Thoughts on Women Entrepreneurship: An Application of Market Basket Analysis with Google Trends Data

Anıl Boz Semerci
  Hacettepe University: Hacettepe Universitesi
Ayşe Abbasoğlu Özgören
  Hacettepe University: Hacettepe Universitesi
Duygu İçen (✉ duyguicn@hacettepe.edu.tr)
  Hacettepe University: Hacettepe Universitesi   https://orcid.org/0000-0002-7940-5064

Thoughts on Women Entrepreneurship: An Application of Market Basket Analysis with Google Trends Data

Anıl Boz Semerci[1], Ayşe Abbasoğlu Özgören[2] and Duygu İçen[3*]

[1]Department of Business Administration,Hacettepe University, Ankara, 06800, Turkey.
[2]Institute of Population Studies,Hacettepe University, Ankara, 06800, Turkey.
[3*]Department of Statistics,Hacettepe University, Ankara, 06800, Turkey.

*Corresponding author(s). E-mail(s): duyguicn@hacettepe.edu.tr;
Contributing authors: anilboz@hacettepe.edu.tr; ayse.ozgoren@hacettepe.edu.tr;

Abstract

This paper focuses on the popularity and awareness of keywords in Google Trends data related to entrepreneurship of women in a global and cross-regional setting by using market basket analysis. Google Trends is one of the digital data platforms that provides a time series index of the volume of queries users enter into Google in a given geographic area. It is the most popular tool for gathering any information, and it has been used in several topics. Market basket analysis indicates items that appear/used together and the frequency of these appearances. Such technique is appropriate in finding hidden associations between items, which is also crucial in assessing individuals' thoughts on a specific topic. This study contributes to the literature in terms of being the first study to use market basket analysis on Google trends data in the context of women's entrepreneurship finding hidden associations between items, which is crucial in assessing individuals' thoughts on a specific topic. The results of the analysis are interpreted through the lens of gender-responsive strategies, equality, efficiency and social justice in different country and region contexts.

Keywords: Women Entrepreneurship, Apriori Algorithm, Market Basket Analysis, Google Trends data

1 Introduction

Quantitative research methods clearly dominate the entrepreneurship study field [33],[40],[49]. However, it has been agreed on that quantitative studies in this field need improvement in terms of sampling and analytical techniques [22]. McDonald et al. [39] in their review of research methods in entrepreneurship 1985-2013 indicated that there is a survey method dominance as data collection method. Short et al. [48], in their interview-based research on editorial board members of Organizational Research Methods indicated difficulties related to measurement techniques, sampling and analysis techniques. According to editorial board members, unsophisticated measurement tools and data gathering by survey methods only permit traditional statistical techniques to be used. Kyrö and Kansikas [33] explained the dominance of relatively classical statistical techniques as means, standard deviation, t-tests and linear correlations. Two years later, Dean, Shook and Payne [13] reported that the entrepreneurship field had a number of areas where new methodological practices were needed. Short et al. [48] have also revealed that even if there are some new approaches in quantitative data collection and

analysis, there is still need to conduct more empirical research by using new analytical procedures in the entrepreneurship field. While the experimental designs [29] , structural equation modelling, and multilevel modelling are the three most frequently mentioned research methods reflecting current entrepreneurship researchers' interests, they would also like to gain new insights on data mining [13].

Data mining, as a new developing research technique for making data, is defined *'as the process of discovering "hidden message," patterns and knowledge within large amounts of data'* [36] which called as big data. Although using big data and related analysis techniques are quite new in entrepreneurship field, there are some studies in management and economic fields that highlight the opportunities and challenges of big data. Gutmann, Merchant and Roberts [18] examined what big data means to economic history and forthcoming sources of big data. Sheng and Wang [47] focused on growing awareness of big data's business values and managerial changes led by data-driven approach. Mesquita, Antonialli and Carvalho [41] intended to explain big data role in marketing research. Adomako and Amoah [2] shed light on decisions and processes in facilitating or hampering firms' ability to harness big data to mitigate the cause of business failures. Since association rules as one of the data mining methods are useful and easy to understand, there have been many successful business applications, including finance, telecommunication, marketing, retailing, and web analysis [24, 30, 51, 53].

Derived from all this literature review and methodological gaps in entrepreneurship, this study aims to bring the new methodological approach into entrepreneurship research, within the women entrepreneurship context.

2 The Reason to Choose Women Entrepreneurship as a Context

Entrepreneurship has often been touted as a viable alternative for women in participation into the economic activities. Having your own business, being your own boss, having autonomy in decision making and flexible working schedules are some features associated with entrepreneurship as a career option for women. Women entrepreneurship is a growing trend in both developing and developed countries. Although a lot of countries make every effort to increase the level of women entrepreneurship in their country, certain regions and countries are more successful in achieving this goal compared to others. A central debate in explanations of these differences between countries concerns on two main explanations: (I) environmental and (II) individualistic factors [8]. While environmental factors are composed of economic, governmental, educational and cultural indicators, individualistic factors consist of psychological profile of individuals. Although previous studies on women entrepreneurship have explained the variety of these concepts and their effects on women's entrepreneurial decisions, most of them have approached it from the same methodological point of view. This methodological shortcoming can be explained in two main aspects based on sampling and analysis technique.

In sampling, most of the researches on intentions, preferences and opinions of individuals on women's entrepreneurship, use cross-sectional survey studies, structured questionnaires [3] and in-depth interviews [25]. On the other hand, the emergence of big data provides new opportunities for monitoring and modelling attitudes towards social and economic issues. Specifically, employing trend analysis using online search data brings rewarding input to evaluate and assess changes in public opinion and perception, which can be thought as a proxy for the level of public knowledge and awareness of specific terms [34].

