International Journal of Information Management Data Insights 2 (2022) 100093

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Building Textual Fuzzy Interpretive Structural Modeling to Analyze Factors of Student Mobility Based on User Generated Content

Ronak Razavisousan*, Karuna Pande Joshi

Information Systems Department, University of Maryland Baltimore County, Baltimore, USA

A R T I C L E I N F O

Keywords:
Student Mobility
Fuzzy interpretive Structural modeling (FISM)
Fuzzy ISM
Textual Fuzzy Interpretive Structural Modeling (TFISM)
Complex decision-making problem

A B S T R A C T

Many factors influence student mobility across regions and countries. The roles of these factors, along with their interrelationship and interaction, make student mobility a complex decision-making issue. Many textual data generated on social media can answer many open questions about factors affecting human behavior, particularly social mobility. We have developed a novel methodology, called Textual Fuzzy Interpretive Structural Modeling (TFISM), that automatically analyses large textual datasets to identify the internal and external relationships between management or decision-making problems. This computational social science methodology enhances Interpretive Structural Modeling (ISM) approaches to allow the input to be textual data. It is multi-disciplinary and integrates ISM with Artificial Intelligence, Text extraction, and information retrieval techniques. TFISM is a domain-free method, while we have validated this methodology on two different datasets from social media and academic articles. In this paper, we present the results of our study to identify the critical factors and most influential factors for global student mobility.

1. Introduction

Economists and social scientists traditionally study student mobility. Consequently, the motivations and pushing factors for studying this phenomenon are discussed from the economic or sociological point of view. Quantitative analysis and extensive data analysis can enable us to have a new look at this concept. This research focused on the quantitative analysis of social media and examined how social media data can help study student mobility and affect the process. We also tried to overcome studying students' mobility from social media data.

Over half of the world's population is currently online. Mobile users, social media applications, and sensors generate a large volume of online data. Researchers recognize that these "digital traces" present an enormous opportunity to complement traditional migration data sources and improve our knowledge of various migration aspects. Social media platforms capture the geo-location of their users and provide a valuable dataset to study mobility patterns and immigration. Increased rates of human mobility all around the world (United Nations, D. o 2017) push the general population to recognize more geographical destinations than they have in the past. Traveling facilities have accelerated this phenomenon. The internet provides data, knowledge, experiences, and concerns for immigrants traveling to foreign locations. Our motivation for this research is to gather critical insights from social media data to analyze the decisive factors for student mobility. Social Science often uses different data sources for studying behavior patterns. In this study, we have used social media to understand deep processes and complicated decisions in student mobility. The results also demonstrate how Big Data can clarify the hidden angles of social science and help us understand human behavior.

There are three roles involved with the concept of migration, including "origin country," "immigrant," and "host country." These three roles function for student mobility too. The "origin" refers to the location or country where migration starts. The host is the target country that a person chooses as a destination. An immigrant is a person who moves from the origin to the host. This research analyzed the barriers and motivations of student mobility/ migration, known in this field as pulling and pushing factors.

Researchers introduced market forces as one factor for the increasing rate of educated immigrants (Guruz, 2011). Some host countries such as Germany, Sweden, Switzerland, and France have plans to seek and attract more international students. These countries update the process of long-term residency for international students to attract them easily. Hence, highly educated immigrants will join the new population with fewer challenges and conflicts than people with low education. One of the examples of enticing international students is equal tuition fees for them and open pathways to work opportunities (Choudaha, 2017). Eu-

Abbreviations: ISM, Interpretive Structural Modeling; TFISM, Textual Fuzzy Interpretive Structural Modeling; LDA, Latent D Allocationirichlet; MICMAC, Cross-Impact Matrix Multiplication Applied to Classification.

* Corresponding author.
E-mail addresses: Ronak2@UMBC.edu (R. Razavisousan), karuna.joshi@umbc.edu (K.P. Joshi).

https://doi.org/10.1016/j.jjimei.2022.100093
Received 16 January 2022; Received in revised form 16 June 2022; Accepted 16 June 2022

ropean countries projected for 2025 to increase the hosted international students from 125 million to 263 million (Guruz, 2011).

Using social media as a data source to study student mobility has challenges. Considering the growth of immigration and student mobility worldwide pushes us to seek a sustainable data source to study this subject. The primary question to start the project is: Can social media's big data serve as a basis for modeling human migration, specifically international student mobility? Another critical question is, "How can we develop a model to understand the productive factors of student mobility based on user-generated content? And how do the elements of student mobility affect each other?

Exposure to new cultures and providing sociological changes are considered positive effects of accepting immigrants. Therefore, understanding the flow of immigration and forecasting human behavior for mobility is valuable for societies, governments, and all participating parties (Choudaha, 2017).

Our proposed approach is inspired by one of the established decision-making and modeling methods known as Interpretive Structural Modeling (ISM). We customized the ISM method to identify the critical factors for student mobility and extract different related variables to structure them into a comprehensive systematic model. John Warfield was invented in 1979, and it has developed a wide area of decision-making and problem-solving (Warfield, 1979, Warfield and Cárdenas, 1994). It helps impose order and direction on the complexity of relationships among various elements. It is primarily intended as a group learning process, but individuals can also use it (Attri, Dev and Sharma, 2013).

The uniqueness of the ISM is the creation of logical links between elements, which helps form a high-level visual map of the system or problem. The core competency of this method can be summarized into two prospects. The first is simplicity in not requiring advanced mathematical knowledge from the user's viewpoint, and the second is efficiency in terms of computing (MSc, 2000).

The input data for the ISM method is the knowledge of the domain experts. For identifying the interaction between two factors, a person who has a deep understanding and expertise of domain values of the relationships puts a value on the sets of relationships. The main contribution of our work is considering the textual resources for each domain to find the connections between factors. This novel integration enables us to enhance the decision-making process based on noticeable textual data quickly. In this paper, we propose an approach to identify the factors and analyze the relationships and effectiveness of each element.

The rest of the paper is organized as follows: Section 2 covers the literature review. Section 3 describes our methodology in detail. Sections 4 and 5 present the results of our experiments and the validation. Finally, we conclude the paper in Section V and describe some of our ongoing work.

2. Literature Review

A rich list of studies focused on human migration from economics and sociologic point of view. Student mobility can be studied as a mobility category under immigration subject research. Student mobility overlaps with human migration, and both these topics share factors. The definition of migration and legal definition of international students prohibits considering student mobility as a subsection of human migration.

2.1. Human migration and student mobility in economic and social sciences

This category contains all the studies we reviewed on student mobility from economic, sociological, psychological, cultural, and demographical perspectives. Pulling and pushing factors are extracted from this set of articles. All these articles use a specified data source to identify the reason for student mobility and trends. None of these studies used social media data.

Verbik and Lasanowski (Verbik and Lasanowski, 2007) combined the official central government sources and extracted the trend of the student movement. They categorized student mobility based on the country and economic status. Compared to this article, we could not categorize the countries based on the host or origin roles while we extracted a list of highly effective countries. We also identified the power of impact and effectiveness for each country based on other factors. Interestingly our results are close to the former article. Another research project run by Rajika Bhandari (Bhandari, 2017) uses a project atlas as a data source. In this research, the author tried to identify the critical factors of student mobility in higher education. The sample country was the United States, and the results expanded to the world. In another research, Choudaha and Kono (Choudaha, 2012) identified the critical factors of student mobility to guide U.S institutions in recruiting international students. The results from our research have a meaningful closeness to the impact of this research. The same authors have other research to consider economic and political events to highlight the three waves of student mobility over time.

Moreover, demographic changes cause student mobility patterns and influencing factors. These factors include economic, cultural, political, and sociological perspectives. A recent paper on this was published by Guruz (Guruz, 2011).

The main difference between this group of research and our project is the methodological differences. All the studies mentioned in this category used a qualitative research method, while in our study, we applied a quantitative method for analyzing big data. We finally compare the extracted factors by the qualitative approach with the elements extracted from quantitative methods.

In the proposed methodology, we feed the method with the list of reference text articles to extract the relationships between factors. For this dataset, we used articles covering human migrations or student mobility from different perspectives. Table 1 presents the pieces that we used as input data sources. The discussion about choosing these articles presents in the last part of this project (Razavisousan and Joshi, 2019, December).

2.2. List of socio-economics articles used as a data source

The second group of literature we reviewed referred to the studies that used text data to analyze the social factors of migration. Hughes et al. (Hughes et al., 2016) presented a wide range of data sources to study immigration from traditional methods to modern ones. They also provide a list of data sources and their features. We used the feature list of data sources to choose the most appropriate authorities. This research focuses on comparing the differences between traditional and modern methods. In our previous research (Razavisousan and Joshi, 2019, December), we used the results of the conventional method as the ground truth and then compared it with the data collected using modern methods. McGregor et al. (McGregor and Siegel, 2013) studied the feasibility of reviewing migration through social media. They studied the role of social media and its influences on the concept of human migration from four perspectives. Those four groups include 1) the influence of social media on migration, 2) migration integration and the role of social media in this process, 3) immigrant networks and the role of social media, and 4) the role of social media in studying migration (McGregor and Siegel, 2013). Our research is distinct from this work because we explicitly focus on student migration. Moreover, the core concept of their research highlights the role of social media in the field of migration. In contrast, we focus on the migration model and use social media as a data source to identify the power of each factor. The results from our research confirm that social media and UGC can be a data source for studying human behavior and extracting influential factors.

Researchers have also modeled human mobility. Hawelka et al. (Hawelka et al., 2014) leveraged the geo-location of tweets to extract the mobility pattern. They also considered the seasonal timing and community network in this pattern. Another study conducted by Pinto et al. (Llorente Pinto, Garcia-Herranz, Cebrian and Moro, 2015) examined user behavioral activity and social media action to justify migration be-

Table 1
List of academic articles on socio-economics used as a data source.

	Article	Author	Year
1	International Migration Report 2017 (United Nations, D. o 2017)	UnitedNations	2017
2	Website: Big data, migration and human mobility(migrationdataportal.org/themes/big-data-migration-and-human-mobility)	MIGRATION DATA PORTAL-The bigger picture	2019
3	The immigration debate: studies on the economic demographic and fiscal effects of immigration (Isbister, 1998)	Isbister, John	1998
4	Globalization and the inward flow of immigrants: Issues associated with the inpatriation of global managers (Harvey, Kiessling and Moeller, 2011)	M. Harvey, T. Kiessling, and M. Moelle	2011
5	A universal model for mobility and migration patterns (Simini et al., 2012)	F. Simini, M. C. González, A. Maritan, and A. L. Barabási	2012
6	Migration: Ravenstein Haythornthwaite and beyond (Tobler, 1995)	W. Tobler	1995
7	International opportunities: searching for the meaning of student migration (Findlay, Stam, King and Ruiz-Gelices, 2005)	A. M. Findlay, A. Stam, R. King, and E. R. Gelices	2005
8	Social media and migration: Virtual community 2.0 (Komito, 2011)	L. Komito	2011
9	Determinants of the international mobility of students (Beine, Noël and Ragot, 2014)	M Beine, R Noël, and L Ragot	2014
10	A gravity model analysis of international migration to North America (Karemera, Oguledo and Davis, 2000)	D. Karemera, V. I. Oguledo, and B. Davis	2000
11	Social media and migration research (McGregor and Siegel, 2013)	E McGregor, and M Siegel	2013
12	Migration of professionals to U.S. (State et al., 2014)	B. State, M. Rodriguez, D. Helbing, and E. Zagheni	2014
13	A world on the move: Trends in global student mobility. (Bhandari, 2017)	R. Bhandari	2017
14	Introduction to Gravity Models of Migration & Trade (Crymble, 2019)	A. Crymbl	2019
15	The gravity model of migration: the successful comeback of an aging superstar in regional science (Poot et al., 2016)	J. Poot, O. Alimi, M. P. Cameron, and D. C. Maré	2016
16	Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone Social Media and other Big Data: Feasibility study on Inferring (labor) mobility and migration in the European Union from big data and social media data (Hughes et al., 2016)	C. Hughes, E. Zagheni, G. J. Abel, A. Wisniowski, A. Sorichetta,I. Weber, and A. J. Tatem	2016

tween cities. They also considered the tweets' geo-locations. Geo-located social media activity, such as posts on Twitter and LinkedIn, has been used to infer international migration flows (Hawelka et al., 2014). The main difference between these previous approaches and ours is the type of data we analyzed. These studies have accessed users' geo-location and identified the location changes in their approach while we extract the pulling and pushing factors for movement based on the social media content shared by users.

