard coefficient to evaluate the similarity of top rank keywords within and between different datasets in trimester level to confirm the qualitative result derived before.

(a) Kendall's τ coefficient (K- τ Coeff): Kendall's τ similarity is a metric that counts the number of pairwise agreements between two ranking lists [7].

Intuitively, the Kendall correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank between the two variables, and low when observations have a dissimilar (or fully different for a correlation of -1) rank between the two variables.

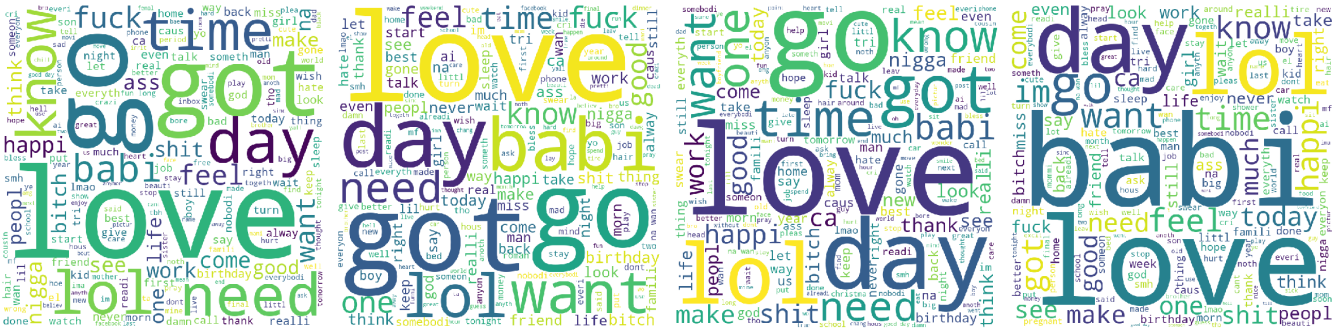
(b) Jaccard coefficient(J-Coeff): The Jaccard coefficient is a statistic used for comparing the similarity and diversity of sample sets, defined as the size of the intersection divided by the size of the union of the sample set.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

where $0 \leq J(A,B) \leq 1$, A and B are sets.

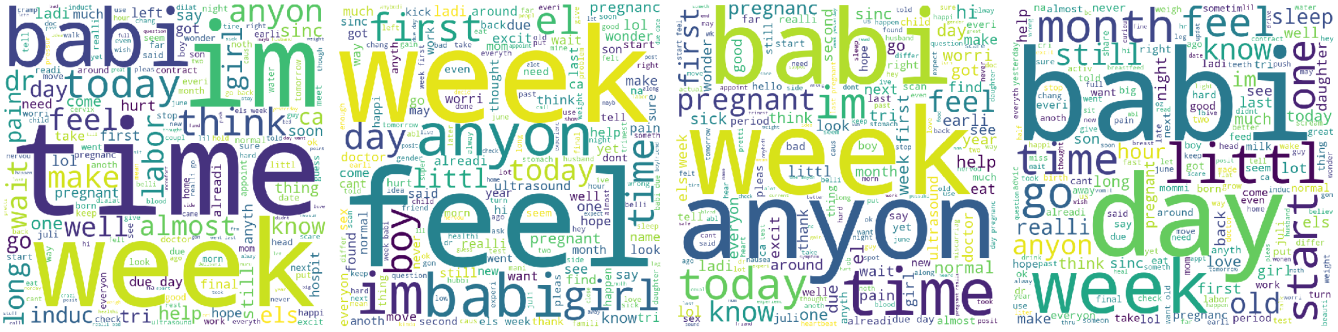
After using Latent Dirichlet Allocation(LDA) to automatically extract the top rank keywords that characterize each phases, we use two coefficients to evaluate the similarity of top 10 and 20 keywords between MedHelp and Facebook in trimester level. Seen from Table 3, two similarity coefficients are significantly small, which indicates pregnant women have different patterns of behaviors and are interested in different themes between different platforms.