Technically, many studies in women entrepreneurship field are based on general linear model-based technique. This technique is bounded by assumptions (such as linearity, normality etc.) and provide results based on standard errors, correlation coefficients, and regression coefficients. More specifically, Ahl [3], in her study related to women entrepreneurship, suggests new research directions that produce more and richer aspects of women's entrepreneurship. Ahl [3] has also stated the dominance of correlation tests, t-tests, multiple regression, manova, anova, factor analysis, clustering, discriminant analysis, and logit models as analysis techniques in women entrepreneurship field. These findings also show consistency with previous investigations on research methods in

women entrepreneurship (e.g., Brush [9]; Coviello and McAuley, [11]). More recently, Kuckertz and Prochotta [32] have searched current topics and methods in entrepreneurship by collecting data from current researchers. Therefore, this study aims to examine a new methodological approach and explore its usage in women entrepreneurship context.

2.1 The Reason to Apply Regional Approach in Women Entrepreneurship Context

Gender gap in entrepreneurial activities still remains as challenge for all countries consisting developed, developing and emerging ones. The past 30 years of research has shown that there are stylized facts related to women entrepreneurship. The most apparent one is that there are relatively less women entrepreneurs compared to men [42]. The main reason for this, is not higher rates of failure of women but rather that fewer women start businesses. For instance, in only 6 of 48 economies women's total early-stage entrepreneurial activity (TEA) rates among adults of age 18-64 are higher than or equal to that of men's according to Global Entrepreneurship Monitor (GEM) Adult Population Survey, 2018 [17]. Why are these rates low? In the environmental point of view, economic, legal and socio-cultural characteristics of a society can be the reason behind the low rate of women entrepreneurship.

GEM reports since 1999, have indicated that entrepreneurship does not exist uniformly in the world, but rather has "extraordinary variation". One reason of variation among regions is due to the differences in the level of development. In developing countries, the prevalence women entrepreneurship is higher than in developed countries [42]. Europe and North America regions of the globe have experienced considerable industrial re-structuring in the last three decades, which also helps to develop new entrants and small firms [6]. Women in such developed economies are more likely to start businesses, while those in less developed economies have lower rate of entrepreneurship activities. The reasoning of these varied women entrepreneurship rates can be meaningful when the regional evaluation is used. However, most of research on women entrepreneurship is very "Western-centric". Studies from Latin America, Asia, the Middle East and Eastern Europe are only now emerging [10]. Hence a cross-regional analysis is well-suited to understand and analyze entrepreneurship dynamics of women, and this study at hand employs a regional approach.

In this vein, this study aims to examine new methodological approach and explore its usage in women entrepreneurship context by using market basket analysis on Google Trends data. It contributes to the literature in several ways. First, to the best of authors' knowledge, this study is the first one which employs market basket analysis to Google trends data in women entrepreneurship context. Second this research analyses the trend searched topics related to women entrepreneurship in a given region for the time period of January 2009-November 2019. Third, the findings of these research are important as they reveal the connectivity of keywords related to women entrepreneurship and their inter-associations as being consecutives or descendants. Last but not the least, analysing the popularity and awareness of keywords related to entrepreneurship of women in a global and cross-regional setting provide comprehensive empirical ground for practical suggestions. The following sections detail the Google Trend data set and analytical procedure that are used in current study.

3 Methodology

3.1 Data

Google Trends have become very popular and been widely used in assessing public opinions about several kinds of subjects. It is one of the digital data platforms that provides an index of the volume of queries users enter into Google Search Engine based on a given geographic area (country, state/province and city levels) and time. Since Google is the most widely used search engine in the world [45], it provides compilations of big data related to people's search terms. The search index is a compilation of all Internet queries ever submitted to Google's search engine since 2004. The index for each query term is calculated as the search volume for a specific query in a given geographical location divided by the total number of

queries in that region at a given point in time [1] . Thus, the index is always a percentage from 0 to 100 [52].

For this study, it is investigated whether the key words indicating the attitudes and opinions of women entrepreneurship in the world are popular between January 2009 and November 2019. It was recorded whether the terms considered were popular at least three times on the specified period or not. Popularity was defined as a binary variable which takes a value of 1 when the index is equal to or greater than 75.

In determination of keywords, the computer-based indexing system that depends on frequency analysis of the text was used. The widely used entrepreneurship scales related to 'entrepreneurial intention' (developed by Liñán and Chen [35]), 'entrepreneurial opportunity recognition' (developed by Mom, Van Den Bosch and Volberda, [43]), 'entrepreneurial attitudes' (developed by Kolvereid [31]) and other scales related to women entrepreneurship (such as 'gender stereotypes' (developed by Van Egmond et al. [50]), 'work family conflict'(developed by Marshall and Barnett [37]), 'social support' (developed by Flood [15] that were used in order to reveal supporting and hindering factors in women entrepreneurship context were used. All these scales are combined and written as text. Moreover, the scale on the measurement of environmental incentives (such as governmental support, education and training, cultural and social norms used in GEM Report, [17]) was also included into the text. In the text mining process, stop words such as 'the', 'at', 'and', 'or' etc. were excluded from the keyword data. Twenty-four keywords were obtained, and the results were presented with 'word cloud' visualization (Figure 1).