Grover and Kar (Grover and Kar, 2020) extracted a practical solution for promotional marketing-based users' social media activities. They built a Twitter engagement model for understanding user dynamics. Hong and Davison (Hong and Davison, 2010, July) studied topic modeling for microblogging datasets such as Twitter. McCallum et al. (McCallum, Wang and Mohanty, 2006, June) proposed a model to identify groups and topics in the text. Mikolov et al. (Mikolov, Chen, Corrado and Dean, 2013) worked on the approaches to find a similarity between words. They proposed a model to compute a vector for each word. Maas et al. (Maas et al., 2011, June) built a model that uses unsupervised and supervised techniques to learn word vectors that capture semantic terms. Our novel approach differs from these studies as it combines topic modeling, clustering, and word embedding to determine the closeness of topics from different data sources. The main contribution of our work is to study student mobility based on analyzing UGC on Twitter and model the extracted factors by our developed method of textual fuzzy interpretive structural modeling (TFISM). We will explain the details of the proposed TFISM method in the rest of this article.

2.3. Interpretive Structural Modeling (ISM)- Newmodeling for computational social science

ISM is an established methodology to identify the relationship between variables in decision-making or complex problem-solving. ISM was invented by Warfield (Warfield, 1979, Warfield and Cárdenas, 1994) for "interactive management" and "structural approach" to system design. ISM has an enhanced version called Fuzzy-ISM (RAGADE, 1976)which Ragade mentioned for the first time. Fuzzy-ISM uses Fuzzy logic, which plays a significant role in dealing with vagueness and uncertainty in human language and thoughts in decision making. Integrating individual opinions, experiences, ideas, and motivations become essential to translating linguistic judgments into fuzzy numbers (Zadeh, Klir and Yuan, 1996). The integration of ISM with fuzzy sets allows decision-makers to understand further the level of influence of one criterion over another, which was earlier present only in the form of binary (0,1) numbers (Khatwani, Singh, Trivedi and Chauhan, 2015). We extended Fuzzy-ISM by enabling textual data as input.

ISM is an established and known methodology used in various fields. There are multiple extensions to the ISM method to adjust it for various domains. Fuzzy ISM is one of the extensions that inspired us to implement the ISM model based on textual data.

Fuzzy theory plays a significant role in dealing with vagueness and uncertainty in human language and decision-making. Human sentences for decision-making are influenced by their past knowledge and experiences, and often their estimations are articulated in uncertain linguistic terms. However, integrating individual experts' various opinions, experiences, ideas, and motivations become important to translate linguistic judgments into fuzzy numbers. (Zadeh, Klir and Yuan, 1996).

The integration of ISM with fuzzy sets provides flexibility to decision-makers to further understand the level of influence of one criterion over another, which was earlier present only in the form of binary (0,1) numbers (Khatwani, Singh, Trivedi and Chauhan, 2015). 0 represents no influence, and 1 illustrates influence. Due to this, the decision-maker is left with only the option of saying 0 or 1 irrespective of the level of influence, whether low, high, or very high. The proposed Fuzzy ISM approach takes care of this issue and gives more comprehensive flexibility to express the level of influence using fuzzy numbers. Fuzzy ISM is a better match for our purpose. In this approach, instead of a binary point of view to identify the relationship between variables, the connections are ranked based on a range of numeric values between 0 and 1, such as 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 1.

ISM and its extensions are used in various domains. For instance, Jiang et al. (Jiang et al., 2018) use ISM and fuzzy analytic network processing to identify risk factors and calculate risk for Artic shipping strategic alliance. They can highlight the risk factors that should be focused on. Another research used the ISM method to identify a sustainable production system's critical success factor (CSF) (Kota et al., 2021). Moreover, the ISM is applied for analyzing pr-vacuation time under the building in a fire situation by Yixue Liu et al. (Jiang et al., 2018). They extracted the critical factors that should be considered for a fire emergency. In another research run by RR Menon et al., the barriers to supply chain management in the electronic industry were studied with the help of the ISM (Menon and Ravi, 2021). Due to the recent attention to the cybersecurity domain for organizations, Rajan et al. identify the parameters that affect cybersecurity within an organization and analyze relationships among these factors (Rajan et al., 2021). They modify the ISM to a new approach called Modified total Interpretive Structural modeling (M-TFISM). Shrimali (Shrimali, 2019) applied the ISM model to the manufacturing domain to extract factors that affect the successful implementation of lean manufacturing; she also identified the interaction between elements. For the construction domain, Sandbhor and Botre (Sandbhor and Botre, 2014) used TISM to study affecting factors on labor productivity to enhance the productivity of the construction projects and improve associating factors. Another domain addressed by the ISM method is seismology. Ahmad, Mahmood, et al. (Ahmad, Tang, Qiu and Ahmad, 2019)used the ISM method to highlight the soil factors and the relationships with earthquakes. Besides, e-commerce took advantage of Fuzzy ISM. Valmohammadi and Dashti (Valmohammadi and Dashti, 2016) applied FISM to identify and highlight inherent interactions among these barriers to e-commerce. For the technology domain, the ISM integrate by NPD New Product Development (NPD) to evaluate various technologies for the new product (Lee, Kang and Chang, 2011).

2.4. Text analysis for decision-making and management

Another aspect of TFISM relates to analyzing big text data to build a vision for managers or policymakers. An extensive range of Natural Language Processing (NLP) techniques has been applied to various domains comparable to TFISM. We have reviewed the most recent research and identified the significance of TFISM. For instance, Kumar et al. (Kumar, Kar and Vigneswara Ilavarasan, 2021) comprehensively reviewed text mining in management services. They recognize the study focusing on text mining or NLP in different domains and review their methodology, text mining techniques, and visualization tools. Another review appears with the systematic analysis methodology to analyze the management disciplines and the help of text mining to support it from different domains (Kushwaha, Kar and Dwivedi, 2021). These two studies reveal that no process advanced the ISM method by leveraging text mining techniques. Jamal Abdul Nasir and his team (Nasir, Khan and Varlamis, 2021) ran another study in the field of text analysis for social media. They invented a hybrid deep learning method combining a convolutional and recurrent neural network. This novel approach can detect fake news with significant performance. Another research that has a comparable policy to our proposed system is run by JisooAhn et al. (Ahn, Son and Chung, 2021). They used topic modeling techniques to analyze the collected tweets. We also have the same approach when the input text comes from social media and don't have predefined variables from previous ISM studies.

3. Methodology

By looking deep at the motivation for student mobility, we realized that international students have similar encouragement as people who migrate. Fig. 1 presents the extracted critical factors for humans migrating based on economic and sociological research. Students consider the same characteristics regarding traveling overseas, but their priorities and the weight of elements are varied. In Fig. 1, items on the left side push the students to leave their homes. These factors are generally negative. Conversely, factors that attract students to the destinations are present on the right side. Most of the elements from destinations are positive, except the distance, which is a negative factor for immigrants and international students.

All these factors can affect choosing the destination. Suppose selecting a goal is a decision that students have to make. We can consider all the variables in Fig. 1 as the factors that influence decision-making processes and their results. In other words, if any of these factors change, it can push the students to choose another destination for their education. For example, when a country provides a job opportunity for graduated people, it attracts more people to study in the country. However, if the quality of education gets poor, many international students will eliminate this country from the destination list. Therefore we can look at student mobility as a decision-making process. In this research, we propose a solution to automate aspects of this decision-making problem. It is essential to be aware that deciding for the student is not our target in this research. At the same time, we want to understand the process of making decisions for students because it can be helpful for the governments, policymakers, and universities to reach appropriate thair students and catch their attention. The proposed solution can benefit those who try to influence the decisions or understand the role of factors in the final results.

3.1. New modeling for Computational Social Science

This paper tries to identify the relationship between factors with the help of well-established modeling approaches. As is shown in Fig. 1, several factors affect human migration. These factors are not significantly different for student mobility, while their weight for decision-making can be different. Interpretive Structural Modeling (ISM) is a methodology to identify the relationship between variables in the decision-making process or complex problem-solving. The designed method for process flow is presented in Fig. 2. We customized the methodology for modeling social media data. It is necessary to mention that ISM has an enhancement version called Fuzzy ISM. In our proposed approach, we implement Fuzzy ISM for using textual data from social media and automate this process.

The primary adaptation in the original ISM version that helps us solve the problem happens in two first steps. These two steps with new updates have been explained. Having access to the knowledge of experts for a specific domain is the critical principle of ISM methodology. Without the specialist's knowledge, the first step of ISM methodology will not be complete. Moreover, finding an expert in the domain who is familiar with the concept of ISM will make the process harder. We developed an enhanced version of ISM that uses text data as an input to analyze the problem, and we can have the same type of results. Based on what we are looking for, we can select the input textual data. We must have appropriate data that contains information and knowledge related to the problem. We can choose a list of scholarly articles related to the problem for this aim.

Moreover, When we are interested in general opinion about our question, a collection of UGC from social media can be the input data for the system. Based on what we are looking for, we can select the input textual data. We have more detail about the data source in section 3.4 and 3.5.

The main contribution of our research is customizing the process in a way that helps to build the ISM matrices automatically based on text data and without the knowledge of experts. We also adjust the process to have a meaningful analysis of the results. We consider two different data sources, including social media and academic articles. We also changed the approach to analyze data based on the frequency of the factors in text data. Details about the new method for analyzing the results and building the plots are presented in section 3.5.

Our proposed approach (TFISM) implements a process flow to analyze text data, extract the variables, and fille the matrices. TFISM is im-

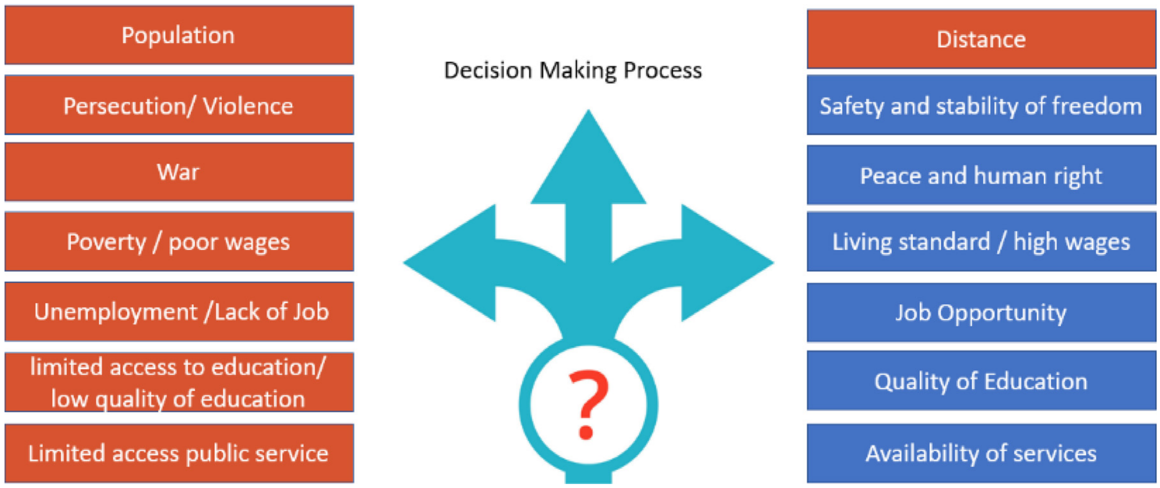

Fig. 1. Pulling factor and pushing factors of student mobility. The red boxes show the pushing factors from the origin countries, and the blue boxes show the pulling factors from the host countries. The students decide about education and destination countries based on these factors.