Within MedHelp and Facebook, we also compare phase-wised similarity of top 10 keywords using Jaccard coefficient in Table 4 5.



(a) Trimester Before Pregnancy (b) First Trimester (c) Second Trimester (d) Third Trimester

Figure 1: Four phases for Facebook



(a) First Trimester (b) Second Trimester (c) Third Trimester (d) Trimester After Baby Birth

Figure 2: Four phases for MedHelp

Phase		1st-Tri	2nd-Tri	3rd-Tri
K-$\tau Coeff$	Top10	-0.11	0.24	-0.51
	Top20	0.03	-0.09	-0.16
J-Coeff	Top10	0.0	0.0	0.1
	Top20	0.0	0.05	0.05

Table 3: Similarity in same phase among top rank topics between two datasets

Phase	1st-Tri	2nd-Tri	3rd-Tri	Tri-AB
1st-Tri	1.00	0.54	0.54	0.43
2nd-Tri	0.54	1.00	0.54	0.25
3rd-Tri	0.54	0.54	1.00	0.33
Tri-AB	0.43	0.25	0.33	1.00

Table 4: Phase-wised Jaccard similarity in Medhelp

Phase	Tri-BP	1st-Tri	2nd-Tri	3rd-Tri
Tri-BP	1.00	0.54	0.54	0.54
1st-Tri	0.54	1.00	0.43	0.54
2nd-Tri	0.54	0.43	1.00	0.67
3rd-Tri	0.54	0.54	0.67	1.00

Table 5: Phase-wised Jaccard similarity in Facebook

Intuitively, if two sets have half same elements, Jaccard coefficient will be 1/3. The result indicates that both MedHelp and Facebook have high similar themes among different phases. High similarity appears among different trimester phases within single platform. On contrast, significantly low similarity between two data sets in same phase presents that people have common traditional agreement on topics discussed in the one platform. This result keeps consistent with word-cloud overview of corpus.

3.3 Analysis based on categories

To analyze intents and desire of pregnancy community, posts are classified into pregnancy-related categories based on themes, including pregnancy, exercise, support, body-image, diet, advice and addict. The distribution of most common themes identified in these posts from two dataset are showed in Table 6, 7 respectively. From result, proportions of all themes in posts of MedHelp forum are significantly higher than those of Facebook, which keeps consistent with common sense that pregnancy is only part of self-expression in life. On the other hand, high proportions of posts in these themes from Medhelp forum indicate their focus on pregnant-related information. For MedHelp forum, Pegnancy, Exercise and Support rank

Jiazhao Li, V.G.Vinod Vydiswaran
University of Michigan
Ann Arbor, Michigan 48109
{jiazhaol, vgvinodv}@umich.edu

top, compared with Support, Exercise and Pregnancy correspondingly. Such difference on rank demonstrates the important role of social network for pregnant community to seek support while role of health forums as professional information providers.

For each category in two datasets, we notice the proportions are stable among different phases, though there are some varies on specific words. One significant vary is noticed(bold in Table 6): higher frequency of posts in 'Body-image'category in 3rd-Tri and Tri-AB for MedHelp forum. Table 8 shows the frequency of top rank words in four phases, word 'little' is much higher than the other phases which mostly used to describe newborn after analyzing semantics distribution shown in 9. Another explanation on this is that some women view weight gain and changes in their body as positive signs of healthy fetal status while some viewed weight gain as negative and commented on the desire to return to their pre-pregnancy body shape quickly after delivery. [10]

Themes	1st-Tri	2nd-Tri	3-Tri	Tri-AB
Pregnancy	39.7%	39.2%	33.2%	44.0%
BodyImage	28.8%	28.2%	**37.6%**	**35.1%**
Support	14.8%	15.5%	14.4%	23.4%
Exercise	13.1%	9.9%	5.4%	7.2%
Advice	10.8%	9.4%	7.1%	14.4%
Diet	7.9%	5.5%	8.3%	16.2%
Addict	0.4%	0.3%	0.2%	1.8%