The geographical regions were selected on the basis of the Global Entrepreneurship Monitor (GEM) categorization. GEM is the world's foremost research on entrepreneurship that assess national level of entrepreneurial activities annually and provides a large-scale database for international comparison [7]. The countries were grouped according to their geographical location

Fig. 1 The word cloud of search words

and it generates four regions with forty-nine countries [2] given as:

- Europe and North America
- East and South Asia
- Latin America and Caribbean
- Middle East and Africa

3.2 Analytical Procedure

Data mining methods that analyse the co-occurrence of events are called Association Rules. These methods present the rules of the events that occur together with the calculated probability values. Association rules are an approach that supports the analysis of past data and the determination of association behaviours within these data and conducting future studies. Association rules are first introduced by Agrawal [4]. Market basket analysis is an example of the application of association rules, which have been very popular since then in data mining applications [16]. The purpose of the association rule used in market basket analysis is to find the relationship between the customers during the shopping process and to determine the buying habits of the customers. Thus, vendors have an effective and profitable marketing and sales opportunity thanks to the relationships and habits discovered.

---

[1]For more information please refer to the Google Trends website at http://www.google.com/insights/search/#

[2]**Europe and North America** : Austria, Bulgaria, Croatia, Cyprus, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Poland, Slovakia, Slovenia, Spain, Sweden, The United Kingdom, Turkey, Russian Federation, Switzerland, United States, Canada.
East and South Asia : India, China, Indonesia, Japan, Thailand, Republic of Korea, Taiwan.
Latin America and the Caribbean : Argentina, Brazil, Chile, Colombia, Guatemala, Panama, Peru, Puerto Rico, Uruguay.
Middle East and Africa : Angola, Egypt, Iran, Israel, Lebanon, Madagascar, Morocco, Qatar, Saudi Arabia, Sudan, United Arab Emirates.

Some of the algorithms used in the literature for Association Rules Analysis are: Apriori, Carma, Sequence, GRI, Eclat, FP-Growth algorithms. Since Apriori Algorithm is the most popular and widely used algorithm in data mining methods, we use this algorithm in our study. In order to implement the Apriori algorithm, the minimum support and minimum confidence values determined by the user must be determined. These minimum requirements are called strong rules [20]. In some sources these rules are also called interesting rules.

An association rule has two parts: an antecedent (if) and a consequent (then). An antecedent is an item found within the data. A consequent is an item found in combination with the antecedent. An example of an association rule for the products of A and B can be shown as "$[A \rightarrow B]$". This means that in the case where A occurs, B will most likely occur. The left side of the expression (may be more than one) is called an antecedent and the right side is called a result (consecutive or descendent). If the expression on the left side is true, the expression on the right side is also true [21]. The importance of this rule can be measured by two different measures. These are support and confidence. The support level of this association rule is given below:

$$\text{support}(A \rightarrow B) = \frac{n(A \cup B)}{N} \qquad (1)$$

In the formula, $n(A \cup B)$ represents the number of transactions where A and B are together, and N represents the total number of transactions. If this value is one, it means that A and B occur together in each process in the examined data set, and that zero means that A and B do not occur together in any process in the data set. On the other hand, the confidence value indicates that what percentage of transactions involving A includes B as well. The confidence level of this association rule is given below:

$$\text{confidence}(A \rightarrow B) = \frac{n(A \cup B)}{n(A)} \qquad (2)$$

The formula given above, unlike the support formula, contains the total number of observations of A in the denominator. An output of this value indicates that each process containing A also contains B.

It is possible to generate too many rules in a medium-sized data set. However, it is also a fact that the rules that do not cover the overall data set are not worth examining. Therefore, it is necessary to list not only every rule but only those rules that meet certain requirements. It is possible to say that minimum confidence and minimum support values limit the number of rules. The highest level of support and confidence values is 100% (or 1), indicating that the relevant rule is included in the entire data set.

For the algorithm to use, the user needs to decide minimum support and minimum trust values. Keeping these values high results in a low number of rules. Likewise, keeping these values too low will result in too many rules. In other words, it is possible to say that minimum support and minimum confidence values control the number of rules. In this study we take confidence and support cut-off level as 85%.

In this study, the important rules are obtained by Apriori algorithm. In order to do this, R studio, which is a popular programming language and software environment for statistical computing and graphical representations of the rules, is used.

4 Results

The results are given in different forms with Table 1, Figure 2 and Figure 3. Market basket analysis results revealed 38 rules in total (Table 1). Support and confidence levels of each rule are given according to the four regions. These rules give us the probabilities of searched terms together and their order of occurrence.

For Europe and North America region, 12 association rules are determined according to the market basket analysis of online search keywords within the context of women entrepreneurship. The results in Table 1 indicate that the most popular terms are [social media] (rule 1), [training] (rule 5) and [firm] (rule 9), respectively. When sorted by the confidence (strength) values, it is seen that [social media] awareness is highest with no antecedent (support value 1.000, and confidence value of 1.000), and with an antecedent as [firm] or [training] or [training, firm]. [Training] is another popular term with a 95% of probability to be searched at least for three times with a Google popularity index above 75. Awareness on

