Topic modeling analysis of Community Savings Groups: evidence from the combined literature

Remilien, Emmanuel[1]; Vivar-Miranda, Rufino[1]; Servín-Juárez, Roselia[2]; Valtierra-Pacheco, Esteban[1]; Pérez-Hernández, Luz María[1]

[1] Colegio de Postgraduados-Campus Montecillo, Carretera México-Texcoco km 36.5, Montecillo, Texcoco, Estado de México, C. P. 56264.
[2] Colegio de Postgraduados Campus Córdoba, Carretera Federal Córdoba-Veracruz km 348, Manuel León, Amatlán de los Reyes, Veracruz, México, C. P. 94953.
* Correspondence: rvm@colpos.mx

ABSTRACT

Objective: To identify topics of study related to Community Savings Groups (CSGs) from the literature available.

Design/Methodology/Approach: Some phases of natural language processing were implemented. The data were obtained from scientific databases and gray literature. The analysis included 228 abstracts of papers, theses, working papers, and reports. The Latent Dirichlet Allocation model was used to identify the topics.

Results: Six topics have been the focus of the discussion about the existence and impact of CSGs in the last two decades.

Study Limitations/Implications: Texts without abstracts in English were excluded.

Findings/Conclusions: The topics found deal with various aspects on which CSGs have had an impact, such as health, empowerment, primary sector, violence, and other aspects related to existence. Additional evidence is required to consider CSGs as an effective and sustainable mechanism that facilitates well-being.

Keywords: Savings Groups, automatic literature review, impact, sustainability, participation.

INTRODUCTION

Scientists generally use Scopus, Web of Science, or PubMed to analyze the scientific literature on a given topic, because these databases contain high-impact publications (Falagas *et al.*, 2008). However, valuable information can also be found in gray literature (theses, working papers, or reports). Some systematic literature reviews use both sources (Sollis *et al.*, 2022). Overall, scientific databases are used for bibliometric analysis, seeking to identify authors, journals, impact of publications, or relationships between authors and topics, in order to identify knowledge gaps (Donthu *et al.*, 2021); however, this analysis ignores the knowledge provided by gray literature.

Combining both sources of information (*i.e.*, scientific and gray literature) is important in literature review. Topic modeling (TM) analysis —which is part of text mining— was used in the identification of topics in documents from both sources.

Citation: Remilien, E-. Vivar-Miranda, R., Servín-Juárez, R., Valtierra-Pacheco, E., & Pérez-Hernández, L. M. (2024). Topic modeling analysis of Community Savings Groups: evidence from the combined literature. *Agro Productividad*. https://doi.org/10.32854/agrop.v17i17.2519

Academic Editors: Jorge Cadena Iñiguez and Lucero del Mar Ruiz Posadas

Received: February 20, 2023.
Accepted: December 17, 2023.
Published on-line: February 16, 2024.

Agro Productividad, 17(1). January. 2024. pp: 35-41.

TM is used to handle the large amount of information available as texts in sources such as Twitter, web pages, news, or books (Blei, 2012).

The most popular TM algorithm for the detection of topics is Latent Dirichlet Allocation (LDA), which is part of an Unsupervised Machine Learning (Griffiths and Steyvers, 2004; Kao and Luarn, 2020). TM makes use of artificial intelligence with natural language processing and machine learning. Previous studies used the LDA algorithm to discover topics in textual data. Rivera Delgado *et al.* (2021) analyzed 125 documents related to institutions and participants to find topics on the direct uses of geothermal energy and technological innovation systems in Mexico. Kao and Luarn (2020) focused on Twitter data to find social enterprise topics.