Table 6: Frequence of posts in different themes for MedHelp

Themes	Tri-BP	1st-Tri	2nd-Tri	3rd-Tri
Support	10.8%	12.1%	11.6%	12.1%
Pregnancy	5.7%	5.9%	4.9%	9.8%
Bodyimage	2.7%	3.1%	2.6%	**6.0%**
Exercise	1.7%	2.0%	2.1%	2.7%
Diet	1.7%	2.1%	1.9%	3.5%
Advice	1.5%	1.5%	1.2%	2.3%
Addict	0.2%	0.1%	0.2%	0.1%

Table 7: Frequence of posts in different themes for Facebook

Phase	Top rank words
1st-Tri	little (15.3%), big (6.2%), weight(3.2%)
2nd-Tri	little (14.9%), gain (4.3%), weight(3.7%)
3rd-Tri	little (10.2%), gain (3.7%), weight(3.1%)
Tri-AB	little (32.1%), big (6.9%), weight(6.1%)

Table 8: Distribution of top words in Body-image categories after baby birth

Phase	Facebook	MedHelp
Tri-BP	8.70%	–
1st-Tri	17.14%	4.74%
2nd-Tri	16.67%	5.58%
3rd-Tri	29.60%	1.83%
Tri-AB	–	9.01%

Table 9: Distribution of newborn related word in each phase

4 RELATED WORK

Many researches study behavior and intents of pregnant women by analyzing survey of the user profiles and their activities on Internet. [4]found women with higher education were three times more likely to seek advice than women with less than a high school education, and also that single and multiparous women were less likely to seek advice than married and nulliparous women. [6] found that the majority of pregnant women (92.8 %) reported the stages of birth as the first most searched topic, following by fetal development (81 %) and nutrition in pregnancy (58.3 %). The most often mentioned topics of interest included fetal development, nutrition in pregnancy, medications in pregnancy, pregnancy complication and antenatal care [11]. [13] systematically assess that social media (Twitter) can be used to discover cohorts of pregnant women and to develop and deploy a natural language processing and machine learning pipeline for the automatic collection of cohort information.

4.0.1 Communities in Facebook. By tracking and analyzing behavior of specific groups online can help understand as well as support these groups. For instance, [2] showed that fathers' motivations for using social media include documenting and archiving fatherhood, learning how to be a father, and accessing social support from other fathers. Fathers coming from diverse family environments rely on online spaces to find fathers in similar experiences. Facebook page is proved place to gather cohorts. Cohorts whose children had rare conditions were especially likely to embrace and give back to their online communities SLS Facebook Group. Because they typically could not find extensive resources online or in bookstores, they shared experiential information online with other parents worldwide [1]. [8] found Facebook offer a new and promising platform for new mothers to enact and receive validation of good mothering, both critical processes in the early months of new motherhood.

In our study, interests and behavior of community of pregnant women gets further studied by performing mining on online Facebook posts, including body image, food, exercise, and influences during pregnancy.

4.0.2 Communities in health forums. Health forums are popular venues frequented by patients and caregivers seeking information and support[15]. Such sources are preferred by multi-area study. For instance, in user behavior,[16] identified influential users in an online healthcare community by incorporating users' message similarity and response immediacy; [5] provided a clear picture of the network's social structure by incorporating the block modeling research technique for graphically representing the community member interactions as social positions. In user intent prediction, [17] analyzed the distribution of intents of different topics in the forum by using a multiclass support vector machine (SVM) classifier to classify original thread posts.

5 CONCLUSION

Our study is the first to do cross-platform contents comparison between and within health forum (MedHelp) and social media (Facebook) in trimester period level. Significant difference between two platforms indicates that social media plays importance role in offering support while health forums focus on professional pregnancy-related information. In different phases/ trimesters, high Jaccard coefficient within single platform indicated similar themes discussed, reflecting social traditional agreement on topics on platform. We conclude that Facebook is chosen as a self-expression place to seek emotional support while health forum serves as professional information provider.