Regions	Rules	Antecedent	Consequent	Support	Confidence
Europe and North America	1	{}	Social media	1.000	1.000
	2	Firm	Social media	0.909	1.000
	3	Training	Social media	0.955	1.000
	4	Training, Firm	Social media	0.864	1.000
	5	{}	Training	0.955	0.955
	6	Social media	Training	0.955	0.955
	7	Firm	Training	0.864	0.950
	8	Social media, Firm	Training	0.864	0.950
	9	{}	Firm	0.909	0.909
	10	Social media	Firm	0.909	0.909
	11	Training	Firm	0.864	0.905
	12	Training, Social media	Firm	0.864	0.905
East and South Asia	1	{}	Opportunity	1.000	1.000
	2	Firm	Network/Networking	0.857	1.000
	3	Network/Networking	Firm	0.857	1.000
	4	Firm	Opportunity	0.857	1.000
	5	Network/Networking	Opportunity	0.857	1.000
	6	Failure	Opportunity	0.857	1.000
	7	Network/Networking, Firm	Opportunity	0.857	1.000
	8	Opportunity, Firm	Network/Networking	0.857	1.000
	9	Opportunity,Network/Networking	Firm	0.857	1.000
	10	{}	Firm	0.857	0.857
	11	{}	Network/Networking	0.857	0.857
	12	{}	Failure	0.857	0.857
	13	Opportunity	Firm	0.857	0.857
	14	Opportunity	Network/Networking	0.857	0.857
	15	Opportunity	Failure	0.857	0.857
Latin America and the Caribbean	1	{}	Network/Networking	1.000	1.000
	2	{}	Failure	1.000	1.000
	3	Network/Networking	Failure	1.000	1.000
	4	Failure	Network/Networking	1.000	1.000
Middle East and Africa	1	{}	Network/Networking	1.000	1.000
	2	Training	Network/Networking	0.909	1.000
	3	Firm	Network/Networking	0.909	1.000
	4	{}	Training	0.909	0.909
	5	{}	Firm	0.909	0.909
	6	Network/Networking	Training	0.909	0.909
	7	Network/Networking	Firm	0.909	0.909

Table 1 Support and confidence values of the rules and their antecedent – consequent terms according to each region

[training] is dependent on antecedents as [social media], [firm], and [social media, firm] with confidence values above 95%. [Firm] is also a popular consequent anteceded by [social media], [training], [training, social media] terms. If the term [firm] is a popular search term, then the probability that the term [social media] is a popular search term is 100% (confidence value) for this region ([firm → social media], rule 2). The confidence value of the rule 12 indicates that if the terms [training] and [social media] are popular search terms, then the probability that the term [firm] is a popular search term is 90.5% ([training, social media → firm]). The support value indicates that the terms [firm] and [social media] co-occur in 90.9% of the data in Europe and North America region ([firm → social media], rule 2).

(a) Europe and North America

(b) East and South Asia

(c) Latin America and Caribbean
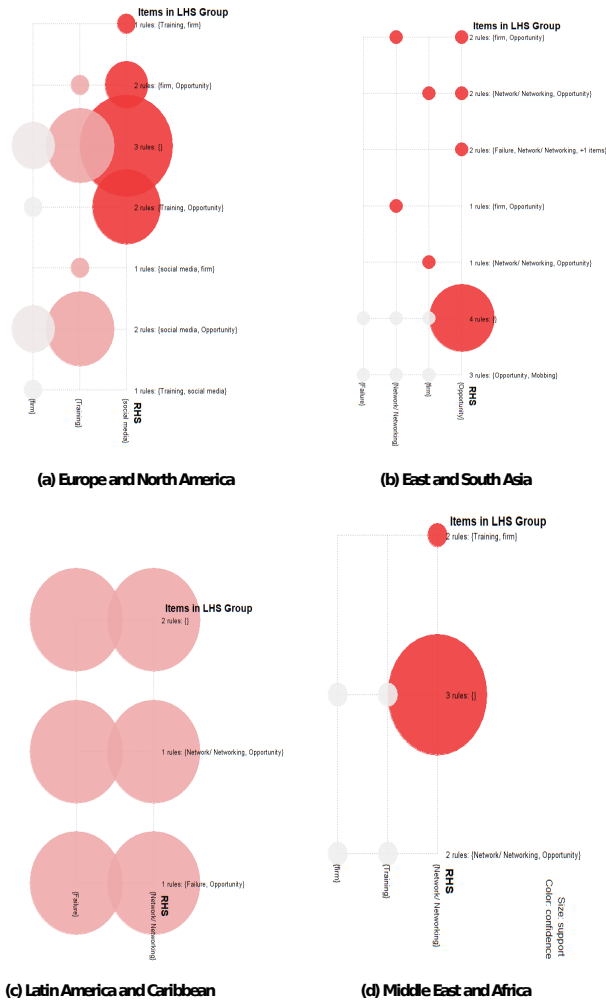
(d) Middle East and Africa

Fig. 2 Grouped matrix-based visualization of the rules

Figure 2 [3] is a grouped matrix-based visualization of the rules of market basket analysis results that are presented in Table 1. Each region is represented in different panels. While left-hand side (LHS) means antecedent terms, right-hand side (RHS) represent the consequent term of the rule. Each of these circles represents the association rules for each region. The colour of the circle shows the confidence value of the rule where the darker ones have higher confidence values compared to others. Also, the larger the size of the circle, the higher support value of the rule [19].

For Europe and North America region four in 12 association rules have [social media] as a

descendant, four have [training] as a descendant and four have [firm] as a descendant term. However, [social media] has higher support and confidence levels as a descendant (right hand side) term as suggested by larger sizes and darker colours of the circles in panel 'a' of Figure 2. Higher confidence values (or darker colours) indicate that [social media] awareness is highly likely once awareness on [firm] and/or [training] is achieved.

Figure 3 summarizes the rules of market basket analysis results with the graph-based visualization, in which each region is represented in different panels. In this graph, search terms and rules are connected to each other with directed arrows. This graph enables to examine how the rules are composed of individual terms and how rules are associated. The colour of the circle indicates confidence level and the size of the circle gives the support value of this rule, similar to Figure 2.