This study focused on the application of LDA to the Community Savings Groups (CSG) literature. CSGs are one of the informal financial institutions created by people excluded from the formal financial system (Amponsah *et al.*, 2023). CSGs are set up with the help of a facilitating institution (FI) which is generally a Non-Governmental Organization (NGO) that carries out CSG establishment projects. After a certain period (usually one year), CSGs are considered to be consolidated and its members are expected to be able to maintain the operation from that point onwards. In 1990, the CARE (Cooperative for Assistance and Relief Everywhere) NGO started the CSG movement in Mali; subsequently, other NGOs replicated that model with some modifications in other regions of Africa. After a decade, the CSGs —which had already spread to several regions of Latin America— were the subject of several researches. The objective of this research was to use the LDA to identify the topics of the publications made about CSGs.

MATERIALS AND METHODS

Documents about CSGs published until 2022 were searched in several scientific databases (Scopus, Web of Science, Google Scholar, PubMed, etc.), in university repositories, and reports from NGO dedicated the creation of CSGs. The abstracts of each document were considered, since they provide the maximum information contained in the document about a given subject (Griffiths and Steyvers, 2004). In total, 228 abstracts were retrieved and considered for the machine learning process.

The phases of natural language processing (NLP) were implemented in the R statistical package, using the tm and topicmodels packages. In the first phase, text processing (cleaning) was carried out through the following steps: text segmentation, tokenization, stop word, stemming, and lemmatization. In the second phase, a document term matrix (DTM) was developed. Defining the number of topics can be a subjective task; however, four available algorithms (CaoJuan2009, Deveaud2014, Arun2010, and Griffith2004) were implemented in the ldatuning package to obtain better results. This process suggests between four and seven topics. Finally, LDA was applied to the matrix (DTM) with the specification of six topics.

A post hoc analysis was implemented to assign names to the topics created to avoid a subjective proposal (Chang *et al.*, 2009). This algorithm is part of the lda package and it functions as the term frequency-inverse document frequency (TF-IDF), where the words with greater frequency in the documents are less important for the formation of topics.

RESULTS AND DISCUSSION

The word cloud was developed using words whose frequency was greater than 70 to determine which concepts were repeated the most in the publications. In addition, a bigram was used to present the relationship that exists between two concepts. The top word correlation has been previously observed to be: "Saving Group", "Saving Loan", "Village Saving", and "Loan Association" pairs, which are grouped under the heading that the literature provides for the CSG, after the word "and" was eliminated in the NLP. Figure 1 shows only the first 15 pairs of concepts —some of which are useful for the analysis of economic empowerment, women empowerment, food safety, etc.

Bigrams indicate relationships between two words. The topics were detected through the implementation of LDA with six topics, yielding satisfactory results. In LDA, each concept influences the formation of the topic, as represented by the Beta value. Figure 2 shows each topic and the Beta value for the concepts. The graph was limited to 10 concepts.

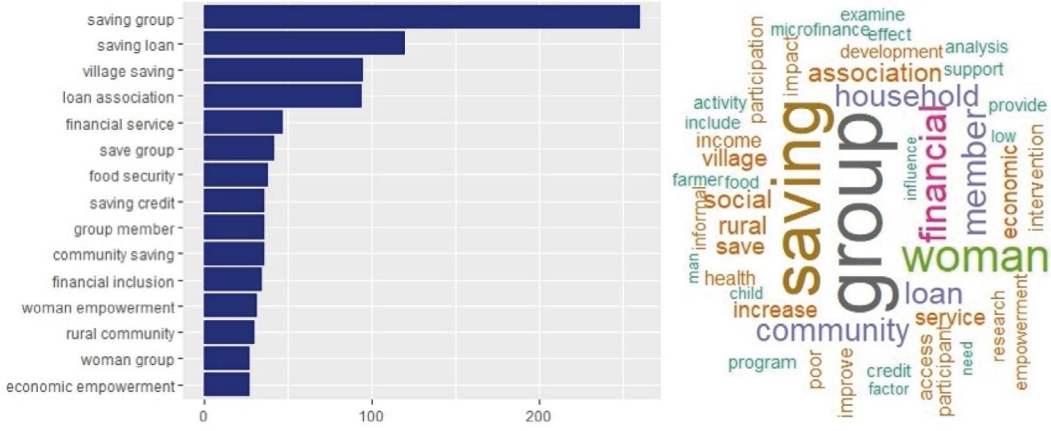

Figure 1. Bigram and word cloud.