REFERENCES

[1] Tawfiq Ammari and Sarita Schoenebeck. 2015. Networked Empowerment on Facebook among Parents of Children with Special Needs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.*

[2] 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.*

[3] Stefan G. Hofmann Ashwini Nadkarni. 2012. Why do people use Facebook. In *Personality and Individual Differences.*

[4] Ludmila N Bakhireva, Bonnie N Young, Jeanne Dalen, Sharon T Phelan, and William F Rayburn. 2011. Patient utilization of information sources about safety of medications during pregnancy.. In *Journal of Reproductive Medicine 56,7-8(2011):339-343.*

[5] Katherine Y. Chuang and Katherine Y. Chuang. 2013. How Do E-Patients Connect Online? A Study of Social Support Roles in Health Social Networking?. In *Journal of Lecture Notes in Computer Science / Information Systems and Applications. Vol. 7812, 193-200.*

[6] Bert1 Fabrizio, Maria Rosaria Gualano, Silvio Brusaferro, Elisabetta De Vito, Chiara de Waure, Giuseppe La Torre, Lamberto Manzoli, Gabriele Messina, Tullia Todros, Maria Valeria Torregrossa, and Roberta Siliquini1. 2013. Pregnancy e-health: a multicenter Italian cross-sectional study on internet use and decision-making among pregnant women.. In *Journal of Epidemiology and Community Health 67,(2013),1013-1018.*

[7] Jean Dickinson Kendall, Maurice; Gibbons. 1948. Rank Correlation Methods. Charles Griffin Book Series (5th ed.). In *Oxford: Oxford University Press.*

[8] Priya Kumar and Sarita Schoenebeck. 2015. The Modern Day Baby Book: Enacting Good Mothering and Stewarding Privacy on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work.*

[9] Briege M Lagan, Marlene Sinclair, and George Kernohan. 2010. Internet Use in Pregnancy Informs WomenâĂŹs Decision Making: A WebâĂŘBased Survey. In *Journal of Birth issues in perinatal care,37(2)(2010),106-15.*

[10] Elizabeth Marshall, Margaret Abigail Moon, Anicia Mirchandani, Grace Smith, Lauren P Nichols, Xinyan Zhao, VG Vinod Vydiswaran, and Tammy Chang. 2018. âĂIJBaby wants tacosâĂİ: Analysis of health-related Facebook posts from young pregnant women. In *52nd Society of Teachers of Family Medicine Annual Spring Conference.*

[11] Sayakhot Padaphet and Carolan-Olah Mary. 2016. Internet use by pregnant women seeking pregnancy-related information: a systematic review. In *Journal of BMC Pregnancy and Childbirth,2016,16:65.*

[12] Pew Research Center. 2014. Teens Fact Sheet. In *http://www.pewinternet.org/fact-sheets/teens-fact-sheet/.*

[13] Abeed Sarker, Pramod Chandrashekar, Arjun Magge, Haitao Cai, Ari Klein, and Graciela Gonzalez. 2017. Discovering Cohorts of Pregnant Women From Social Media for Safety Surveillance and Analysis. In *Journal of Med Internet Res,19(10)(2017),e361.*

[14] UCSF Clinics Centers. 2014. Pregnancy Periods. In *https://www.ucsfhealth.org/conditions/pregnancy/.*

[15] V.G.Vinod Vydiswaran, Yang Liu, Kai Zheng, David A Hanauer, and Qiaozhu Mei. 2014. User-created groups in health forums: What makes them special?. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM).*

[16] Christopher C. Yang and Xuning Tang. 2012. Estimating User Influence in the MedHelp Social Network.. In *IEEE Intelligent Systems, 27(5), 44-50.*

[17] Thomas Zhang, Jason H. D. Cho, and Chengxiang Zhai. 2014. Understanding User Intents in Online Health Forums. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics0.*