A Study on Social Network Analysis: case study of UAE expats relationship to their home country through social media interactions

دراسة حول تحليل الشبكات الاجتماعية: دراسة علاقة المغتربين في دولة الإمارات ببلدهم الأم من خلال التفاعل عبر وسائل التواصل الاجتماعية.

by

REEM WARRAK

**A dissertation submitted in fulfilment
of the requirements for the degree of
MSc INFORMATION TECHNOLOGY MANAGEMENT
at
The British University in Dubai**

October 2018

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

_____
Signature of the student

COPYRIGHT AND INFORMATION TO USERS

Abstract

This paper is studying the influence of social media and displaying the significant role social media data is currently playing for benefit of both research & business context. Additionally, this paper will discuss social network analysis, challenges and benefits to manifest the emerging role of this new field of study.

As UAE community consists of 88% of immigrants and is ranked as one of the top countries in mobile social media penetration in the gulf region, it is interesting to analyze how much the expats are showing their connection to their home country by interacting with news accounts on social media platforms.

To achieve this13 countries were selected to reflect the diversity of UAE's demographics, the study used Facebook data and collected 155,694 posts from 65 news accounts, the analyzed data extended to use 8,848,404 comments and 4,666,774 identified users.

Results showed that social network analysis can reveal interesting relationships between the countries that reflect current political situations. Additionally, Pakistan was found to be the top nationality among UAE's expats who have the highest social interactions with their home countries through social media.

Finally we discussed challenges in this approach and suggested future work to be conducted.

Abstract (Arabic)

تدرس هذه الورقة تأثير وسائل التواصل الاجتماعي وتعرض الدور الهام الذي تلعبه حاليًا وسائل التواصل الاجتماعي للاستفادة من سياق البحث والأعمال. بالإضافة إلى ذلك ، ستناقش هذه الورقة تحليل الشبكات الاجتماعية والتحديات والفوائد لإظهار الدور الناشئ في هذا المجال الجديد في البحث العلمي.

بما أن مجتمع دولة الإمارات يتكون من 88٪ من المقيمين المغتربين ويصنف كأحد أكبر الدول في استعمال وسائل التواصل الاجتماعي في منطقة الخليج ، فمن المثير للاهتمام تحليل مدى إظهار الوافدين اتصالهم ببلدهم الأم من خلال التفاعل مع حسابات الأخبار على منصات وسائل التواصل الاجتماعي المختلفة.

ولتحقيق هذا ، تم اختيار 13 دولة لتعكس التنوع الديموغرافي لدولة الإمارات العربية المتحدة ، واستخدمت الدراسة بيانات موقع فيسبوك وجمعت 155،694 مشاركة من 65 حسابًا إخباريًا ، وامتدت البيانات التي تم تحليلها لاستخدام 8،848،404 تعليقًا و 4،666،774 مستخدمًا تم تحديدهم.

أظهرت النتائج أن تحليل شبكات التواصل الاجتماعي يمكن أن يكشف عن علاقات مثيرة بين الدول التي تعكس الأوضاع السياسية الحالية. بالإضافة إلى ذلك ، أظهرت النتائج أن باكستان على أنها أعلى جنسية بين المغتربين في الإمارات العربية المتحدة الذين لديهم أعلى تفاعلات مع أخبار أوطانهم من خلال وسائل التواصل الاجتماعية.

وأخيرا ناقشنا التحديات التي يمكن مواجهتها عند دراسة وتحليل شبكات التواصل الاجتماعية و أضفنا مقترحات لتطوير هذا العمل في المستقبل.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

GCC has attracted many labor forces in the past 50 decades due to the consistent infrastructure, social and urban development (Gardner 2015). With this move, the population growth has increased more than tenfold over the past half century, from 4 million in 1950 to 40 million by 2005 which is some of the most rapid growth.

In UAE the population has more than doubled in the past five years from fours millions in 2005 to eight million in 2010. (Forstenlechner & Rutledge 2011).

Country	2010 Population (mn)		Workforce		National Resource Wealth [a]		
	National	Non-national	National	Non-national	Oil [b]	Gas [c]	SWF [d]
Bahrain	0.51	0.54	36.1%	63.9%	...	.00037	16,852
Kuwait	1.04	2.43	16.9%	83.1%	0.98	.00173	284,615
Oman	2.39	1.02	28.7%	71.3%	0.02	.00029	3,431
Qatar	0.22	1.46	5.7%	94.3%	1.18	.11500	386,364
Saudi Arabia	20.94	7.75	50.5%	49.5%	0.13	.00038	22,818
UAE	0.95	7.24	4.2%	95.8%	1.03	.00632	759,053
GCC	26.05	20.45	38.3%	61.7%	0.19	.00161	61,313

Notes: [a] This is per national capita; [b] In millions of barrels; [c] In billion cubic metres; [d] In USD per national capita.

Table 1. GCC Demographic Imbalances & National Resource Wealth (Forstenlechner & Rutledge 2011).

Forstenlechner, I. and Rutledge, E.J. (2011) has discussed GCC demographic imbalance, their study showed that UAE has the highest non-national workforce in GCC with 95.8% as shown in table 1.