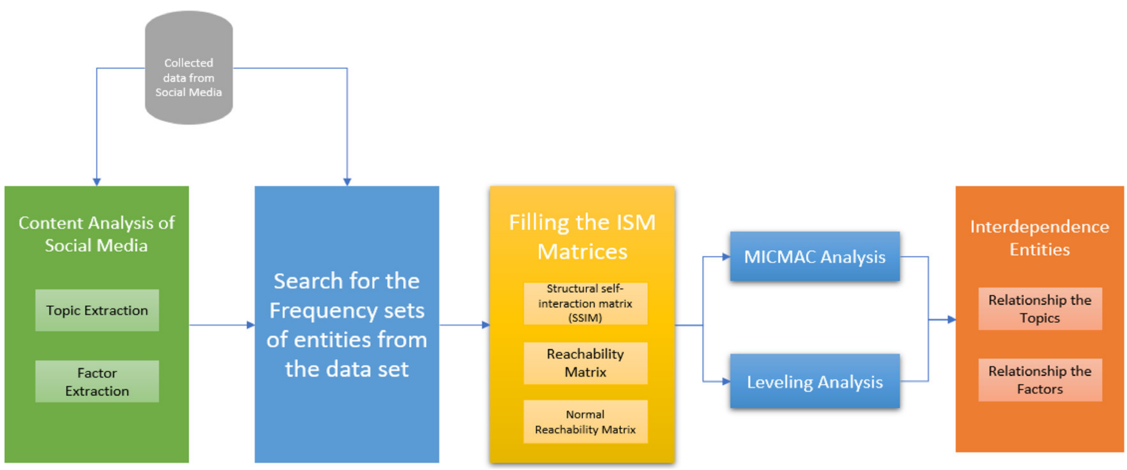

Fig. 2. Methodology of implementing ISM and Process Flow.

Table 2
Keyword list for student mobility.

Immigration	f2visa	GRE
Visa	Student migration	Student green card
Student visa	Foreign student	U.S. university
Educated immigrant	International education	Student work permit
Graduate student immigrant	International student	OPT
f1visa	TOEFL	CPT

plemented based on the frequency of factors in text data. We have a list of factors that are involved in the problem. We search for these factors in each unit of text data and find the frequency of each set of two factors. In other words, the Structural Self-Interaction Matrix (SSIM) and reachability matrix in the proposed approach evolve based on social media or academic articles (extracted from big data). The process flow of creating SSIM based on the type of data has some changes. The process flow of implementing TFISM for two data types explains separately, and the quality of results and the method's validity will discuss in the validation part.

3.2. Data Collection

Previously explained that input text data plays an essential role in TFISM methodology. In other words, we want to examine how text data can cover the role of expert knowledge in the ISM method. The deployment of two data sets gave us a better view of what we can reach from TFISM. These two datasets include: 1) a list of academic articles focused on student mobility and human migration. All these articles came from the sociology and economy sections. This list carries sixteen pieces of published research and reports which are presented in Tables 1, 2) Social media data is a collection of tweets related to student mobility that contains more than 7 million tweets. All of These tweets contain at least one of the keywords related to student mobility. Creating the keyword list, collecting articles, and building the datasets refers to another research run by the same authors (Razavisousan and Joshi, 2019, December).

As we mentioned earlier, we chose Twitter as a social media data source because of its accessibility, popularity, and number of active users. Moreover, Twitter has various ways to extract data to study users' mobility (Hughes et al., 2016). Twitter APIs are used for collecting data. Our sample data shows that people who talk about migration on social

Each party in the migration process has its concerns; in the table, the concerns of migration roles are present separately.

	Host	Origin	Immigrant
1	Skills and quality of education	Skills and quality of education	Skills and quality of education
2	Job opportunities and a better life situation	Cost of migration (living cost, distance, language, and network.)	Cost of migration (living cost, distance, language, and network.)
3	Globalization and ease of migration and access to the education	The fact of common host list countries	immigrant background (Gender-race, Socio-economic situation, language component, personality)
4	Mobility Culture	Fact: rate of returning international student	Unemployment rate
5	Educated labor	Job opportunities and a better life situation	Job opportunities and a better life situation
6	Population	Globalization and ease of migration and access to the education	Globalization and ease of migration and access to the education
7	Brain Gain	Facilitate immigration	Family and financial history
8	Economic factor (distance and GDP)	Mobility Culture	facilitate immigration
9	stress level	Educated labor	Mobility Culture
10	Unemployment rate	Job hunt by immigrant	Educated labor
11	Fact: rate of returning international student	Living Cost	Job hunt by immigrant
12		Population	Life skills and personal experiences
13		Rules and policies	Rules and policies
14		Brain gain	education cost
15		Colonial power	Economic Factor (distance and GDP)
16		Economic Factor (distance and GDP)	Student financial support
17		Financial benefit	Students' motivation and background
18		Students and education providers background	The success of the previous generation
19			A transformation that students faced

media have a vast range of points of view. To narrow the research scope, we limit the topics to student mobility and student migration.

A list of related keywords is needed for data collection. To compile a comprehensive list of keywords, we conducted interviews with international students and staff in the University of Maryland, Baltimore County international student office. The results of the interviews were categorized as:

1 Common questions topics:
   - Work permission (Permit) for post-graduation
   - Tend to stay or find a way to stay
   - CPT- Duration – extension- employer
   - Permanent resident (PR) – green card
   - OPT – STEM OPT – traveling
   - Carrier
   - Internships – coups
   - International faculty _ EB1 / EB2
   - Passive working – other sources of income

2 Visa-related questions:
   - Embassy for an interview– tips for the interview- get a visa in a first place
   - Visiting home – returning visa- traveling (vacation – conference)- documents- safety

3 Updating the immigration record:
   - Updating /Changing position / employer

4 Applicant questions:
   - University ranking
   - location: diversity in the area/ Job hob
   - living cost
   - safety

5 Countries where the process of getting a visa may take longer=> 7 countries in Muslim Ban and specific people of Venezuela.

Table 2 presents the keywords extracted from the interviews and other resources. The data collection process runs based on the keyword list, and we collect all tweets with any of the keywords. Tweeter API is used for data collection, which was done from Nov 2018 to Feb 2019 (four months). We collect almost 10 million tweets.

3.3. Data cleansing

One of the essential steps of data preparation is cleansing the data before the analysis process. For data cleansing, we performed these steps: 1) URL removal, 2) retweets removal, 3) stop words removal, 4) steaming, 5)tokenizing and de-tokenizing.

Although the contents of URLs that people share for each topic have valuable information about the users' concerns, in this work, we decided to remove all URLs and focus on the users' generated contents. For the same reason, we remove the retweet posts because the duplication of the posts is not a focus of this study, and we focus only on the content of the tweets. The original tweets were saved, but the retweeted ones were removed.

Stop words removal and steaming are primary steps for text analysis to omit the words that do not add semantically and turn the words into their roots. Tokenizing is a common task in text mining to break strings into words. It is also necessary for our cleaning process to prepare the input data for LDA (next step). After running the LDA, there were some topics in our results in which the de-tokenization was required. In fact, some phrases or bonded words (that should come together) are broken with the tokenizing process, and we must take them back together with the process of de-tokenizing.

Finding appropriate data that guided us is significant; however, this data is not accessible to reach. Using an interview can be an option, but we think we can talk to the student who has already made their decision and relocated to the host. Using interview data may cause students to support their decision and highlight the positive variables. Additionally, Face to face Interviews may cause dishonesty because people are not comfortable talking about their limitations and genuine concern. Consequently, we decided to use data from the platform where people speak freely and anonymously share their true thoughts—filtering valuable data between the growing data generated online called for a comprehensive list of filters. The details for reaching data for this research are presented here (Razavisousan and Joshi, 2019, December), and considering all the limitations; this is the most appropriate data that we can reach for this research because this is from the target populations, and this is what they worry about with minimum consideration.

After preparing the data, we must consider the computational section that uses collected text data as input for the proposed method.

3.4. Process flow of using TFISM in data modeling of textual data

ISM method has a process flow to run. However, we must add new sections for accepting text data and analyzing text to extract information and adjust the process to build results based on further details. The updated diagram of TFISM for collected data is presented in Fig. 3, and

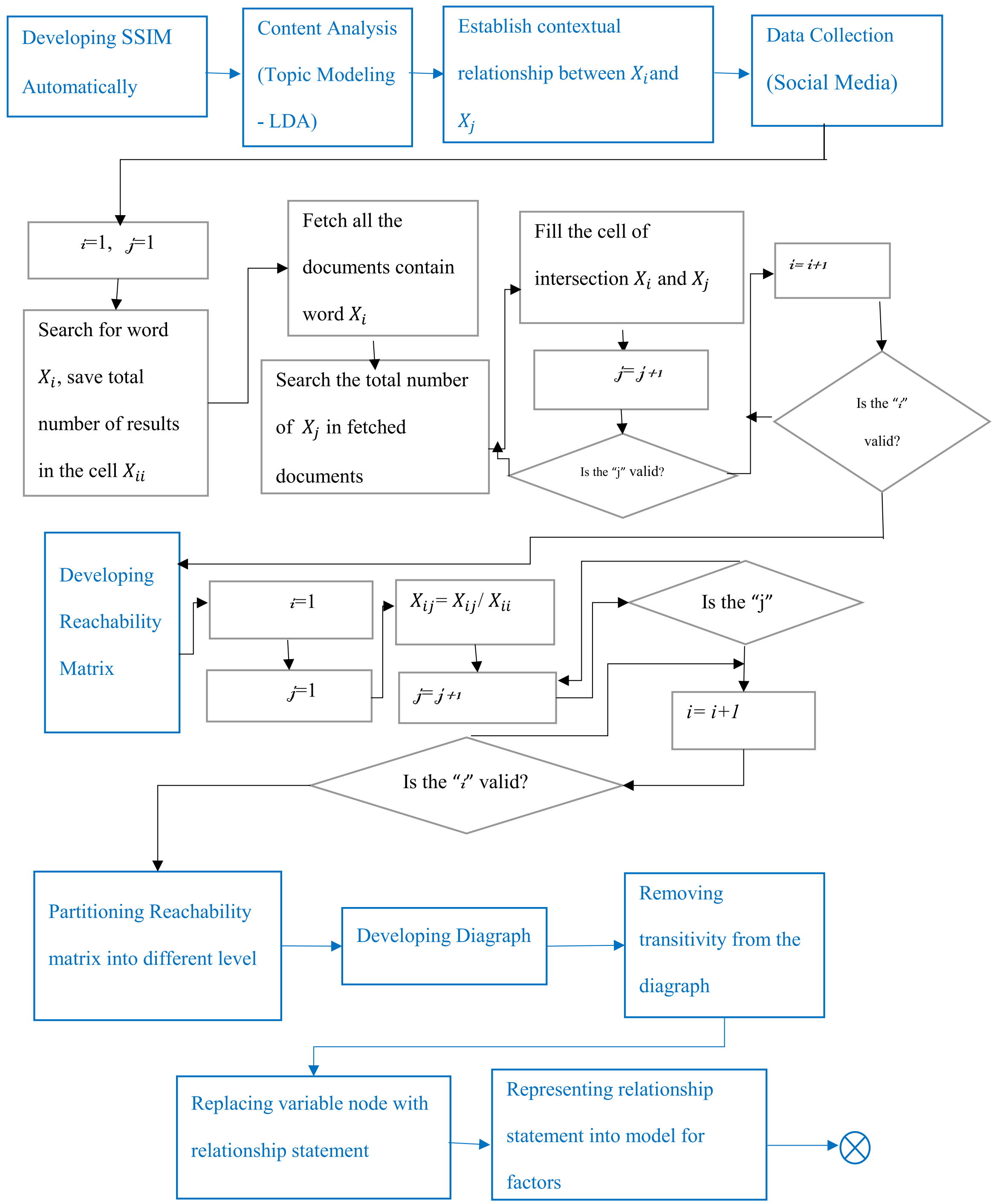

Fig. 3. Process flow of TFISM for textual data- The blue shapes are the original process of ISM and the gry ones are customized for TFISM.

for simplifying the process explaining the steps are designed like the Warfield-ISM version. Gray-colored items in Fig. 3 are the parts intended for TFISM implementation. These parts are the main contribution of the research. The blue items in Fig. 3 represent the initial steps of the ISM method, and the black objects are the ones that are added to enhance the methodology.

