

Parenthood Mining Using Hashtag Social Network Mining Approach

Xiaowen Wang
Center of Computer Vision and
Signals Analysis
University of Oulu
Oulu, Finland
xiaowen.wang@oulu.fi

Mourad Oussalah
Research Unit of Medical
Imaging, Physics and Technology
University of Oulu
Oulu, Finland
mourad.oussalah@oulu.fi

Mika Niemelä
Center for Life-long Health
Research Unit
University of Oulu
Oulu, Finland
mika.niemela@oulu.fi

Tiina Ristikari
Itla Children's Foundation
ITLA
Helsinki, Finland
tiina.ristikari@itla.fi

Abstract—Low birth phenomenon is a global trend with an important economic and societal consequences that stress global and holistic initiatives in western countries. This paper attempts to contribute to this issue by developing an original approach to mine parenthood pattern from Twitter. The relevant hashtags are employed to develop a social network-based approach to distinguish relevant communities. In the three communities identified, we conducted topical based analysis to comprehend the discussion trend and sentiment analysis to further monitor issues of parenthood concerns. Examples of concerns raised by the community analysis include anxiety in autism parenting, stress in single parenting, college loans, and debts in pandemic time.

Keywords—parenthood, social network analysis, hashtag, Twitter

The type of the submission as "Full/Regular Research Paper". The paper is being submitted to "Research Track on Social Network Analysis, Social Media, & Mining (CSCI-RTNA)".

I. INTRODUCTION

The declining fertility rate has emerged as a global trend in 21st century but also as a serious concern to western economies, which is set to have a "jaw-dropping" impact on society [1]. For instance, Finland's birthrate has dropped to its lowest level of 1.35 children by woman in 2019 according to Statistics Finland [2]. Motivated by the urgent need to comprehend the various components that influence the occurrence of low-birth rate, many researchers and policy-makers have raised the paramount importance of parenthood and also its effect on parental well-being (e.g., [3, 4, 5, 6, 7, 8]). In the 2010s, the contexts of parenting are not only related to the rising economic insecurities and inequalities but also a diffusion of intensive parenting ideology based on the recent review work of parenthood and well-being from Nomaguchi and Milkie [3]. Indeed, parents face several parenting challenges often linked to lack of preparedness, new psychological, societal and economic burden as well new parenthood responsibilities although some evidence has shown that generally parenthood has a positive effect on parents' subjective well-being (e.g., [9, 10]).

With the emergence of World Wide Web in 1990s, new solutions and interaction modes are provided to parents to seek help and support from the various online services (i.e., websites, online discussion forums, blogs) [11]. These online services have acted as an important role in supporting child-raising

mothers [12-16] or fatherhood [17-20], especially with special needs kids. As one of the most popular online activities, social media has been developed enormously and widely used in people's daily life. Parents can access the social media platforms quite easily to share and exchange experiences with others regardless their physical locations owing to the advent of mobile devices [11]. Investigating the social patterns from online platforms becomes crucial to understand people's concerns about parenthood with the purpose of providing possible solutions for parents from the perspective of policy making. Despite the acknowledged noisy nature of social media, valuable insights can be inferred from the data exploration and analysis that can benefit researcher, policy maker and the general-public.

Among these social media platforms, Twitter is ranked as one of the most popular one with merely 400 million users currently. Driven by emotions and desire to influence the community, twitter users freely post their individual's opinions in tweets instead of just responding to a question [21]. Twitter data is relatively easier to reach with free API available when compared to other social media platforms. These features provide an edge to capture public perception through Twitter data. Hashtags, consisting of phrase that follows symbol "#", emerged as one the leading forms of communication in Twitter to raise support and take position in a debate concerning various contemporary issues. Twitter hashtag analysis has been reported for events detection [22, 23]. Based on hashtag analysis specially by leveraging event-related hashtags, a mutually generative LDA model was proposed by Xing et al. Hashtags was highlighted to play a role as a semantic representation of the corresponding tweets [22]. In recent years, Social Network Analysis has been employed to study the social structure of hashtag network on Twitter. Habibi, et al. used Social Network Analysis to find the most influential hashtag related to Indonesian COVID-19 on Twitter by measuring various centralities of hashtag network [24]. A network using Twitter hashtag #BersatuLawanCovid19 was examined through Social Network Analysis metrics by Wiiava and Handoko [25]. Most of the papers simply focused on measuring the graph attributes of the network to identify the most influential node without covering the topic of "parenthood". This paper seeks to provide a more comprehensive study of the hashtag network to get insights into public perception of parenthood (i.e., concerns, sentiments) which may influence parental well-being and studying the inferred online communities.