For Europe and North America region, it is seen that social media, firm and training terms are likely to be seen as both an antecedent and a consequent term/s of the rules. Among two-term associations, the probability of co-occurrence of [social media] and [training] terms are highest with a level of 96%. [Social media] and [firm] co-occurrence, and [firm] and [training] appearing together, on the other hand, are less prevalent compared to other rules. This indicates a hierarchy of association rules within the region.

For East and South Asia region there are 15 rules are given in Table 1. Among the rules term [opportunity] has the highest confidence and support values. This is a different term among the rules compared to other regions. The probability that awareness on [opportunity] will be achieved given there is awareness on [firm], [network/networking], [failure], [network/networking, firm] is 100%, and the probability of co-awareness on these term groups is 85.7% for each. [Firm], [network/networking] and [failure] terms are also popular alone with a support and confidence level of 85.7% for each.

Finally, [opportunity] is also an important antecedent. There is an 85.7% probability (support level) that opportunity, firm, and network / networking terms are seen as popular search terms together ([Opportunity, Firm → Network / Networking], rule 8). In addition, if the terms opportunity and network/networking are popular search terms, then the probability that the term

---

[3]Figure 2 also shows the association rules with confidence and support values less than but close to the cut-off level of 85%.
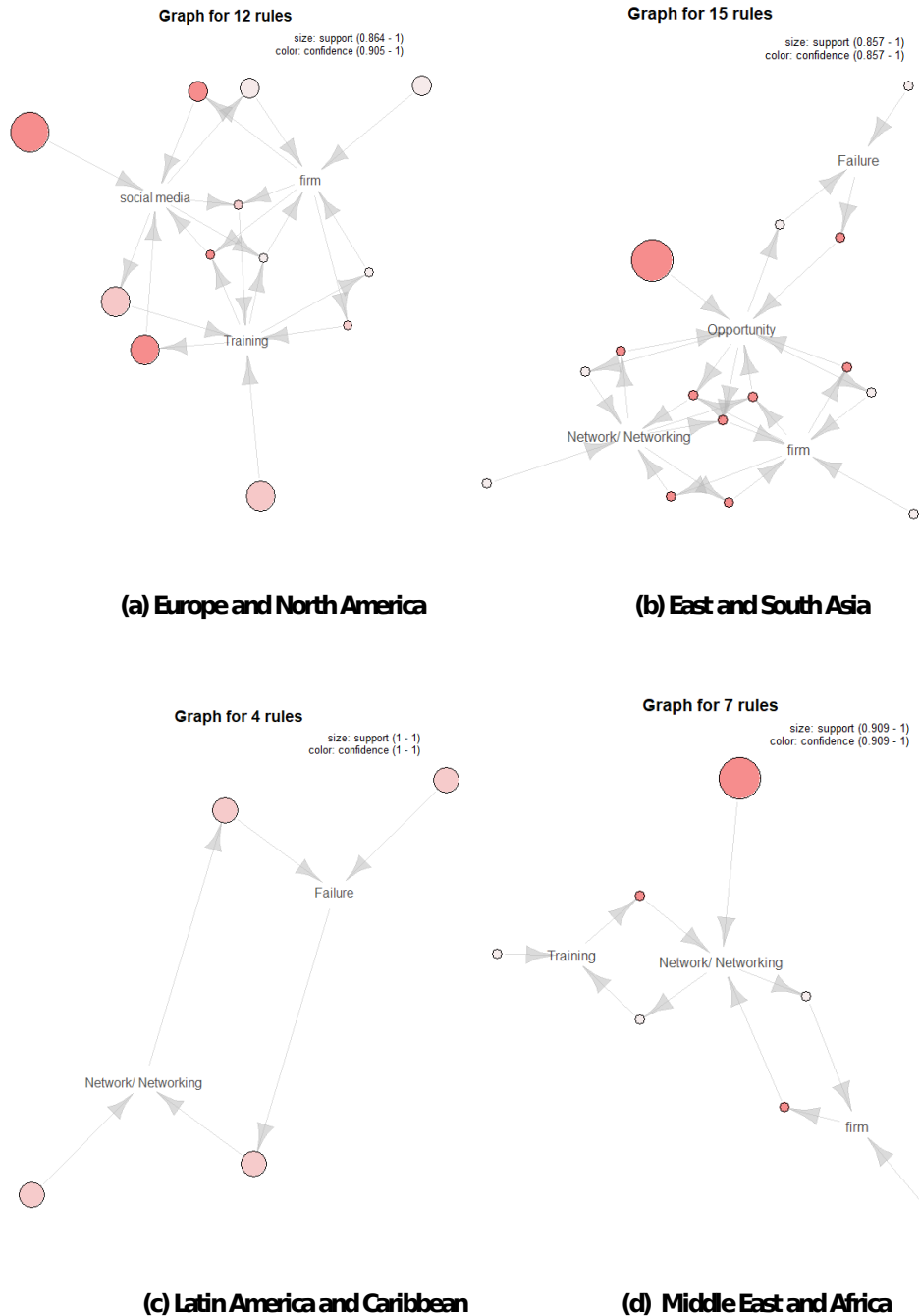
(a) Europe and North America

(b) East and South Asia

(c) Latin America and Caribbean

(d) Middle East and Africa

Fig. 3 Graph-based visualization of the rules

firm is a popular search term is 100% (confidence level) ([Opportunity, Network / Networking → Firm], rule 9). Finally [firm], [network/networking], and [failure] with support and confidence values of 85.7% for each rule (rules 13, 14, 15).

Among all 15 association rules for the region of East and South Asia, the rule with the highest confidence and support values involves only [opportunity] as a descendant with no antecedent (Figure 2). Other rules have the same support values indicated by the same size of circles in panel

'b' of Figure 2. [Failure] is included in fewer number of rules compared to other right-hand side terms which is visible in Figure 2 and Figure 3. In another terms, [network/networking], [firm], and [opportunity] are more connected to each other comparing to other terms.