The enhancement is built based on two concepts. First is the data unit, which means the data size for each process cycle that can be varied from sentences and paragraphs to the whole document. The second concept is frequency. Frequency means the number of the appearance of two variables in a data unit.

3.5. How to extract driving power and dependency?

The main burden is interpreting concepts of driving power and dependency from the frequency of factors. The simple idea extracted from prior versions of ISM is the definition and relationship between driving power and dependency. In other words, the relationship of variable i with other variables identifies driving power and dependency. Fig. 4. shows the association between these two concepts.

As demonstrated in Fig. 4. dependency of i to j defines as a dependency, and dependency j to i, illustrates the driving power of i and j. Therefore, we can define dependency and driving power for a single

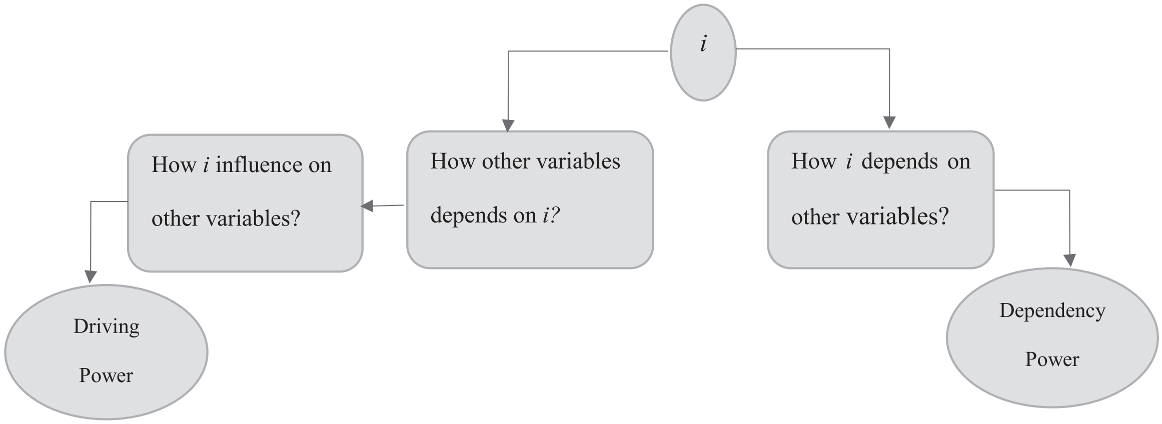

Fig. 4. Variables associations - Driving Power and Dependency.

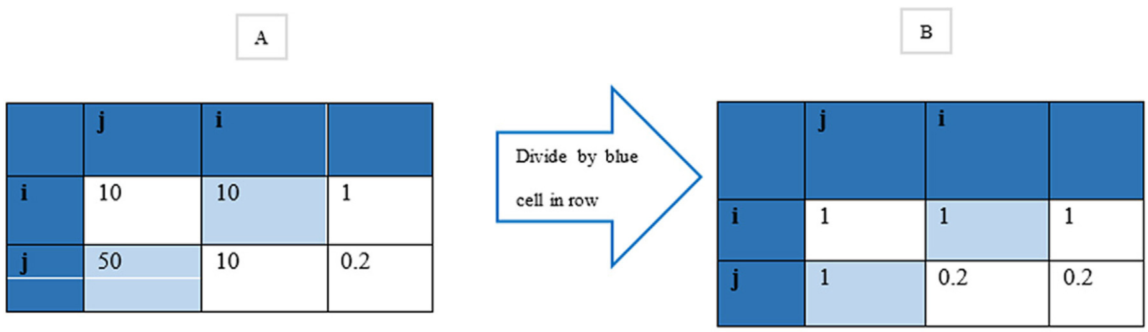

Fig. 5. Sample for explaining the concept of Dependency and Driving Power.

variable based on the dependency of the index variable to others and the dependency of other variables on the index variable. With the help of this explanation, we need to define the two-way dependency of the variable. Two-way dependency means how i depend on j and how j depends on i. In the presented case, finding the dependency of two variables from the appearances in the text is the goal. The proposed solution is based on the simultaneous appearance of two variables in the collected data. We know that the number of times i appear with j is equal to the number of j appears with i; this number calls by X. The total number that i is shown up is named I, and the total number that j is coming up is J.

Based on our definition, X dividing by I; (X/I) presents dependency j to i, which can be conveyed as the driving power of i. Moreover, the result of X dividing by J; (X/J) shows the dependency i to j., which means the dependency of i. Fig. 5 presents a hypothetical example with only two variables, i and j. The value for I is ten, and J is 50. The number of appearances of i and j together is 10. It means from 10 times i show up always is followed by j. It supports the driving power of i. In this case, driving power for i calculated 1, the maximum driving power rate. Since the 50 times, j appears only ten times comes with i, the dependency of i is 0.2. Consequently, the value in rows shows the driving power, and the values in columns interpret as a dependency power.

4. Experiment

Section 4 presents the proposed TFISM applied to the collected data in section 3.3. Principle steps for building TFISM are explained, and data adjustment parameters are present.

4.1. Step 1: Create Structural Self-Interaction Matrix (SSIM) based on the extracted topic from social media

Automatically extracting the variable is one of the ultimate goals in this approach; for this purpose, we decide not to apply the opinion of experts in student mobility and enhance the method by extracting the critical factors of content from social media. Using the extracted topic from topic modeling techniques is our solution because we do not have a list of influential factors. Consequently, we will reach a list of topics and use them as variables to build the SSIM. In this process, we take advantage of the Latinate Dirichlet Allocation (LDA) technique for extracting a list of topics, details about student mobility, and extracting a list of factors; please review this research (Razavisousan and Joshi, 2019, December).

For running LDA, two different settings were tried. First, having a high number of clusters (in this case 8) and a smaller number of topics in each cluster (in this case 5); for the second set, having a low number cluster (in this case 2) and a high number of topics in each cluster (in this case 20). Prototyping the methods with sample data shows that the second set provides better results.

4.2. Step 2: Automatically Fill the SSIM based on social media data

An essential enhancement in implementing the TFISM methodology for social media data is filling the SSIM by an automated process. In other words, we get the advantage of general knowledge of a large population in social media when we don't have access to the experts. The frequency of appearance-specific topics considered for identifying a relationship between two topics of i and j. The number of times that topic i and j come in each reference document will be put in the cross-section cell of i and j in SSIM.; in Fig. 6, the highlighted cell, which has the value of 2, shows the number of times "policy" appears when "issue" is already showing up. For filling the ISSM, the topic in each row considers the main topic, then the topic's appearance in columns will enumerate. Besides, the values in gray cells show the total number that the rows' topic occurred. For explanation, the topic "issue" was repeated six times in all sample tweets, and it comes with a policy two times. It is not hard to find out that the SSIM is a diagonal matrix because display two words in

		schools	family	Internationalstudent	History	education	University	work	Canada	International	policy	Gvernment	created	week	around	issues	House	Speaker	UK	Laredo	border	Nancy Pelosi	security	discusses	TAMIU	US
		25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
US	1	0	0	0	0	2	4	2	0	0	0	4	4	0	2	1	0	0	2	0	0	0	0	0	0	15
TAMIU	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	8	8	0
discusses	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	10	8	0
security	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	8	8	0
Nancy Pelosi	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	8	8	0
border	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	8	8	0
Laredo	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	0	8	8	8	8	8	8	0
UK	8	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	2
Speaker	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	10	0	8	8	8	8	8	8	0
House	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	6	0	6	6	6	6	6	6	0
issues	11	3	0	0	0	0	0	0	0	0	2	0	0	0	2	6	0	0	0	0	0	0	0	0	0	1
around	12	2	0	0	0	0	0	2	0	0	2	0	0	0	8	2	0	0	0	0	0	0	0	0	0	2
week	13	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
created	14	0	0	0	1	1	4	1	0	0	0	4	6	0	0	0	0	0	0	0	0	0	0	0	0	4
government	15	0	0	0	0	1	4	0	0	0	0	5	4	0	0	0	0	0	0	0	0	0	0	0	0	4
policy	16	3	0	0	0	0	0	1	0	0	5	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0
International	17	1	0	3	0	1	1	0	3	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Canada	18	0	0	2	0	2	0	0	6	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
work	19	1	2	0	0	0	0	12	0	0	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	2
University	20	0	0	0	0	1	10	0	0	1	0	4	4	0	0	0	0	0	3	0	0	0	0	0	0	4
education	21	0	0	1	1	13	1	0	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
History	22	0	0	0	4	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
internationalstudent	23	0	0	3	0	1	0	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
family	24	0	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
schools	25	11	0	0	0	0	0	1	0	1	3	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0

Fig. 6. sample of fully filled SSIM by TFISM for social media.

the dataset are not associated with the column and row; the appearance of "issue" with "policy" is similar to the arrival of "policy" with "issue." It is good to know that the SSIM in the original version also has the same feature.

In the preliminary version of the ISM method for having a fully filled SSIM, we should copy the value of the filled cells symmetrically based on the diameter then update the matrix with and th the adjacency relationships that experts decide about it. Every value in the SSIM (V, A, O, X) shows the direction of the relationship; consequently, when we have one side of the association, we can use the duplication to find the other side. However, TFISM considers both directions of associations. For the same factors are mentioned above, M computes "issue" with the appearance of "policy" and puts it in the same cell that we highlighted. Then for having the other direction (relationship of "policy" and "issue") needs that TFISM calculates the appearance of "issue" in the documents when "policy" is already appeared and puts the mirror cell. The values in these two cells can be similar, but also they can be different. The main difference between ISSM built based on the TFISM method is that all the cells' values are calculated, and we cannot use the mirroring option.

Fig. 6 shows the final SSIM matrix reloaded with sample Twitter data. The significant difference between ISM and TFISM is in the type of cells' values. Formerly, the SSIM contains the value of (V, A, X, and O), which converted to 0 and 1 (0,1) and for fuzzy extension converted to the rational number (Q) between 0 and 1 [0,1], while in this approach, SSIM can be filed with any natural numbers (N) as presented in Fig. 6.

4.3. Step 3: Building the reachability matrix and normalizing the reachability matrix

The reachability matrix in the ISM method is a numeric matrix that shows the dependency and driving power between i and j variables. Each value shows how i depend on j and how i can drive j. For extracting this concept from the numeric data, the values in each row divided by the total number of topic appearances have shown in dark-gray cells in Fig. 6. For instance, number 2 (highlights for "issue" and "policy") will be divided by 6.

5. Result and Validations

We implemented a series of updates and changes in the original version of ISM to customize the process for social media data. Based on these adjustments, we designed the validation methodology and used other validation techniques to evaluate the reliability of TFISM. The following three types of validation are considered, and for each type of validation, our possible methods are mentioned below.

1) Factor coverage validation
   a. Comparing source of factors (Comparing expert opinion and social media)
   b. It starts from a large number of factors and Narrows down to an appropriate list
   c. Using reference text data (Academic Articles)
2) Method validation
   a. Compare TFISM matrices with the experts' opinion
   b. Identified relationship validated by association rule mining
   c. Validating adjacency relationship by a search for sets of three factors frequency
3) Results validation
   a. Using completed research-based ISM results as ground truth and all related references and articles have used it as a reference for TFISM input, then comparing the results.
   b. Expert opinion needs for validating from specific domains.
   c. Ground truth and test data

The high-level view of the proposed method is present in Fig. 7. The methods we used to analyze the student mobility dataset will be discussed in detail.

5.1. Expert Opinion Validation

There are a group of people that we contacted directly and indirectly.