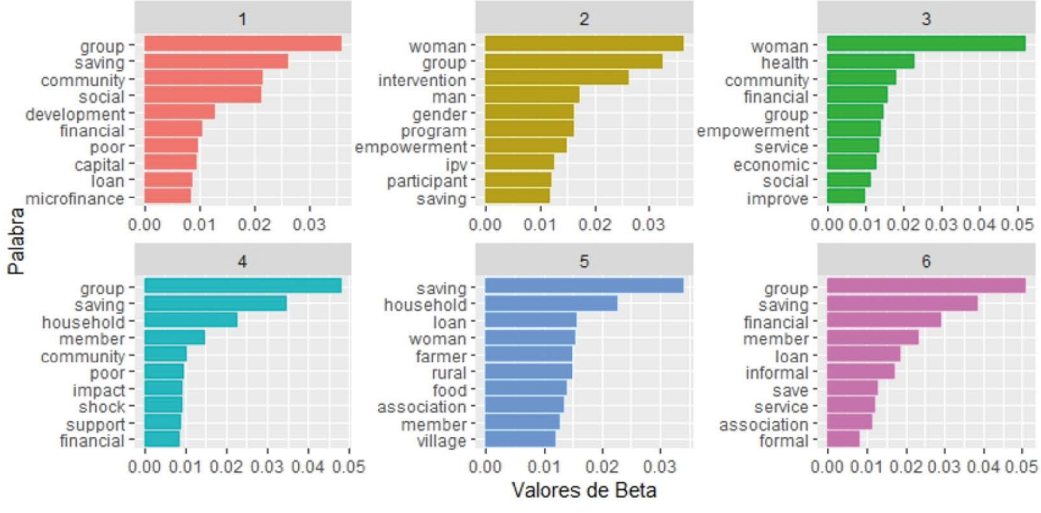

Figure 2. Result of the LDA model.

The names of the topics are defined according to the result of the post hoc test. Table 1 shows the first ten concepts presented in Figure 2 and five additional terms that were not presented in the same Figure.

Topic one takes the essence of CSGs as an alternative model for the financing of rural development in places where formal financial services are lacking. According to Ashe and Neilan (2014), CSGs constitute a revolution that will achieve development through the personal efforts of the poor. There are two types of objectives targeted by development finance: productive activities and infrastructure. CSGs, along with microfinance institutions, play an important role in the financing of activities that generate income in rural areas. Karlan *et al.* (2017) found that CSGs improve the businesses of the poor.

Topic two constitutes the importance of the gender approach in the study of CSGs, since the word "woman" has the highest Beta value in the graph. Several works found that CSGs empower women (Karlan *et al.*, 2017; Pamuk *et al.*, 2022). The CSGs constitute a mechanism that trains women on various topics. Some experimental works include men in workshops given at CSGs. For example, Dunkle *et al.* (2020) found that intimate partner violence (IPV) was reduced in couples who participated in workshops.

CSG training may vary depending on the interest of the NGO that facilitates its creation. Health is the third topic and it is frequently included in the workshops. PubMed is, in fact, one of the databases used to find papers about CSGs, which justifies including the health topic. While the previous topic involves the analysis of a single variable (*i.e.*, women empowerment and consequently the reduction of violence), this topic considers socioeconomic level and health simultaneously. The topic of maternal care is included in this group. Women's participation in a CSG is associated with four or more prenatal care visits, use of qualified midwives, and postnatal care no later than 48 hours after birth (Tura *et al.*, 2020).

Participation and impact on the resilience capacity of CSG members is another important topic for the analyses of these groups. The impact of the CSG can be detected in several aspects, both within the group and outside it (*i.e.*, participants and non-participants).