First, we constructed a hashtag network using the hashtags identified in the Twitter data collected using Twitter Academic API. Second, we conducted community detection by implementing NetworkX algorithms. Third, we performed sentiment analysis and keywords extraction for each community (see Fig. 1). The rest of the paper is organized as follows. Section II presents the details of methodology. Section III presents the results obtained by the proposed method. Section IV discusses the results.

II. METHODOLOGY

A. Data collection

In this study, the dataset was gathered through Twitter academic API by querying the predefined three leading hashtags (i.e., *#parenting*, *#parenthood* and *#momlife*) related to parenthood as keywords, which were chosen based on their popularities as per best-hashtags website¹. Initially, a total of approximate 5 million relevant posts were retrieved from 01.01.2018 to 31.12.2020. The main attributes were extracted from API outputs and named as 'id', 'username', 'date', 'text', 'hashtags', 'location', 'lang', and 'followers_count'. Due to the huge size of the twitter data, the filtering strategy includes i) Remove the posts in non-English; ii) Remove the count of followers less than 10, 000; iii) Remove the posts with null and duplicated text; iv) Remove the posts with null values of hashtags; v) Remove the posts with null values of locations. As a result, 254,094 tweets were achieved and used for the following analysis in our study.

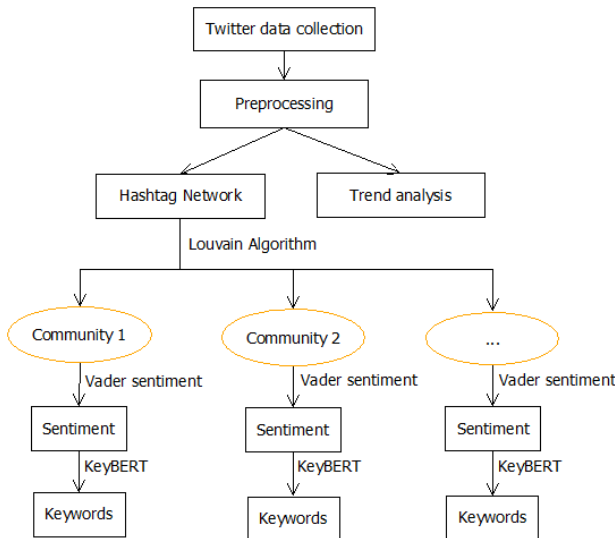


Fig. 1. The diagram of methodology

B. Overall methods

Fig. 1 shows the diagram of the methodology used in this study. As clarified in the introduction part, the purpose of our study is to perform parenthood mining from the perspective of hashtag network analysis. To construct a hashtag network with less noise, all the hashtags were preprocessed by lowering the

case, and manually checking for potential misspelling and simple linguistic errors, for instance, hashtags “#momhood” and “#motherhood” are deemed to be equivalent. Thereafter the communities were detected from the constructed hashtag network by implementing Louvain algorithm (as implemented in NetworkX). As a heuristic method, Louvain’s algorithm is used to extract the communities from large networks based on modularity optimization [26]. The original paper from Blondel et al. demonstrated that this algorithm is faster than other known community detection methods. Then each community was investigated through sentiment analysis and topical analysis. In our case, Vader sentiment [27] which is specifically developed in analyzing sentiments expressed in social media content was implemented to perform sentiment analysis. As one of the recent most popular tools in extracting key phrases from text, KeyBERT [28] was utilized to perform topical analysis to extract keywords and key phrases that most comply with the content of document(s) based on BERT-embeddings language model. The details of network construction and community analysis is described in the following subsections.

C. Network construction and community detection

First, a general trend analysis consisting of extracting the location and the top active users based on the numbers of tweets posted. Specifically, the top influencers in each hashtag class (i.e., *#parenting*, *#parenthood*, *#momlife*) can be captured by investigating the profiles of top-active users (i.e., the number of tweets posted, the number of followers and location) from the collected data. This step allows us to distinguish genuine users from the fake ones, as well as the potential activity associated to the tweet user, if any.