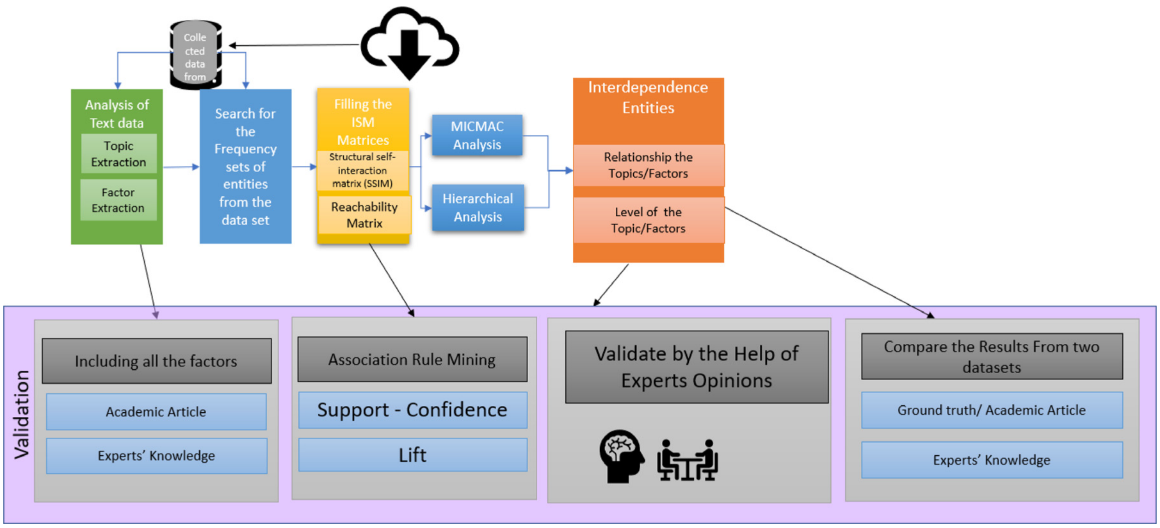

Fig. 7. Validation Methodology.

Directly contact: First, we talked to an officer of the international students at two universities (University of Maryland Baltimore County and Towson University). They are experts in accepting international students, their visa application and requirements, their concerns before and during their educations, and their possible options after graduation. We have interviews and contact directly via email. These interviews helped us have broad categories with the list of questions and genuine concerns of people in the process of mobility as students. After finalizing the keyword list for data collection, these officers also check and confirm this list.

We also attended to UMBC RESEARCH FORUM with the subject of IMMIGRATION AND MOBILITY IN HIGHER EDUCATION on 12 April 2019. This gathering gives us a chance to share our thought and achievement with the domain experts and have their suggestions and opinions during the discussion.

Indirectly, we get the advantage of expert publications for having their opinions about the research question. All the academic articles from socio-economic science are part of our attempt to understand the factors and their relationships in sections 2.1 and 2.2.

5.2. Association Rule mining

In this case, the possible relationship between sets of two topics is validated. The frequency of two words is searched in the whole Twitter data sets with 7 million tweets. With the help of the frequency of the topics, Fig. 8 can present the support values for each set of words or for every single topic in diameter cells which are highlighted by green color. The value for lift is calculated for topics and presented in Fig. 9, and all the values greater than one are shown by red color.

To calculate each factor's support, first, we count the number of the appearance of each set of two topics in the dataset. Each row contains the repetition of the topic in the row with each topic in the column. The value in the diameter cells demonstrates the summation of all the values in the rows. Each value in the diameter cells is divided by the total number of topic sets and gives us the support values (total number of topic sets reached by the summations of all the diameters values). Topics with the highest support are the "border," "illegal," "law," "OPT," and "Visa."

The confidence and lift were calculated with the same logic for the topics. The red cells in Fig. 9 are the topics with a lift greater than 1. The topics which have the highest lift value in the row are "The U.S.," "Trump," "border," "illegal," "international student," and "University."

The association rule mining results show the expected values, and the relations between factors are close to the previous results. We can compare these results with different methods' outcomes in the other part of the validation phase.

5.3. Validating phase by comparing social media and academic articles

For this comparison, in addition to our collected dataset from Twitter, we use sixteen scholarly articles that mainly focus on student mobility over the world. The list of these articles comes from another research. (Razavisousan and Joshi, 2019, December).

5.3.1. TFISM applied to the related academic articles

We must use the same variables and indicators to compare two sets of results. We use the topics from Twitter and articles in previous research, then find the similarities between these two sets of topics. At this point, we use a unique set of topics and run TFISM based on the frequency of these topics in two datasets. The topics are the most frequent topics from Twitter and articles. The topic list contains 76 topics that are presented in Table 3.

Fig. 10 presents the topic distributions for the articles, and Fig. 11 presents the topic distribution for the tweets. These plots are built based on MICMAC analysis. As shown in Figs. 10 and 11 considerable number of topics have a low dependency and driving power, so they are located at the left-bottom of the plots. In Fig. 10, "foreign", "student", "US" and "work" are the outliers, and in Fig. 11, "foreign", "Family", and "student" are the outliers. MICMAC analysis outliers indicated the variables with high dependency, driving, or both. It is not surprising that the words "foreign" "student" is in the outliers' range in both articles and tweets dataset.

For having a clear view of the topics' dependency and driving power, we remove the outliers from the plot, and the rest of the topics are shown in broader space: f and Fig. 12 present topic distributions without the outliers. In Fig. 13, both data are presented in one plot that gives us a view of distributions of topics from two datasets. The blue dots present Twitter's topic, and the orange dots are presented with articles.

73 topics which have non-zero value are categorized into 13 categories including "Abroad"," Background"," Countries", "Demographics", "Economic", "Financial", "Globalization", "Host", "Job", "Network", "Origin", and "Political". Table 4 shows all the topics and the categories they assigned.

When we run MICMAC analysis for categorized data, the results can examine easier. Fig. 14 presents the categorized data from both Twitter and article datasets. The value of the driving power and dependency power of the categorized variables is very close to each other (The distance between the same topic from two datasets is smaller than 5) and is

Suport	T=206	visa	cpt	opt	school	Pathway	family	student	History	education	University	work	World	Canada	International	policy	government	law	week	sexual	cost	security	House Speaker	UK	illegal	border	financial	global	Trump	US
		29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
US	1	0.003	0.000	0.001	0.000	0.000	0.000	0.003	0.000	0.001	0.006	0.005	0.001	0.001	0.001	0.002	0.002	0.002	0.000	0.000	0.002	0.000	0.000	0.001	0.005	0.005	0.000	0.002	0.013	0.102
Trump	2	0.003	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.012	0.001	0.000	0.000	0.004	0.001	0.002	0.001	0.000	0.009	0.001	0.001	0.000	0.016	0.013	0.000	0.002	0.153	
global	3	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.012		
financial	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005			
border	5	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.012	0.002	0.000	0.000	0.009	0.002	0.000	0.007	0.077				
illegal	6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.008	0.001	0.002	0.000	0.073					
UK	7	0.001	0.000	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.003	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.033						
House Spe	8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011							
security	9	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.013								
cost	10	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.025									
sexual	11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.007										
week	12	0.003	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.017											
law	13	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.001	0.068												
governme	14	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.015													
policy	15	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.018														
Internatio	16	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.001	0.000	0.000	0.000	0.000	0.013															
Canada	17	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.023																
World	18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006																	
work	19	0.003	0.000	0.002	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.071																		
University	20	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.040																			
education	21	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.008																				
History	22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001																					
student	23	0.001	0.000	0.000	0.001	0.000	0.000	0.030																						
family	24	0.000	0.000	0.000	0.000	0.000	0.013																							
Pathway	25	0.000	0.000	0.000	0.000	0.000																								
schools	26	0.000	0.000	0.001	0.012																									
opt	27	0.000	0.000	0.076																										
cpt	28	0.000	0.024																											
visa	29	0.053																												

Fig. 8. Support for sets of topics.

lift	T=206	visa	cpt	opt	school	Pathway	family	student	History	education	University	work	World	Canada	International	policy	government	law	week	sexual	cost	security	House Speaker	UK	illegal	border	financial	global	Trump	US
		29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
US	1	0.644	0.047	0.088	0.412	1.824	0.293	0.988	3.157	1.129	1.422	0.681	0.805	0.380	0.634	1.050	1.454	0.354	0.233	0.000	0.942	0.117	0.388	0.406	0.738	0.610	0.140	1.371	0.808	
Trump	2	0.469	0.003	0.086	0.032	0.000	0.850	0.091	1.590	0.138	0.106	1.622	1.392	0.102	0.016	2.410	0.674	0.343	0.609	0.227	3.547	1.038	1.232	0.025	2.187	1.717	0.080	1.827		
global	3	0.018	0.001	0.097	0.040	0.000	0.000	0.036	0.000	0.254	0.008	0.011	1.143	0.012	0.099	0.012	0.000	0.006	0.009	0.009	0.004	0.602	0.000	0.022	0.006	0.100	0.013			
financial	4	0.001	0.000	0.002	0.005	0.000	0.005	0.014	0.000	0.007	0.006	0.002	0.005	0.011	0.007	0.002	0.004	0.000	0.009	0.005	0.010	0.000	0.003	0.000	0.008	0.000				
border	5	0.022	0.000	0.128	0.024	0.000	0.099	0.061	0.047	0.004	0.004	0.063	0.040	0.072	0.016	0.385	0.050	1.766	1.022	0.000	0.048	6.765	1.797	0.013	0.929					
illegal	6	0.072	0.000	0.035	0.103	0.000	0.131	0.008	0.071	0.045	0.002	0.866	0.060	0.064	0.007	0.111	0.307	0.674	0.020	0.019	3.085	0.863	1.932	0.057						
UK	7	0.095	0.020	0.106	0.055	0.365	0.503	0.162	0.142	0.295	0.110	0.145	0.075	0.076	0.233	1.415	0.473	0.114	0.098	0.090	0.302	0.022	0.011							
House Spe	8	0.002	0.000	0.004	0.003	0.000	0.050	0.029	0.047	0.000	0.010	0.019	0.005	0.001	0.012	0.009	0.008	0.024	0.007	0.000	0.062	0.078								
security	9	0.008	0.000	0.110	0.003	0.000	0.005	0.019	0.000	0.019	0.002	0.010	0.482	0.000	0.037	0.040	0.017	0.013	0.870	0.000	0.003									
cost	10	0.286	0.010	0.032	0.053	0.365	0.084	0.026	0.000	0.101	0.003	0.431	0.005	0.032	0.051	0.067	0.544	0.042	0.007	0.000										
sexual	11	0.005	0.000	0.060	0.098	0.000	0.002	0.006	0.000	0.127	0.008	0.007	0.000	0.003	0.005	0.063	0.529	0.002	0.063											
week	12	0.475	0.035	0.028	0.166	0.000	0.042	0.188	0.190	0.093	0.176	0.080	0.020	0.023	0.115	0.044	0.035	0.056												
law	13	0.120	0.019	0.100	0.544	0.000	0.233	0.110	0.142	0.064	0.042	0.647	0.477	0.052	0.032	0.257	0.405													
governme	14	0.022	0.000	0.024	0.016	0.000	0.144	0.612	0.071	0.105	0.379	0.131	0.005	0.124	0.053	0.190														
policy	15	0.012	0.001	0.096	0.000	0.000	0.074	0.036	0.024	0.131	0.000	0.136	0.030	0.068	0.062															
Internatio	16	0.019	0.001	0.006	0.354	0.000	0.156	2.096	0.095	1.215	0.029	0.047	0.413	0.077																
Canada	17	0.109	0.000	0.015	0.016	0.365	0.183	0.161	0.000	0.045	0.039	0.279	0.060																	
World	18	0.058	0.008	0.015	0.000	0.000	0.002	0.012	0.119	0.019	0.048	0.019																		
work	19	0.545	0.026	0.276	0.541	0.000	0.731	0.381	0.926	0.258	0.192																			
University	20	0.001	0.000	0.009	0.399	0.365	0.050	0.474	1.044	0.172																				
education	21	0.006	0.026	0.086	0.533	0.365	0.020	0.198	0.000																					
History	22	0.001	0.000	0.004	0.005	0.000	0.002	0.008																						
student	23	0.099	0.004	0.035	0.892	0.000	0.159																							
family	24	0.091	0.010	0.011	0.016	0.000																								
Pathway	25	0.000	0.000	0.000	0.000																									
schools	26	0.012	0.001	0.074																										
opt	27	0.024	0.006																											
cpt	28	0.001																												
visa	29																													

Fig. 9. Lift for sets of topics.