For Latin America and the Caribbean there are only four rules that consist of [failure] and [network/networking] terms as sole concepts; as antecedents or as descendants (in Table 1). All association rules have confidence and support values of 1 indicating no hierarchy within the rules. Table 1 indicates that the most popular terms in this region are [network/networking] (rule 1) and [failure] (rule 2). Table 1 also indicates that if the term [failure] is a popular search term, then the probability of the term [Network/Networking] is a popular search term is 100% ([Failure → Network/Networking]) and vice versa ([Network/Networking → Failure]). Moreover, there is a 100% probability (support level) that [network/networking] and [failure] terms are seen as popular search terms together ([network/networking, failure]). In this vein, panel 'c' of Figure 2 and Figure 3 has symmetrical shapes.

For Middle East and Africa there are seven rules in Table 1. The first rule indicates that the probability that the keyword of [network/networking] is a descendant with no antecedent is 100%.In other words, the most remarkable popular search term obtained from market basket analysis is [network/networking]. Given that there is awareness on [training] or [firm], awareness on [network/networking] is guaranteed in this region ([training → network/networking], rule 2), ([firm → network/networking], rule 3). Terms of [training] and [firm], are also important descendants, individually or as anteced by the term of [network/networking]. The probability of co-occurrence of the terms or the support values of the rules is the same with the value of 90.9% in this region. For all the rules, except the first one, the size of the circles is identical (Figure 2 and Figure 3). It is also seen in Figure 3 that [training], [firm] and [network/networking] terms are likely to be seen as an antecedent or a consequent term of the rules.

5 Discussion

This study was conducted to explore a new methodological approach in women entrepreneurship context. It examined women entrepreneurship concepts by using market basket analysis on regional Google Trends data. Market basket analysis is one of the used methodological approaches for working on big data. It indicates items that appear/used together and the probability of these appearances and relations. Such technique is appropriate in finding non-obvious and hidden associations between items, which is also crucial in assessing individuals' thoughts on a specific topic.

The varying results confirmed the need for a regional analysis to approach the popularity of terms associated with women entrepreneurship. Social media is the most popular trend term in Europe and North America. There are several studies that have examined the impact of the internet, specifically social media, on entrepreneurial practices. Kamberidou [27] has mentioned that many European projects have focused on the benefits and opportunities of information and communication technologies in business activities of women. Duffy and Pruchniewska [14] who interviewed 22 independently employed female professionals in America, suggested that there is a sharp rise in women entrepreneurship in digital context in USA. [44] also helps to explain social media and entrepreneurship match in this region.

In East and South Asia, the four popular pillars of women entrepreneurship appeared to be opportunity, network/networking, firm, and failure. Indeed, for the context of Europe and North America, the relationship between networks and creation of new firms [26] networks and firm survival (as antonym of failure) [23, 46] , and networks and opportunity, specifically opportunity perception or recognition of individuals, [5] are well documented in the literature. Our study shows that these associations may be valid for East and South Asia as well. One unique feature of East and South Asia region is the high popularity of the keyword of [opportunity] and its associations with other women entrepreneurship concepts. Awareness on opportunity is a fact on its own, but it is also a consequent conditional on firm, network, and even failure. Awareness on

opportunity is also an antecedent of awareness on firm, network and failure.

According to De Vita, Mari and Poggesi [12], access to networks influence female entrepreneurship in developing countries. This is also valid for entrepreneurship in general. Networks of entrepreneurs are effective in identifying business opportunities in over 70% of entrepreneurs consulted in the regions of Latin America and East Asia according to the study of Kantis, Ishida and Komori [28]. However, there are regional differences. East Asian entrepreneurs have networks in the business world like colleagues and business contacts while entrepreneurs in Latin America have social networks like relatives, friends, and acquaintances. This may be the reason for the keyword network to be linked with the keyword failure in Latin America and the Caribbean region in our study.

In the Middle East and Africa region, [network/networking] is the most single popular term, and associations as [training → network/networking], [firm → network/networking], [network/networking → training], and [network/networking → firm] are also important rules according to online search data. It is seen that [social media] in results for Europe and North America is replaced by [network/networking] in the Middle East and Africa region. This is an expected finding as the women entrepreneurs in the Middle East region are mostly in the informal sector and their firms are less likely to use Information and Communication Technology (ICT) tools [38].

These findings provide an opportunity to researchers to note the popular search terms in each region. We suggest that further research on women entrepreneurship should focus on these varied trends and associations in each region. Moreover, these results, revealing the popular terms that individuals have searched on Google search engine, would be helpful in interpreting the other empirical studies' findings related to environmental or individual motives of women entrepreneurship.

6 Conclusion

In spite of the growing keen interest in women entrepreneurship research, there are several methodological limitations that obscure the focus of the research and obstruct the advancement of knowledge in the area. Accordingly, this study carries an importance in terms of its statistical technique, data and findings to help to reveal opinions towards women's entrepreneurship in the world. However, similar to all researches, this study should also be considered in light of its limitations. The first limitation is about search terms and time frame. Even the widely used scales were used in order to produce key words that will be searched within Google Trend data set, it may not capture all searched terms related to women entrepreneurship. Moreover, the time frame was limited as 10 years in this study. Further studies can be conducted with larger keywords pool and for longer period of time. This would be helpful in order to illustrate global context of women entrepreneurship more comprehensively. Second limitation concerns the variation of internet usage by countries. These variations can be seen from the Appendix-1 and further studies can consider these variations as weight of countries or regions.