After the previous preprocessing stage, a social network was constructed to exploit the interconnection between all the hashtags from the collected dataset. Formally, given two distinct hashtags, a link is established between these hashtags if both hashtags appear in the same tweet for ten times. In other words, each hashtag was treated as a node in this constructed undirected network, while the link between two nodes is established if there are at least ten tweet messages where the two hashtags occur at the same time. Besides, the number of tweets where the two hashtags co-occur is used as a weight on the corresponding edge of the network. After the above network construction stage, community detection was employed to identify relevant online communities from the network. For this purpose, we employed Louvain algorithm implemented in NetworkX library. Each community or subgraph was quantified by investigating its graph attributes, e.g., Average path length, Average degree centrality, Clustering coefficient.

D. Topical and sentiment analysis of communities

After the identification of the communities, the content of each community is scrutinized from both topical content and sentiment. Initially, a preprocessing stage is also carried out to reduce the noise content. This is carried out by removing

¹ [Hashtags — best hashtags](#)

uncommon characters, numbers, URLs, new lines marks (“\n”), @ and splitting the compounds. To further explore the topics discussed in each community, sentiment analysis was conducted to analyze the emotions (i.e., positive, negative, or neutral) for a given text by Natural language Processing techniques (i.e., Vader sentiment) and topical analysis was performed by implementing keyBERT to extract keywords and key phrases from the documents.

III. RESULTS

A. Trend analysis

The top 6 countries identified from the collected tweets dataset, which was listed as United States (59.4%), United Kingdom (28.2%), Canada (8.4%), France (1.9%), Spain (1.2%), and Australia (1.0%). United States is ranked as the leading country based on the number of Twitter users. Among those three hashtags, #Parenting (57, 363 tweets) and #Momlife (12, 954 tweets) is shown to be the more popular hashtag used in our dataset compared with #Parenthood hashtag class (4, 419 tweets).

B. Hashtag network results

The top 4 nodes with higher degree were labeled as #parenting, #parent, #parenthood, and #momlife. Three central nodes (i.e., #parenting, #parenthood, #momlife) correspond to the three initial class hashtags used to generate all tweets and subsequent hashtags. According to graph attributes summarized in Table I, the network is well connected since it has about two hops of average path length, but has relatively weak density of ties as revealed by low value of global clustering coefficient and average degree centrality. We shall also mention that low value of average degree centrality is also partially caused by the normalization with the maximum degree of the node (i.e., “#parenting”). The node with maximum betweenness centrality was found to coincide with #parenting, which acts like a bridge between different clusters.

C. Community detection & analysis results

To discover the possible communities or themes related to parenthood from the collected dataset, we have performed community detection by implementing NetworkX algorithms. Doubtless there are numerous algorithms which can be applied to identify the communities and the number of communities may end up with more than thousands. In our case, we have chosen Louvain algorithms to detect the communities because of the ease of their interpretation results and sound findings obtained in other studies. As a result, three communities (i.e., “autism-parenting”, “single-parenting”, and “pandemic-parenting”) were selected to elaborate our findings. Other communities were excluded either because of the very small number of nodes involved or because of difficulty to assign a rational interpretation to the overlapping content of the tweet messages involved. For each of these three communities, we explored the sentiment of tweets and extracted the keywords inside to dig out the topics shared by the users. Table I summarizes the metrics measured for each community network in terms of graph attributes, sentiment scores, and keywords extracted.

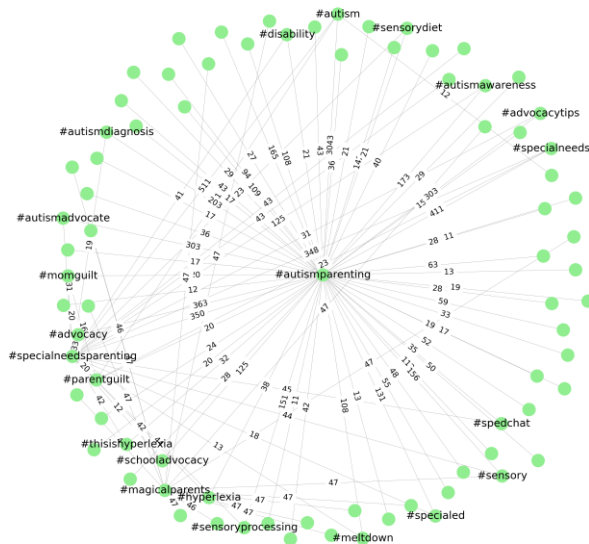


Fig. 2. “autism-parenting” network extracted from the overall hashtag network by implementing Louvain algorithm