Table 3
Frequent topics in Tweeter and article datasets.

"abroad"	"class"	"family"	"home"	"law"	"performance"	"security"	"urban"
"academic"	"cost"	"fee"	"host"	"level"	"policy"	"service"	"US"
"achievement"	"culture"	"foreigner"	"Income"	"link"	"population"	"social"	"utilization"
"Australia"	"degree"	"France"	"India"	"living cost"	"poverty"	"student"	"Vietnam"
"Family background"	"destination"	"GDP"	"International"	"mobile"	"quality"	"tuition"	"work"
"Bologna"	"diploma"	"gender"	"Italy"	"media"	"relatives"	"Turkey"	"world"
"border"	"economic"	"Germany"	"Job"	"network’	"resource"	"UK"	
"Brazil"	"education"	"global"	"Kazakhstan"	"origin"	"rural"	"undergraduate programs"	
"Canada"	"enrollment"	"government"	"knowledge"	"overseas"	"Saudi "	"unemployment"	
"China"	"Europe"	"growth"	"language"	"parents"	"school"	"university"	

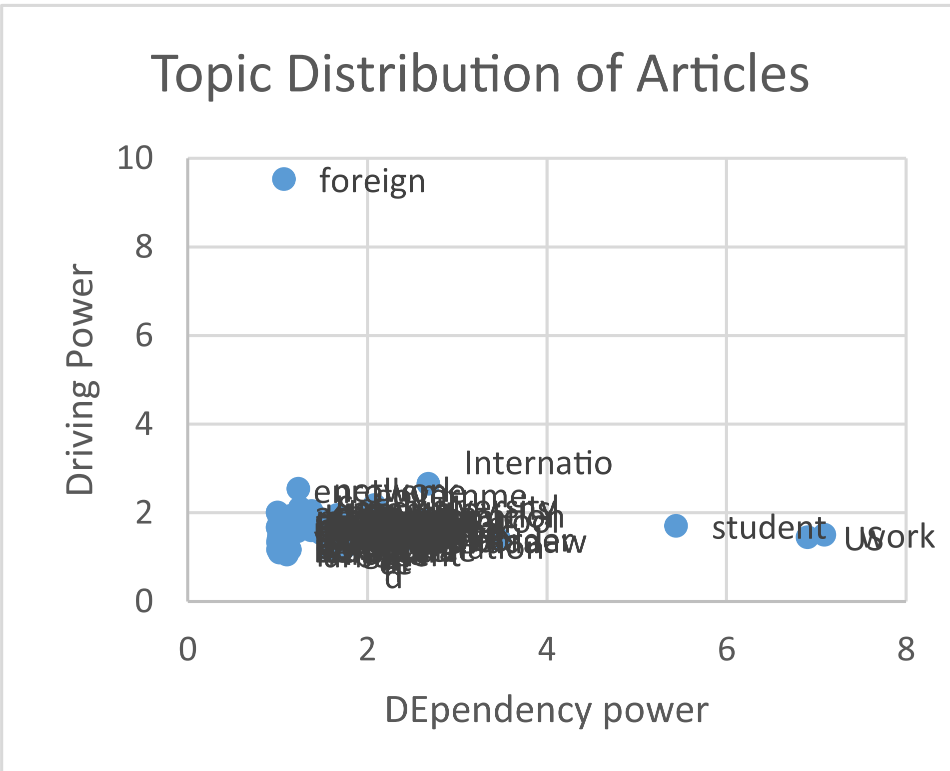

Fig. 10. Topic distribution of articles with outliers.

observable in. There are two exceptions in this plot which are shown by arrows. Exceptions are two factors with a considerable distance, meaning very different driving power and dependency power. The first is "Family background," The second is "Education." Family background in tweets has a Driving power of 20.7 and dependency power of 15.4, while the same factor in the academic article has 5.5 and dependency 5.2. The differences between values indicate that people have more words about the family background when they post on Twitter than the researchers when writing an article about student mobility. In other words, we can say that researchers do not have much concern or do not have access to the family background in their publications. At the same time, the general opinion on Twitter puts this factor in a powerful position. Having limited differences in driving and dependency power of the other 11 factors can verify the TFISM approach results for textual data. In Fig. 14, we can also compare the power of each category with the others. "Countries" is a linkage variable with the highest driving power and dependency power. "Education" also has high driving power and dependency power which group it into linkage factors. "Economic" and "Abroad" also have high driving power that causes put them into independent factors. The other factors can be classified into Autonomous factors.

We also run MICMAC analysis for a list of countries in data. Table 5 presents the list of countries we observed both in academic articles and on Twitter. Some of these countries are the joint host countries, some of them are famous as the origin countries, and there are countries in the list that consider as both host and origin. Therefore, we identify each country’s driving power and dependency power in Fig. 15. As Fig. 15 shown, all the 14 countries from tweets and articles have very close power, and we cannot see any out range between the two datasets. While the U.S. has a significant dependency and average driving power, it classifies as a linkage country. If the U.S. has a considerable change in its policies of international student education, the other countries may be affected drastically. Germany and China have high driving power, but their dependency power is insignificant. The interesting point in Fig. 15 is that there is no variable in the autonomous category that is approved; all these countries have a meaningful role in student mobility. All the countries have driving power between 1.2 and 2, but the range of dependency is varied.

6. Discussions

This section is organized with three main subsections that cover all aspects of this research. The first contributes to the theory that reviews findings that this study adds to the student mobility topic. The second one contributes to the methodology, highlighting the novelty of the created method. The last part is the implication to the practice, which identifies the impact of our work from different aspects.

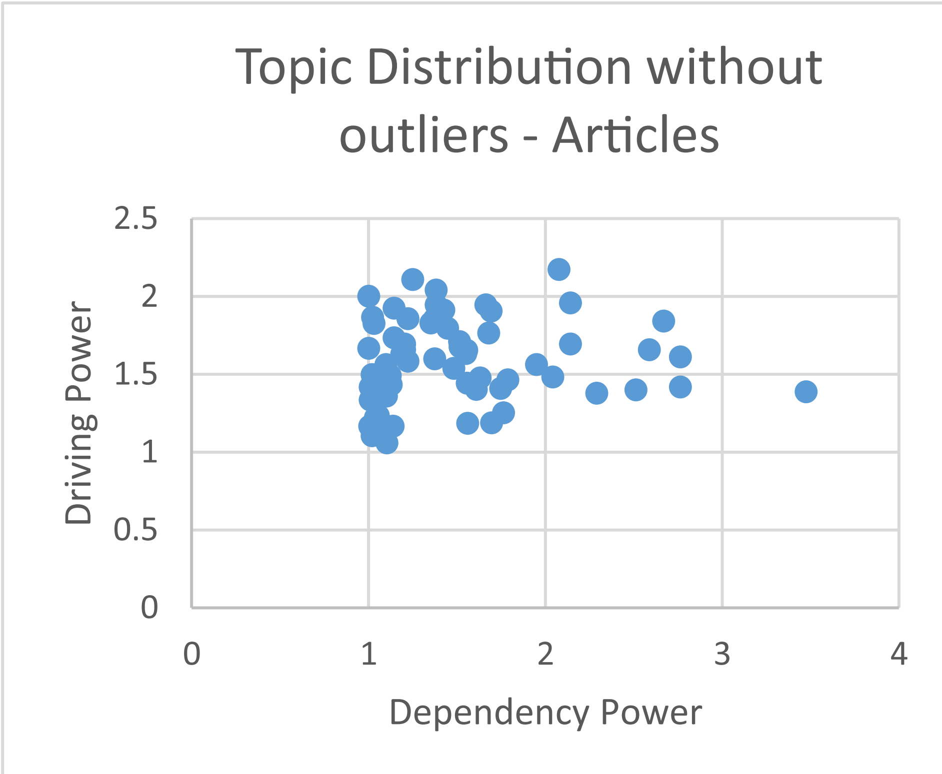

Fig. 11. Topic distribution of Tweets with outliers.

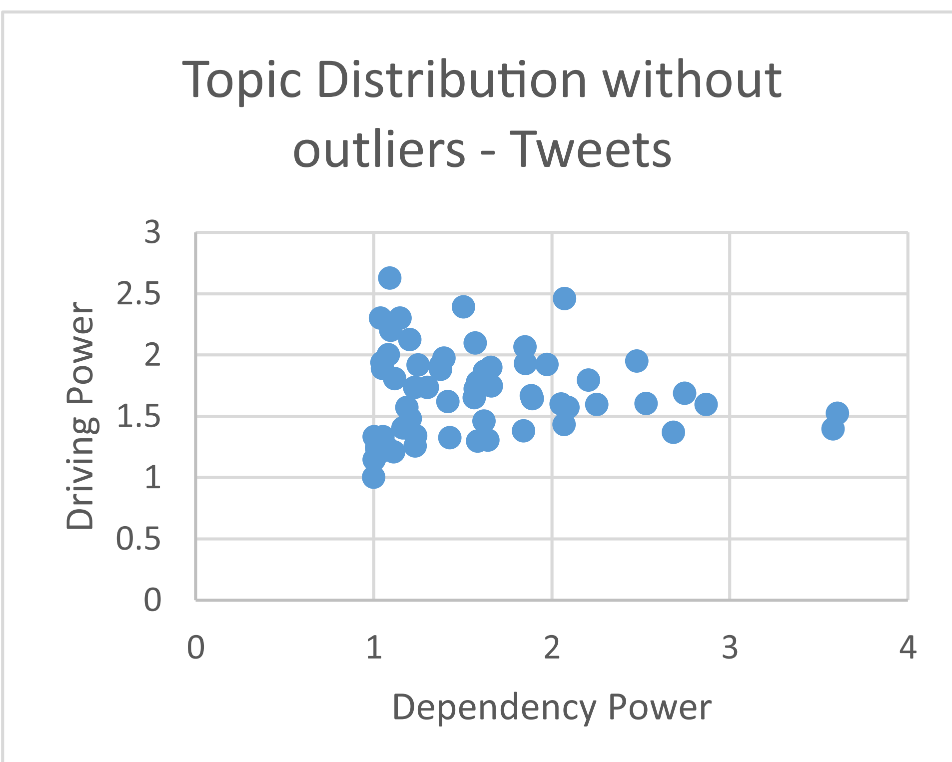

Fig. 12. Topic distribution of Tweets without outliers.