Lastly, contrary to our expectations, fertility, work-family conflict, family support and mobbing terms, which are more related to women entrepreneurship were not found as trend topics within searched regions and the chosen time period. This may be related to gender differences on internet usage which is in favour of males. Abraham, Mörn and Vollman [1] reported that the rate of internet usage of women is still lower than that of men, particularly in developing countries. Men outnumber women among adult internet users in all regions (Asia Pacific, Europe and Latin America) except in North America.

Declarations

Human or animal rights. This article does not contain any studies with human participants performed by any of the authors.

Contribution of authors. All authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

References

[1] Abraham, L. B., Mörn, M. P., Vollman, A. (2010). Women on the web: How women are shaping the internet. ComScore, Incorporated.

[2] Amankwah-Amoah J , Adomako S . (2019) "Big data analytics and business fail- ures in data-Rich environments: an organizing framework". Computers in Industry. 105:204–12 .

[3] Ahl, H. (2006). "Why research on women entrepreneurs needs new directions". Entrepreneurship theory and practice, Vol. 30 No. 5, pp. 595-621.

[4] Agrawal, R., Imielinski, T. and Swami, A.N. (1993). "Mining Association Rules between Sets of Items in Large Databases". Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington DC, 25-28 May 1993, 22(2) of SIGMOD Records, 207-216.

[5] Arenius, P., De Clercq, D. (2005). "A network-based approach on opportunity recognition". Small Business Economics, Vol. 24 No. 3, pp. 249-265.

[6] Baptista, R., Escária, V., and Madruga, P. (2008). "Entrepreneurship, regional development and job creation: The case of Portugal". Small Business Economics, Vol. 30 No. 1, pp. 49-58.

[7] Bosma, N. (2013). "The Global Entrepreneurship Monitor (GEM) and its impact on entrepreneurship research". Foundations and Trends in Entrepreneurship, Vol. 9 No. 2, pp. 143-248.

[8] Boz Semerci, A., Cimen, M. (2017). "Environmental incentives of entrepreneurship: Fuzzy clustering approach to OECD countries". Journal of Global Entrepreneurship Research, Vol. 7 No. 1, pp. 27.

[9] Brush, C. G. (1992). "Research on women business owners: Past trends, a new perspective and future directions". Entrepreneurship Theory and Practice.

[10] Brush, C.G., and Cooper, S.Y. (2012) "Female entrepreneurship and economic development: An international perspective". Entrepreneurship and Regional Development, Vol. 24, No. 1-2, pp. 1-6.

[11] Coviello, N.E., McAuley, A., (1999). "Internationalization and the smaller firm: a review of contemporary empirical research". Management International Review, Vol. 39, No. 3, pp. 223–256.

[12] De Vita, L., Mari, M., and Poggesi, S. (2014). "Women entrepreneurs in and from developing countries: Evidences from the literature". European Management Journal, Vol. 32 No. 3 pp. 451-460.

[13] Dean, M. A., Shook, C. L., and Payne, G. T. (2007). "The past, present, and future of entrepreneurship research: Data analytic trends and training". Entrepreneurship Theory and Practice, Vol. 31 No. 4, pp. 601-618.

[14] Duffy, B. E., and Pruchniewska, U. (2017). "Gender and self-enterprise in the social media age: A digital double bind". Information, Communication and Society, Vol. 20 No. 6, pp. 843-859.

[15] Flood, M (2005). Mapping loneliness in Australia. The Australia Institute, Discussion Paper, No: 76.

[16] Frawley, W. J., Piatetsky-Shapiro, G., Matheus, J. (1991). "Knowledge Discovery in Databases: An Overview", AAAI Press, Menlo Park.

[17] Global Entrepreneurship Monitor (GEM). (2018). Global Report 2017/18. Global Entrepreneurship Research Association (GERA): London, UK.

[18] Gutmann, M., Merchant, E., and Roberts, E. (2018). "Big Data" in Economic History". The Journal of Economic History, Vol. 78 No. 1, pp. 268-299.

[19] Hahsler, M., and Chelluboina, S. (2011). Visualizing association rules in hierarchical groups. In 42nd Symposium on the Interface: Statistical, Machine Learning, and Visualization Algorithms (Interface 2011). The Interface Foundation of North America.

[20] Han, J. and Kamber, M. (2006) "Data Mining Concepts and Techniques". 2nd Edition, Morgan Kaufmann Publishers, San Francisco.

[21] Hand D. J., Padhraic S., and Heikki M., (2001). "Principles of data mining", MIT Press, Cambridge, MA, 2001.

[22] Hill, J. and McGowan, P., (1999). Small business and enterprise development: questions about research methodology. International Journal of Entrepreneurial Behavior & Research, Vol. 5 No. 1, pp. 5-18.

[23] Ingram, P., and Baum, C.A. (1997). "Chain Affiliation and the Failure of Manhattan Hotels, 1898–1980". Administrative Science Quarterly Vol. 42 pp. 68–102.

[24] Insani, R., Soemitro, H. L. (2016). Data mining for marketing in telecommunication industry. In 2016 IEEE Region 10 Symposium (TEN-SYMP) (pp. 179-183). IEEE.

[25] Jännäri, J., and Kovalainen, A. (2015). "The research methods used in "doing gender" literature". International Journal of Gender and Entrepreneurship, Vol. 7 No. 2, pp. 214-231.

[26] Johannisson, B., and Ramirez-Pasillas, M. (2001). "Networking for Entrepreneurship: Building a Topography Model of Human, Social and Cultural Capital". In Frontiers of Entrepreneurship. Wellesley MA: Babson College.