The “autism-parenting” network is connected by relatively weak ties since the average degree centrality is relatively low as shown in Table I. In Fig. 2, two main clusters were found with node #autismparenting and #specialneedsparenting respectively. Regarding to node #autismparenting, the neighbors with higher frequency appearing in the same tweets were listed as #advocacy, #autismadvocate, #autismdiagnosis, #advocacytips, #socialskills, etc. The node #specialneedsparenting has more connection with the neighbors of #autism, #goodreads, #childhoodcancer, #childloss, #hairloss, etc. From the associated tweets, the keywords were also extracted as “tips autism parenting, skills autism parenting”. The sentiment analysis results showed that 55.36% of tweets were positive, 29.33% negative and 15.31% neutral. More negative words were found as anxiety, exhausting, anger, struggle, tough, tantrum, meltdown, etc.

In “single-parenting” community obtained using Louvain algorithm, two central nodes were found as #singleparent and #coparenting (see Fig. 3). The most frequent connected node for #singleparent is #singlemom and the most frequent connected node for #coparenting is #divorce. The network is relatively better connected than “autism-parenting” community network since the average path length becomes shorter and max degree centrality becomes higher (see Table I). Based on the keywords, the tweets posted have more concerns on single parenting. From sentiment analysis perspective, as expected most of the tweets are rather positive. The negative tweets are found to be more associated with the stress in single parenting (see Table I).

Metrics	Overall hashtag network	Community networks		
		Autism-parenting	Single-parenting depression	Pandemic-parenting

IV. DISCUSSION

In this paper, we have explored the parenthood from the collected Twitter dataset during the setting periods (2018-2020) from the perspective of hashtag network analysis. We constructed the overall hashtag network and identified three communities from hashtag network using Louvain algorithms. The communities are named “autism-parenting” community, “single-parenting” community, and “pandemic-parenting” community. For each community, we analyzed the sentiments of the relevant tweets and extracted keywords using keyBERT. The sentiment analysis showed that most of the tweets are positive in these three communities. However, this does not mean that negative tweets should be ignored, but rather should be paid greater importance, especially given the acknowledged result that negativity spreads faster than positivity on Twitter [29].

In autism-parenting community, people have more concerns about parenting special needs kids, for instance, kids with autism. As estimated by World Health Organization ², about one in 100 children has autism around the world. Doubtless, raising children with special needs is quite a challenge for parents since parenting strain is greater not only in terms of required time [30] but also, in terms of financial costs [3, 31]. Parents feel more stressed, frustrated and difficulty to cope with their parenthood life. This agrees with the research work related to stress parenthood in [32] and [33]. In modern life, social media provides a good channel for parents to seek help and support towards their parenting issues. Hashtags related to autism advocacy appeared more often on Twitter indicates advocacy plays an important role in supporting people with Autism or family with Autism children.

In single-parenting community, the main topic is related to single parenting and coparenting in parenthood life. Unfortunately, nowadays divorce has become very common in today’s society. According to the American Psychological Association ³, around 40% to 50% of married couples will end in divorce or separation in the United States. Consequently, this trend leads to single parenting or coparenting situations. In comparison with partnered parenting, in single parenting a single mother or father has to handle double amount of parenting responsibilities which bring in more parenting stress and fatigue [3, 34]. As demonstrated in the previous section, parents feel negatively about single parenting life. Peer experience shared on social media acts an important role in giving support for parents who face these challenges.

Besides, the community related to parenting in pandemic time was also detected from the overall hashtag network since

the period of data collection is from 2018 to 2020. Covid-19 crisis dramatically changes people’s daily life and brings more challenges in various fields like health, education, job market, etc. In “pandemic-parenting” community, the discussion is mainly associated with parenting teens in pandemic situation specifically towards college loan and debts. As reported by CNBC ⁴, around 67% of college students worry about the financial situations in the future. Due to pandemic, students could struggle with repaying the loans since the pandemic makes many families financially vulnerable. In addition, “parenting-support” network presents a general picture about the main content related to parenting discussed on Twitter. This community reveals that people share more information about parenting tips especially regarding to motherhood and mom life. This draws the same conclusion that mothers are more likely to agree that social media could provide a source of useful parenting information [11].