6.1. Contribution to the Theory

Researchers from different domains still doubt social media's trustworthiness as a research data source. The feasibility of using Twitter data for student migration is one of the contributions. Although many researchers previously confirmed the feasibility of social media, content analysis in this domain is not shared. One of the exciting results that should be discussed is the closeness of results from the test data set, which is social media data, and ground truth, which are the articles, such as topic distributions for both data sets that are presented in Fig. 13. This figure is built after removing the outliers topics. They were considering the outliers also important. There are two common

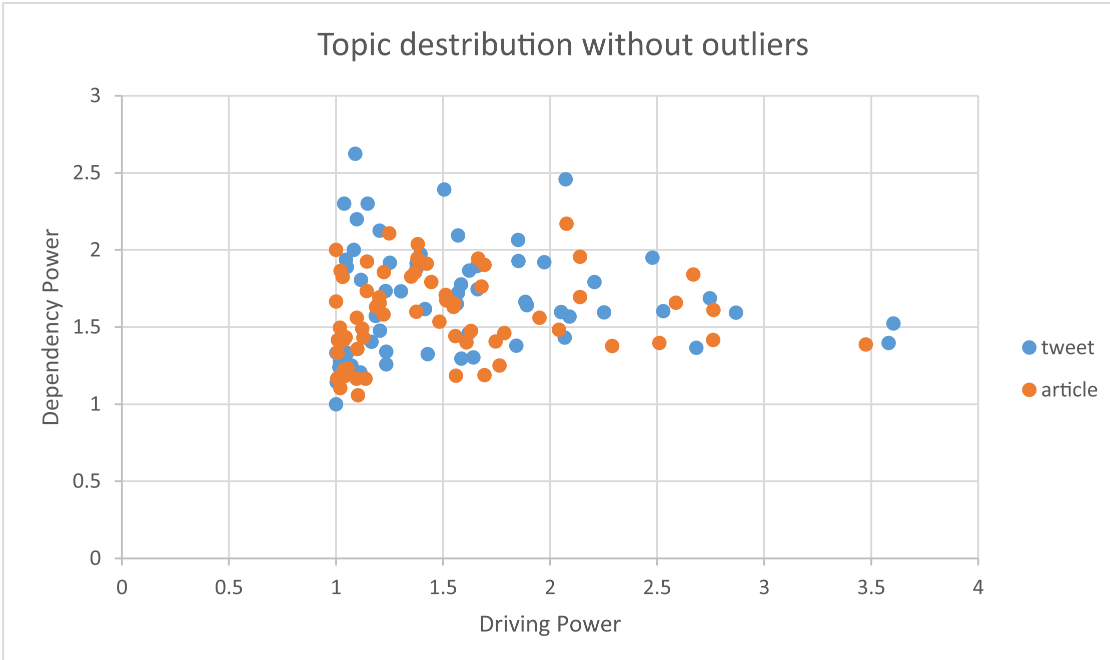

Fig. 13. Topic distribution without outliers- Twitter and Articles.

Table 4
Topics and assigned categories.

No.	Topic	Category	No.	Topic	Category
1	Abroad	Abroad	38	enrollment	Education
2	Foreigner	Abroad	39	knowledge	Education
3	International	Abroad	40	school	Education
4	Overseas	Abroad	41	student	Education
5	Australia	Country	42	university	Education
6	Brazil	Country	43	background	Family Background
7	Canada	Country	44	family	Family Background
8	China	Country	45	parents	Family Background
9	France	Country	46	performance	Family Background
10	Germany	Country	47	cost	Financial
11	India	Country	48	fee	Financial
12	Italy	Country	49	growth	Financial
13	Kazakhstan	Country	50	cost	Financial
14	Saudi	Country	51	fee	Financial
15	Turkey	Country	52	growth	Financial
16	UK	Country	53	Income	Financial
17	US	Country	54	tuition	Financial
18	Vietnam	Country	55	global	globalization
19	Culture	Demographic	56	world	globalization
20	Gender	Demographic	57	destination	Host
21	Language	Demographic	58	Europe	Host
22	Economic	Economic	59	host	Host
23	GDP	Economic	60	Job	Job
24	Population	Economic	61	unemployment	Job
25	Poverty	Economic	62	work	Job
26	Quality	Economic	63	link	Network
27	Resource	Economic	64	media	Network
28	Rural	Economic	65	Network	Network
29	Service	Economic	67	relatives	Network
30	Urban	Economic	68	social	Network
31	Utilization	Economic	69	home	Origin
32	Academic	Education	70	Origin	Origin
33	Bologna	Education	71	border	Political
34	Class	Education	72	government	Political
35	Degree	Education	73	law	Political
36	Diploma	Education	74	policy	Political
37	Education	Education	75	security	Political

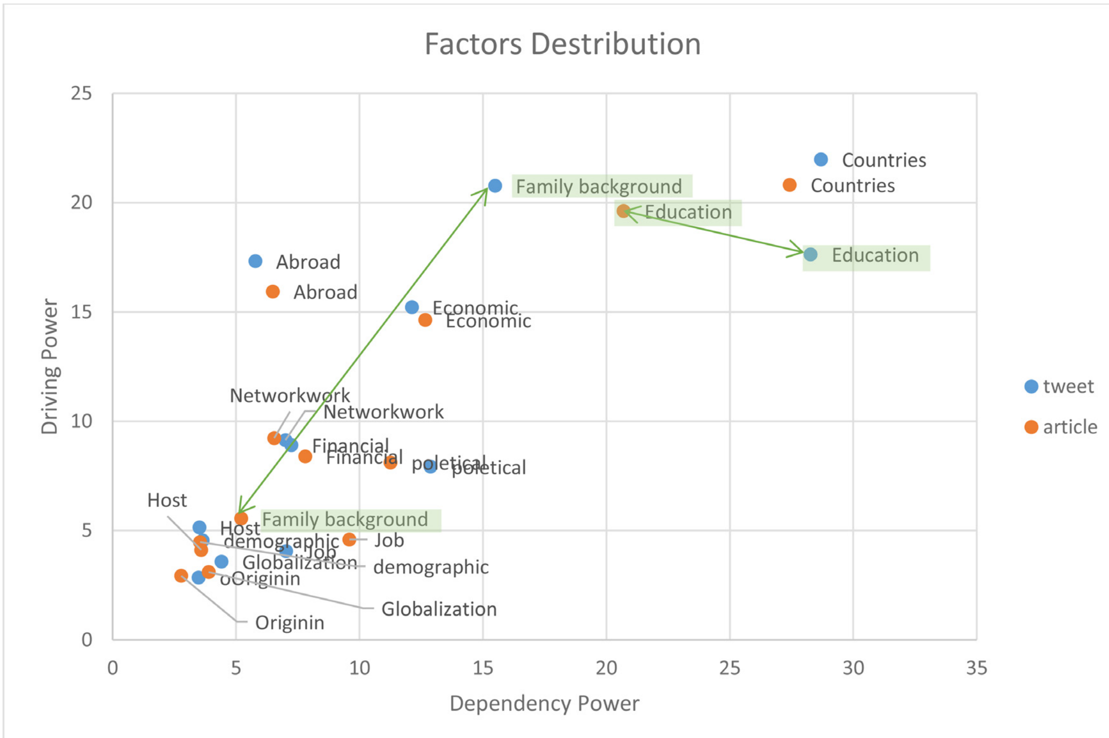

Fig. 14. Factor distribution for Twitter and articles.

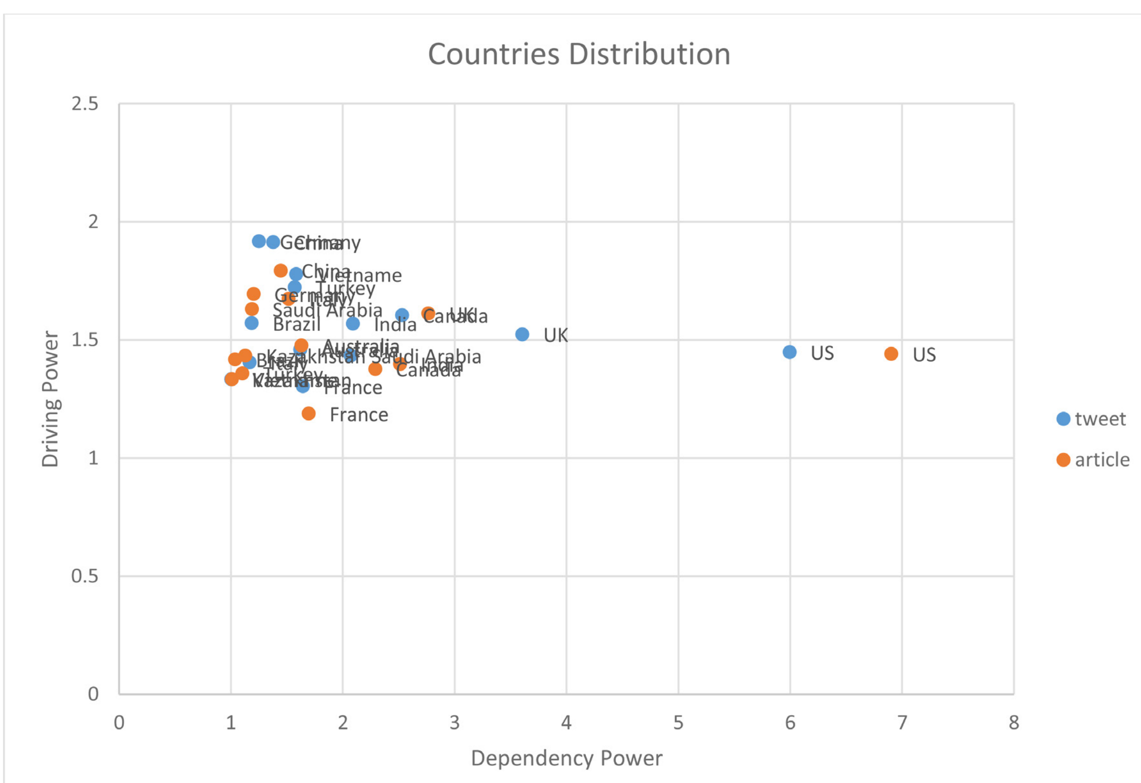

Fig. 15. Countries distributions- Twitter and articles.

Table 5
List of countries observed in datasets.

No.	Country Name	No.	Country Name
1	Australia	8	Italy
2	Brazil	9	Kazakhstan
3	Canada	10	Saudi Arabia
4	China	11	Turkey
5	France	12	The U.K.
6	Germany	13	The U.S.
7	India	14	Vietnam

words in outliers of datasets. They are "foreign" and "student." It is not surprising to have these words repeated because the whole topic is international students or international students. The other two outliers of articles are "The U.S." and "work," and another outlier of tweets is "Family." The differences between these outliers come from a more different point of view of student mobility. In fact, on Twitter, people talk more about the family of international students, while researchers may pay less to this topic because of the access to the data or the hardness of the study. Removing the outliers shows the distributions of the topic for both datasets with the same pattern. Although the pattern of distributions cannot show the closeness of the topics one by one, it confirms that both datasets have a very similar shape of distributions. For comparing the distance of topics from two datasets, we should look deep at Fig. 14. It shows the distance of each set of topics. As shown in Fig. 14, almost all topics from one dataset are close to the same topic from the other dataset since two sets of topics have considerable distance.

"Family background" is the one that has very high driving power and dependency power from tweets in comparison to the data from articles. This fact confirmed that people who share posts on social media rely more on their family background, and they talk about this data. We should be aware this type of data is hard to capture for scientists and researchers. Another topic from Tweeter, which is far from its mirror from articles, is "Education." Although the distance between these two is shorter than the distance between "Family background," it is considerable. "Education" in Tweeter dataset has higher dependency power than "Education" in the article dataset, while their driving power does not have essential differences. This fact indicates that people on social media relate education to more numbers of other topics. In simpler words, people talk about more factors while talking about education. (More factors mean more different factors and more repetition of each factor.)

Another observation that should be discussed is the list of countries. The countries from both datasets are presented in Fig. 15. There is no considerable distance between the name of each country from datasets. The U.S. has the highest dependency power among these countries, which means the U.S. comes after many other factors more than other countries' names. Germany has the highest Driving power, and this fact indicates that many other factors follow the word of Germany. It can explain two different policies of these two countries to attract immigrant and international students. The U.S. has an attraction for international students traditionally because of the famous university and opportunities for having a good job, while in recent years, Germany projected a big plan to attract international students.

6.2. Contribution to the Methodology

ISM can be used for identifying relationships and highlighting the power of variables for problems in any domain. However, the ISM is traditionally used for very limited fields. One of the reasons for not popularity of ISM is limited access to the domain experts who know about ISM. In this situation, TFISM enables all users _without considering expertise_ to leverage related documents and reach the results. In other words, TFISM provides a condition to solve a decision-making problem or any analytical issues for interconnection variables without requiring deep knowledge of the domain. It is necessary to have appropriate document or text data that contains valid information about the analyzing factors.