Table 1. Concepts and headings of the topics

ID	Additional concepts	Post hoc	Name of the Topic
1	save, rural, support, credit, member	Social, institutional, capital, urban, community, scheme	Scheme of Alternative Finance
2	effect, increase, hiv, session, violence	Intervention, ipv, man, woman, empowerment, gender	Impact on gender
3	improve, rural, save, examine, village	Woman, health, empowerment, care, maternal	Socioeconomic empowerment and health
4	child, increase, cope, participant, provide	Shock, cope, household, group, child	Participation and impact on resilience
5	income, participation, farm, agricultural, group	Farmer, food, farm, household, agricultural	Participation and impact on agricultural sector
6	sustainability, performance, institution, bank, access	Informal, sustainability, financial, group, performance, return	Sustainability and performance

Jahns-Harms (2017) found that, through the access they provide to savings and social support, CSGs increase the capacity to face crises resulting from inflation, illness, and poor harvest, among other causes. Other works found that CSG participants take less time to recover from a crisis than non-participants (Wagner *et al.*, 2022; Panman *et al.*, 2022).

The fifth topic is the participation and impact of the CSG in the agricultural sector and it is taken into consideration because the projects are usually implemented in predominantly agricultural territories. Pamuk *et al.* (2022) concluded that CSGs positively affect the adoption of climate-smart agriculture practices. Amponsah *et al.* (2023) found that participation in CSGs significantly increased the agricultural productivity and income of the participants. Another important aspect of this topic is the role that CSGs play in food security, although Lukwa *et al.* (2022) mention that this is not the main reason which drives the members to participated in these groups. Although this topic is included based on the algorithm, in fact it concerns the entire primary sector and there is evidence regarding fishing (Lieng *et al.*, 2018) and livestock breeding (Okello *et al.*, 2020).

Finally, the terms found for topic six suggest that the continuity or performance of the CSGs and their relationship with formal institutions (banks) should be analyzed. CSGs are informal institutions, although their operation is based on written rules. Sustainability is included in the research simply because CSGs must operate on their own when they are no longer assisted by the NGOs that favored their creation. Moret *et al.* (2021) found that CSGs realize changes to survive and the most important factors that enable sustainability are leadership, trust, and love.

In short, these topics reflect the points of view from which the CSGs have been analyzed, including their characteristics, objectives, the benefit they provide to their members (such as empowerment and reduction of violence), and their impact on various sectors of society. CSGs are not only focused on solving women's savings and credit problems; they also constitute a space with multiple purposes.

CONCLUSIONS

Community Savings Groups (CSG) are informal financing mechanisms created by people who do not have access to the formal financial system and are set up with the support of a facilitating institution. The CSG movement, which began in the 1990s, has been the subject of many publications. Summaries of theses, papers, working documents, NGO reports, books, and other sources were compiled to analyze the topics researched in those studies.

The impact of CSGs on health, childcare, agriculture, and women empowerment has been tackled by many studies. However, the main impact is expected to be economic —*i.e.*, granting access to financial resources to carry out productive activities. Many users are attracted to the considerable benefits on savings (passive interest rate) that CSGs offer as part of the strategies aimed at their own existence. Likewise, these groups have to maintain their operations (*i.e.*, they must continue in the long term after the departure of the NGO); therefore, few studies focus on analyzing this aspect.

Further research should be focused on the separate or comprehensive analysis of aspects of the topics determined in this work and even propose new models and more topics.

The CSGs themselves are constantly transforming and growing globally. For example, the Mexican government, through the Programa Sembrando Vida (PSV), promotes the creation, under a more comprehensive approach, of CSGs in the groups known as Comunidad de Aprendizaje Campesino (CAC). Consequently, new topics of interest may potentially arise, including the Government-CSG relationship, the impact on forestry, family well-being, and even sustainability.

REFERENCES

Amponsah, D., Awunyo-Vitor, D., Wongnaa, C. A., Prah, S., Sunday, O. A., & Acheampong, P. P. (2023). The impact of women groundnut farmers' participation in Village Savings and Loans Association (VSLA) in Northern Ghana. *Journal of Agriculture and Food Research*, 11, 100481. https://doi.org/10.1016/j.jafr.2022.100481.