The results obtained in this work present the public concerns and their opinions about parenthood, which might help to understand the effects of parenthood on parental well-being and furtherly reveal one of the reasons behind low birthrate. However, there are still some limitations that needed to be addressed in our study. First, Twitter data, although easily accessible, its analysis is often challenged by the inherent limitations of natural language processing toolkits when dealing with highly noisy data. Second, language translation by Google translate API may bring in some bias in sentiment analysis and keywords extraction due to its accuracy. Nevertheless, it should also be noted that with the development of hashtag-based interaction, this offers a nice setting where several hashtags are shared in several social media platforms.

V. CONCLUSION

Parenthood has been identified as a key pillar to understand the phenomenon of low birth rate observed in western countries and considered as a serious economical and societal challenge. In this study, we described a new approach to mine parenthood related patterns from social media platform within worldwide. The developed approach uses hashtags as a basis for community construction and mining. More specifically, we started with three leading hashtags that are found to be highly associated with parenthood. Then Twitter API was used to collect associated tweets. The latter are scrutinized to both identify their additional hashtag contents as well as conduct topical and sentiment related analysis. A social network graph has been constructed where the nodes correspond to hashtags and edge is established whenever two hashtags occur on the same tweet message. Next, Louvain algorithms are used to identify key communities from the underlined social graph. Especially, three communities are highlighted: autism-parenting, single-parenting, and pandemic-parenting. Each community has been investigated in terms of dominant trend, sentiment analysis by Vader sentiment and topical constructed as revealed by the state-of-the-art keyBERT

² [Autism - WHO | World Health Organization](#)

³ [Marriage and Divorce - American Psychological Association](#)

⁴ [Majority of college students are worried about money due to Covid-19 - CNBC](#)

model. A special focus on negative polarity content, which can provide insights in understanding the declining birthrate.

ACKNOWLEDGMENT

This work is partly supported by Finland Iltä Children's Foundation (Itsenäisyyden juhluvuoden lastensäätiö), which is gratefully acknowledged.