Another contribution of TFISM is proving new looks for analyzing the problems. Although two plots provided by the ISM method (MICMAC analysis and hierarchical analysis) are not innovative, having these views for new fields that ISM is not popular gives us a new vision and insight that provide a better understanding of the problem improved solutions. One example is using TFISM to compare companies' privacy policy and their attitude toward a privacy policy. With the help of TFISM, we analyzed the privacy policy of more than 250 companies and compared them with General Data Protection Regulation (GDPR) (Razavisousan). The plots provided by TFISM gave us a new look at the privacy policy and how companies approach it.

The original version of ISM was built based on the opinion of the human being. Although experts usually develop the ISM matrices, human cognitive errors (Stiegler) constantly threaten their discussion. The considerable enhancement of computational social science is that we can minimize cognitive errors by trusting written references and machine computation with the help of TFISM. We should be aware that TFISM can have cognitive biases if the input data is collected with preferences. However, computational analysis enables us to have big data, which will reduce the chance of biases.

In summary, TFISM provides a new look with the fast results for analyzing conditions with inside variables without dependency on expert knowledge. It is domain-free modeling that relies on text documents in each field.

6.3. The implication to practice

Relying on the text makes TFISM approach domain-free. Providing adequate valid text data enables us to apply TFISM for any domain. We may not have expert knowledge in a specific field. Still, TFISM, with accessing knowledgeable textual sources, will help us have a professional view of the factors, associations, and processes between them.

Furthermore, the proposed method can highlight the differences between data sources. In other words, TFISM shows the power of similar factors from a different document. This feature may help compare the content for the paper with very similar topics. When equivalent factor grades with different driving power and dependency power, it can also help us clear the other points of view of content producers.

7. Conclusion

Student mobility has been a matter of very high importance for the government and humanities. In this paper, we propose a practical method to build a model for student mobility based on textual data. Implementing TFISM as a hybrid methodology enables us to understand the mass text data and extract the relational model between extracted factors. Moreover, this method facilitates analyzing the factors and validating their power that before had been done with the consultations of the domain's expert. In a comparison of other researches, knowing the power of each factor make a valuable change in the process of decision-making. This method can apply to the textual data from different domains. It can also be applied to many fields of study and can help us manage unsupervised textual data. This proposed approach was evaluated by expert knowledge and the multi-level internal process. For student mobility, the results clarify each factor's power and influence. (Figs. 16, 12 and 13).

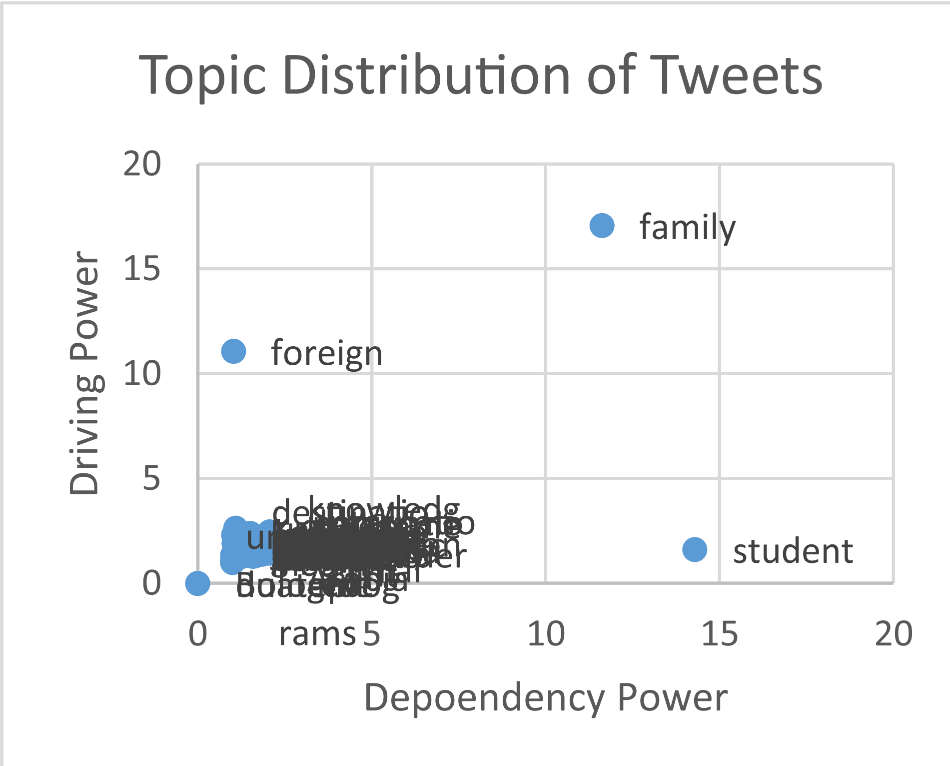

Fig. 16. Topic distribution of articles without outliers.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ahmad, M., Tang, X. W., Qiu, J. N., & Ahmad, F. (2019). Interpretive Structural Modeling and MICMAC Analysis for Identifying and Benchmarking Significant Factors of Seismic Soil Liquefaction. *Applied Sciences, 9*(2), 233.

Ahn, Jisoo, Son, Hyunsang, & Chung, Arnold Dongwoo (2021). Understanding public engagement on twitter using topic modeling: The 2019 Ridgecrest earthquake case. *International journal of information management data insights, 1*(2), Article 100033.

Attri, R., Dev, N., & Sharma, V. (2013). Interpretive structural modeling (ISM) approach: an overview. *Research Journal of Management Sciences, 2319*, 1171.

Beine, M., Noël, R., & Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education review, 41*, 40–54.

Bhandari, R. (2017). A world on the move: Trends in global student mobility.

Choudaha, R. a (2012). *Choudaha, R. and Kono, Y.* New York: World Education Services.

Choudaha, R. (2017). Three waves of international student mobility (1999–2020). *Studies in Higher Education, 42*(5), 825–832.

Crymble, Adam. (2019). Introduction to Gravity Models of Migration & Trade. *The Programming Historian*.

Findlay, A. M., Stam, A., King, R., & Ruiz-Gelices, E. (2005). International opportunities: searching for the meaning of student migration. *Geographica Helvetica, 60*(3), 192–200.

Grover, P., & Kar, A. K. (2020). User engagement for mobile payment service providers–introducing the social media engagement model. *Journal of Retailing and Consumer Services*, 53.

Guruz, K. (2011). *Higher education and international student mobility in the global knowledge economy: Revised and updated second edition.* Suny Press.

Harvey, Michael, Kiessling, Tim, & Moeller, Miriam (2011). Globalization and the inward flow of immigrants: Issues associated with the inpatriation of global managers. *Human Resource Development Quarterly, 22*(2), 177–194.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science, 41*(3), 260–271.

Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80–88).

https://www.migrationdataportal.org/themes/big-data-migration-and-human-mobility

Hughes, C., Zagheni, E., Abel, G. J., Sorichetta, A., Wi'sniowski, A., Weber, I., & Tatem, A. J. (2016). Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data: Feasibility study on Inferring (labour) mobility and migration in the European Union from big data and social media data.

Hughes, C., Zagheni, E., Abel, G. J., Sorichetta, A., Wi'sniowski, A., Weber, I., & Tatem, A. J. (2016). Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data: Feasibility study on Inferring (labour) mobility and migration in the European Union from big data and social media data.

Isbister, John. (1998). The Immigration Debate: Studies on the Economic, Demographic, and Fiscal Effects of Immigration. *Population and Development Review, 24*(3), 636 -636.

Jiang, Xiaodan, et al., (2018). Using interpretive structural modeling and fuzzy analytic network process to identify and allocate risks in Arctic shipping strategic alliance. *Polar Science, 17*, 83–93.

Jiang, Xiaodan, et al., (2018). Using interpretive structural modeling and fuzzy analytic network process to identify and allocate risks in Arctic shipping strategic alliance. *Polar Science, 17*, 83–93.

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied economics, 32*(13), 1745–1755.

Khatwani, G., Singh, S. P., Trivedi, A., & Chauhan, A. (2015). Fuzzy-TISM: A fuzzy extension of TISM for group decision making. *Global Journal of Flexible Systems Management, 16*(1), 97–112.

Komito, L. (2011). Social media and migration: Virtual community 2.0. *Journal of the American society for information science and technology, 62*(6), 1075–1086.

Kota, Srinivas, et al., (2021). Sustainable Production System Critical Success Factors: An Interpretive Structural Modelling approach. *Procedia CIRP, 98*, 324–329.

Kumar, Sunil, Kar, Arpan Kumar, & Vigneswara Ilavarasan, P. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), Article 100008.

Kushwaha, Amit Kumar, Kar, Arpan Kumar, & Dwivedi, Yogesh K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), Article 100017.

Lee, A. H., Kang, H. Y., & Chang, C. C. (2011). An integrated interpretive structural modeling–fuzzy analytic network process–benefits, opportunities, costs and risks model for selecting technologies. *International Journal of Information Technology & Decision Making, 10*(05), 843–871.

Llorente Pinto, A., Garcia-Herranz, M., Cebrian, M., & Moro, E (2015). *Social Media Fingerprints of Unemployment.*

Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142–150).

McCallum, A., Wang, X., & Mohanty, N. (2006, June). Joint group and topic discovery from relations and text. In *ICML Workshop on Statistical Network Analysis* (pp. 28–44). Berlin, Heidelberg: Springer.

McGregor, E., & Siegel, M. (2013). *Social media and migration research (No. 068)*. United Nations University-Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).

Menon, Rakesh R., & Ravi, V. (2021). Analysis of barriers of sustainable supply chain management in electronics industry: an interpretive structural modelling approach. *Cleaner and Responsible Consumption, 3*, Article 100026.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint* arXiv:1301.3781.

MSc., D. M. (2000). Interpretive Structural Modeling (ISM) . White paper.

Nasir, Jamal Abdul, Khan, Osama Subhani, & Varlamis, Iraklis (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights, 1*(1), Article 100007.

Poot, Jacques, et al., (2016). The gravity model of migration: the successful comeback of an ageing superstar in regional science. *Available at SSRN 2864830*.

RAGADE, RAMMOHAN K (1976). Fuzzy interpretive structural modeling. *Cybernetics and System, 6*(3-4), 189–211.

Rajan, Rishabh, et al., (2021). Developing a modified total interpretive structural model (M-TISM) for organizational strategic cybersecurity management. *Technological Forecasting and Social Change, 170*, Article 120872.

Razavisousan, R., & Joshi, K. P. (2019, December). Two Tier Analysis of Social Media Collaboration for Student Migration. In *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)* (pp. 158–167). IEEE.

Sandbhor, S., & Botre, R. (2014). Applying total interpretive structural modeling to study factors affecting construction labour productivity. *Australasian Journal of Construction Economics and Building, The, 14*(1), 20.

Shrimali, A. K. (2019). Interpretive structural modeling of identified success factors to lean implementation in SMEs. *Quality Production Improvement-QPI, 1*(1), 69–77.

Simini, Filippo, et al., (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96–100.

State, Bogdan, et al., (2014). Migration of Professionals to the U.S. In *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014*. Proceedings. Cham: Springer International Publishing.

Tobler, W. (1995). Migration: Ravenstein, thornthwaite, and beyond. *Urban Geography, 16*(4), 327–343.

United Nations, D. o. (2017). *International Migration Report 2017*. New York: UnitedNations.

Valmohammadi, C., & Dashti, S. (2016). Using interpretive structural modeling and fuzzy analytical process to identify and prioritize the interactive barriers of e-commerce implementation. *Information & Management, 53*(2), 157–168.

Verbik, L., & Lasanowski, V. (2007). International student mobility: Patterns and trends. *World Education News and Reviews, 20*(10), 1–16.

Warfield, J. N., & Cárdenas, A. R. (1994). *A handbook of interactive management* (p. 338). Ames: Iowa State University Press.

Warfield, J. (1979). *History and Applications of Interpretive Structural Modeling* 23 May, 1979.

Zadeh, L. A., Klir, G. J., & Yuan, B. (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers (Vol. 6)*. World Scientific.