Ashe, J., & Neilan, K. J. (2014). In their own hands: how savings groups are revolutionizing development. Berrett-Koehler Publishers.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22. Recuperado el 4 de febrero de 2023: https://papers.nips.cc/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296.

Dunkle, K., Stern, E., Chatterji, S., & Heise, L. (2020). Effective prevention of intimate partner violence through couples training: a randomised controlled trial of Indashyikirwa in Rwanda. *BMJ global health, 5*(12), http://dx.doi.org/10.1136/bmjgh-2020-002439.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal, 22*(2), 338-342.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl.1), 5228-5235.

Jahns-Harms, E. (2017). Coping with shocks: The role of Savings Groups in rural El Salvador. Working Paper. The Fletcher School, Tufts University. Recuperado el 10 de febrero de 2023: https://static1.squarespace.com/static/55469067e4b06698c52ceb10/t/58ba82cab3db2b088792cdc9/1488618194719/Jahns-Harms+Coping+and+SG+Working+Paper+25Feb2017+.pdf

Kao, S. W., & Luarn, P. (2020). Topic modeling analysis of social enterprises: Twitter evidence. *Sustainability, 12*(8), https://doi.org/10.3390/su12083419.

Karlan, D., Savonitto, B., Thuysbaert, B., & Udry, C. (2017). Impact of savings groups on the lives of the poor. *Proceedings of the National Academy of Sciences, 114*(12), 3079-3084.

Lieng, S., Yagi, N., Mori, A., & Hastings, J. D. (2018). Savings-group improvements contribute to sustainable community-fisheries management: A case study in Cambodia. *Sustainability, 10*(8), https://doi.org/10.3390/su10082905.

Lukwa, A. T., Odunitan-Wayas, F., Lambert, E. V. & Alaba, O. A. (2022). Can informal savings groups promote food security and social, economic and health transformations, especially among women in urban Sub-Saharan Africa: A narrative systematic review. *Sustainability, 14*(6), https://doi.org/10.3390/su14063153.

Moret, W., Swann, M., & Lorenzetti, L. (2021). What happens when savings groups grow up? Examining savings group sustainability and perceived long-term benefits. *Development in Practice, 31*(4), 462-476.

Okello, D. M., Odongo, W., Aliro, T., & Ndyomugyenyi, E. K. (2020). Access to veterinary services and expenditure on pig health management: the case of smallholder pig farmers in Northern Uganda. *Tropical Animal Health and Production*, 52, 3735-3744.

Pamuk, H., van Asseldonk, M., Wattel, C., Ng'ang'a, S. K., Hella, J. P., & Ruben, R. (2022). Community-based approaches to support the anchoring of climate-smart agriculture in Tanzania. *Frontiers in Climate*, 4, https://doi.org/10.3389/fclim.2022.1016164.

Panman, A., Madison, I., Kimacha, N. N., & Falisse, J. B. (2022). Saving up for a rainy day? Savings groups and resilience to flooding in Dar es Salaam, Tanzania. *Urban Forum, 33*(1), 13-33.

Rivera Delgado, D., Díaz López, F. J., & Carrillo González, G. (2021). Transición energética, innovación y usos directos de energía geotérmica en México: un análisis de modelado temático. *Problemas del desarrollo, 52*(206), 115-141.

Sollis, K., Yap, M., Campbell, P., & Biddle, N. (2022). Conceptualisations of wellbeing and quality of life: A systematic review of participatory studies. *World Development*, 160, https://doi.org/10.1016/j.worlddev.2022.106073.

Tura, H. T., Story, W. T., & Licoze, A. (2020). Community-based savings groups, women's agency, and maternal health service utilisation: Evidence from Mozambique. *Global public health, 15*(8), 1119-1129.

Wagner, S., Thiam, S., Dossoumou, N. I., Hagenlocher, M., Souvignet, M., & Rhyner, J. (2022). Recovering from financial implications of flood impact-The role of risk transfer in the West African Context. *Sustainability, 14*(14), https://doi.org/10.3390/su14148433.