REFERENCES

- [1] Gallagher, J. (2020). Fertility rate: 'Jaw-dropping' global crash in children being born. BBC News, 15.
- [2] Timo N. (2020) Births 2019, Statistics Finland, https://www.stat.fi/til/synt/2019/synt_2019_2020-04-24_tie_001_en.html, Last accessed on 2022-03-31.
- [3] Nomaguchi, K., & Milkie, M. A. (2020). Parenthood and well - being: A decade in review. *Journal of Marriage and Family*, 82(1), 198-223.
- [4] Radó, M. K. (2020). Tracking the effects of parenthood on subjective well-being: Evidence from Hungary. *Journal of Happiness Studies*, 21(6), 2069-2094.
- [5] Matysiak, A., Mencarini, L., & Vignoli, D. (2016). Work–family conflict moderates the relationship between childbearing and subjective well-being. *European Journal of Population*, 32(3), 355-379.
- [6] Luppi, F., & Mencarini, L. (2018). Parents' subjective well-being after their first child and declining fertility expectations. *Demographic Research*, 39, 285-314.
- [7] Aassve, A., Mencarini, L., & Sironi, M. (2015). Institutional change, happiness, and fertility. *European Sociological Review*, 31(6), 749-765.
- [8] Margolis, R., & Myrskylä, M. (2015). Parental well-being surrounding first birth as a determinant of further parity progression. *Demography*, 52(4), 1147-1166.
- [9] Balbo, N., & Arpino, B. (2016). The role of family orientations in shaping the effect of fertility on subjective well-being: A propensity score matching approach. *Demography*, 53(4), 955-978.
- [10] Pollmann - Schult, M. (2014). Parenthood and life satisfaction: Why don't children make people happy?. *Journal of Marriage and Family*, 76(2), 319-336.
- [11] Lupton, D., Pedersen, S., & Thomas, G. M. (2016). Parenting and digital media: from the early web to contemporary digital society. *Sociology Compass*, 10(8), 730-743.
- [12] Hall, W., & Irvine, V. (2009). E - communication among mothers of infants and toddlers in a community - based cohort: A content analysis. *Journal of advanced nursing*, 65(1), 175-183.
- [13] Miyata, K. (2002). Social support for Japanese mothers online and offline. *The internet in everyday life*, 520-548.
- [14] Gibson, L., & Hanson, V. L. (2013, April). Digital motherhood: How does technology help new mothers?. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 313-322).
- [15] Lupton, D., & Pedersen, S. (2016). An Australian survey of women's use of pregnancy and parenting apps. *Women and birth*, 29(4), 368-375.
- [16] Madge, C., & O'connor, H. (2006). Parenting gone wired: empowerment of new mothers on the internet?. *Social & Cultural Geography*, 7(2), 199-220.
- [17] Fletcher, R., Vimpani, G., Russell, G., & Keatinge, D. (2008). The evaluation of tailored and web - based information for new fathers. *Child: care, health and development*, 34(4), 439-446.
- [18] Hudson, D. B., Campbell-Grossman, C., OFE FLECK, M. A. R. G. A. R. E. T., Elek, S. M., & Shipman, A. M. Y. (2003). EFFECTS OF THE NEW FATHERS NETWORK ON FIRST-TIME FATHERS' PARENTING SELF-EFFICACY AND PARENTING SATISFACTION DURING THE TRANSITION TO PARENTHOOD. *Issues in Comprehensive Pediatric Nursing*, 26(4), 217-229.
- [19] Nyström, K., & Öhrling, K. (2008). Electronic encounters: fathers' experiences of parental support. *Journal of Telemedicine and Telecare*, 14(2), 71-74.
- [20] StGeorge, J. M., & Fletcher, R. J. (2011). Fathers online: learning about fatherhood through the internet. *The Journal of perinatal education*, 20(3), 154-162.
- [21] Sulis, E., Bosco, C., Patti, V., Lai, M., Fariás, D. I. H., Mencarini, L., ... & Vignoli, D. (2016). Subjective well-being and social media: A semantically annotated Twitter corpus on fertility and parenthood. In *3rd Italian Conference on Computational Linguistics, CLiC-it 2016 and 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, EVALITA 2016* (Vol. 1749, pp. 1-6). CEUR-WS.
- [22] Xing, C., Wang, Y., Liu, J., Huang, Y., & Ma, W. Y. (2016, March). Hashtag-based sub-event discovery using mutually generative lda in twitter. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, No. 1).
- [23] Yılmaz, Y., & Hero, A. O. (2018). Multimodal event detection in Twitter hashtag networks. *Journal of Signal Processing Systems*, 90(2), 185-200.
- [24] Habibi, M., Priadana, A., & Ma'arif, M. R. (2021). Hashtag Analysis of Indonesian COVID-19 Tweets Using Social Network Analysis. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 275-284.
- [25] Wiiava, S. W., & Handoko, I. (2021, January). Examining a COVID-19 twitter hashtag conversation in Indonesia: a social network analysis approach. In *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)* (pp. 1-6). IEEE.
- [26] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
- [27] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI, June 2014.
- [28] Grootendorst, M. (2020). Keybert: Minimal keyword extraction with bert. Internet]. Available: <https://maartengr.github.io/KeyBERT/index.html>.
- [29] Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affective Science*, 2(4), 379-390.
- [30] Miller, J. E., Nugent, C. N., & Russell, L. B. (2015). Risk factors for family time burdens providing and arranging health care for children with special health care needs: lessons from nonproportional odds models. *Social science research*, 52, 602-614.
- [31] Stabile, M., & Allin, S. (2012). The economic costs of childhood disability. *The future of children*, 65-96.
- [32] Schieve, L. A., Blumberg, S. J., Rice, C., Visser, S. N., & Boyle, C. (2007). The relationship between autism and parenting stress. *Pediatrics*, 119(Supplement_1), S114-S121.
- [33] McStay, R. L., Dissanayake, C., Scheeren, A., Koot, H. M., & Begeer, S. (2014). Parenting stress and autism: The role of age, autism severity, quality of life and problem behaviour of children and adolescents with autism. *Autism*, 18(5), 502-510.
- [34] Meier, A., Musick, K., Flood, S., & Dunifon, R. (2016). Mothering experiences: How single parenthood and employment structure the emotional valence of parenting. *Demography*, 53(3), 649-674.