[27] Kamberidou, I. (2013). "Women entrepreneurs:'we cannot have change unless we have men in the room". Journal of Innovation and Entrepreneurship, Vol. 2 No. 1, pp. 6.

[28] Kantis, H., Ishida, M., and Komori, M. (2002). "Entrepreneurship in emerging economies: the creation and development of new firms in Latin America and East Asia: Summary report".

[29] Kraus, S., Meier, F. and Niemand, T. (2016), "Experimental methods in entrepreneurship research: the status quo", International Journal of Entrepreneurial Behavior & Research, Vol. 22 No. 6, pp. 958-983.

[30] Kronberger, G., Affenzeller, M. (2011). Market basket analysis of retail data: supervised learning approach. In International Conference on Computer Aided Systems Theory (pp. 464-471). Springer, Berlin, Heidelberg.

[31] Kolvereid, L. (1996). "Prediction of employment status choice intentions". Entrepreneurship Theory and practice, Vol. 21 No. 1, pp. 47-58.

[32] Kuckertz, A., and Prochotta, A. (2018). "What's hot in entrepreneurship research 2018?" Hohenheim, Entrepreneurship Research Brief.

[33] Kyrö, P., and Kansikas, J. (2005). "Current state of methodology in entrepreneurship research and some expectations for the future". Entrepreneurship Research in Europe: Outcomes and Perspectives, pp. 121-149.

[34] Lineman, M., Do, Y., Kim, J. Y., and Joo, G. J. (2015). "Talking about climate change and global warming". PloS one, Vol. 10 No. 9.

[35] Liñán, F., and Chen, Y. W. (2009). "Development and cross–cultural application of a specific instrument to measure entrepreneurial intentions". Entrepreneurship theory and practice, Vol. 33, No. 3, pp. 593-617.

[36] Luan, J. (2002). "Data Mining and Knowledge Management in Higher Education-Potential Applications".

[37] Marshall, N. L., and Barnett, R. C. (1993). "Work-family strains and gains among two-earner couples". Journal of Community Psychology, Vol. 21 No. 1, pp. 64-78.

[38] Mathew, V. (2010). "Women entrepreneurship in Middle East: Understanding barriers and use of ICT for entrepreneurship development". International Entrepreneurship and Management Journal, Vol. 6, pp. 163.

[39] McDonald, S., Gan, B. C., Fraser, S. S., Oke, A., and Anderson, A. R. (2015). "A review of research methods in entrepreneurship 1985-2013". International Journal of Entrepreneurial Behavior & Research, Vol. 21 No. 3, pp. 291-315.

[40] McElwee, G., and Atherton, A. (2005). "The case of The International Journal of Entrepreneurship and Innovation Publication trends and patterns in entrepreneurship". Journal of Small Business and Enterprise Development, Vol.12 No. 1, pp. 92-103.

[41] Mesquita, D., Antonialli, F. and Carvalho, D. (2018). "Big Data and Traditional Marketing Research: Comparing methodological approaches" Conference Paper.

[42] Minniti, M., and Naudé, W. (2010). "What Do We Know About the Patterns and Determinants of Female Entrepreneurship Across Countries?" The European Journal of Development Research, Vol. 22 No. 3, pp. 277-293.

[43] Mom, T. J., Van Den Bosch, F. A., and Volberda, H. W. (2007). "Investigating managers' exploration and exploitation activities: The influence of top-down, bottom-up, and horizontal knowledge inflows". Journal of Management Studies, Vol. 44 No. 6, pp. 910-931.

[44] Moreno, Á., Navarro, C., Tench, R., Zerfass, A. (2015). Does social media usage matter? An analysis of online practices and digital media perceptions of communication practitioners in Europe. Public relations review, 41(2), 242-253.

[45] Nielsen Report (2008). https://www.nielsen.com/us/en/insights/article /2008/googles-search-market-share-grows-yahoos-shrinks/ Accessed on: 27 November 2019.

[46] Reese, P., and Aldrich, H. (1995). "Entrepreneurial Networks and Business Performance: A Panel Study of Small and Medium-sized Firms" in the Research Triangle. Birley, S. and Macmillan I. (eds.), International Entrepreneurship. London: Routledge, pp. 124–144.

[47] Sheng, J. and Wang, X. (2017). "A Multidisciplinary Perspective of Big Data in Management Research". International Journal of Production Economics. pp. 191.

[48] Short, J. C., Ketchen Jr, D. J., Combs, J. G., and Ireland, R. D. (2010). "Research methods in entrepreneurship: Opportunities and challenges". Organizational Research Methods, Vol. 13, No. 1, pp. 6-15.

[49] Smith, R., McElwee, G., McDonald, S. and Dodd, S.D., (2013). Qualitative entrepreneurship authorship: antecedents, processes and consequences. International Journal of Entrepreneurial Behavior & Research, Vol. 19 No. 4, pp. 364-386.

[50] Van Egmond, M., Baxter, J., Buchler, S., and Western, M. (2010). "A stalled revolution? Gender role attitudes in Australia, 1986–2005". Journal of Population Research, Vol. 27 No. 3, pp. 147-168.

[51] Woo, J., Xu, Y. (2011). Market basket analysis algorithm with map/reduce of cloud computing. In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

[52] Wu, L., and Brynjolfsson, E. (2009). "The future of prediction: how Google searches foreshadow housing prices and quantities". ICIS 2009 Proceedings, pp. 147.

[53] Xu, Z. M., Zhang, R. (2009). Financial revenue analysis based on association rules mining. In 2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications (PACIIA) (Vol. 1, pp. 220-223). IEEE.