With only 12% of UAE's population is ethnically Emirati, the rest are mainly expats who permanently live in the UAE. With such demographics UAE has been identified as one of the

highest net immigration rates globally. People are attracted to UAE as it is considered to be one of the highest per capital incomes in the Middle East compared to Western regions. (Al-Jenaibi 2011)

Furthermore, in UAE, 88% of the demographics are from immigrants according to UN data as published in the CIA fact book (2017). Nearly three-fourth of UAEs residence use Internet regularly which makes it one of the highest penetrations in the world (Al-Jenaibi 2011).

This density of expats living in the UAE must have the need to relate to their home countries and in this study we are exploring how UAE foreign residences are connected to their origin countries through their interaction in social media.

Social Media platforms are web-based applications that enable users to generate their own content and allow them to exchange it easily.

The availability of Internet and the rapid growth of mobile technology are the primary force behind the move toward social media, providing interactive communication platforms that allow instant sharing of topics, opinions, news and even personal contents spreading in many formats ranging from text based to multi-media data.

This rise of social media has placed the content generated in social media platforms in a critical part of the information ecosystem creating an increasing interest to adopt this data and consume it in research purposes as well as in organizations, government's, businesses.. etc.

In the Middle East, this rise is also driven by the increased adoption of mobile technology and its rapid evolution, statistical studies shows that at least nine in ten nationals in Lebanon, Qatar, Saudi Arabia, and the UAE own a smartphone. (Radcliffe & Lam 2018)

A recent report (Radcliffe. and Lam 2018) showed that since 2013, internet penetration rose by more than 20 percentage points to 84% of all nationals from 7 countries participated in the study. Furthermore, the study states that 64% of users check about their friends and family as a daily activity online, this is almost double the rates in 2013 which was 28%.

The same report shows that "three-quarters or more of all nationals use social media to interact with others: 80% see what friends are family are doing, 76% post messages/ comments via messaging, 82% get messages via messaging, 76% post messages/comments via social media, 77% get messages via social media, 77% post/share their own multimedia content, 77% re-post/share others' links or content"

The UAE in particular have 88% active mobile social media penetration and is ranked the second country in the Middle East after Qatar.

Furthermore, 76% of the population in UAE has access to Internet which is one of the highest rates in the world as per International Communication Union. (Al-Jenaibi 2011)

Usage of social networks varies from advertising contents, fashion, entertainment and even reaches sport, music and reaches business development and news content.

Social networks are becoming valuable source of news in the Middle East region as elsewhere; however some major differences are noticed as per countries and age groups. In the same study, Radcliffe, D. and Lam, A (2018) demonstrated Facebook as the most popular source for news among young Arabs as per the annual Arab Youth survey which showed that 35% of the study respondents get daily news on Facebook, before online sources (31%), TV news

Channels and newspapers (9%)

Dennis, E.E., Martin, J.D. and Wood, R. (2017) have revealed that 66% of nationals in the Middle East get news and information on social media every day. Additionally, News has scored 31% among other content topics and identified as the top social media content users have shared or commented on.

1.1 Research Objective - Using Social Network Analysis to study UAE expat community

In this research, I will be studying the connection between expats in the UAE and their home countries, using social network analysis. The central hypothesis of this paper is that engagement on social networks can reveal insights regarding the connection between an expat and their home country. Current literature studying expat community in the UAE lacks usage of social media as a data source.

This study also seeks to contribute to the analysis of the impact of social media platforms in the daily life, it also demonstrates the value of social media platforms as a source of data for conducting academic studies which provides large amount of easy and frees to access data ready for usage.

1.2 Research Questions

This study focused on answering the following research questions:

- Which UAE's residents shows strong attachment to their home countries through interaction with News accounts in Facebook?

- Can social network analysis be used to reflect political events between the studied countries?

- Will social networks analysis reveal any interconnections between countries?

1.3   Research methodology

This is a quantitative study intending to show the level and strength of UAE expats relationship with their home countries through their interaction to news account in social media.

The collected data set used 155,694 Facebook Posts from 13 countries; retrieved posts had 8,848,404 comments that allowed us to explore common interactions and insightful results.

After data collection, preprocessing was performed to bring the data set to a shape that allows it to be imported into a Data Base for performing queries and gathering the required information. Furthermore, the resulted data set was visualized as a social network to extract further insights and retrieve results.

1.4   Structure of this study

The remainder of this research will have an overview of literature in Chapter 2 followed by detailed explanation of data retrieval and building the data set in Chapter 3. Chapter 4 is explaining steps followed for data preparation and pre-processing, analysis results and discussion are in Chapter 5, conclusions and future work proposals are presented in the Chapter 6.

Chapter 2: Literature Review

2.1   Social Media

In recent years, internet users have massively moved towards web based platforms that allow them to create their own content and share it with their selective set of users or even make it public starting what is now called a social media world. Social media applications allow different type of content such as: text, photos, locations, videos or even live streaming that can be generated, shared and discussed by the user and his/her social community. Lee, I. (2018) defines social media platforms as "online communities via which members seek and share common interests, activities, experiences, and information".

Users chooses social media platforms according to their needs and motivation, while Twitter is a real time network that connects followers that have common interests, Facebook is a network that gives users options to posts multimedia content to their private communities or to the public audience in a non-real-time manner. Instagram on the other hand is specialized in photo-sharing while SnapChat is about short videos. (Lee, I 2018)

Many of those Social Media platforms provide Application Programming Interfaces (APIs) that allow third parties to gain access into their data. Such feature provides access to user generated content and also to the users' demographic data whenever it's made public. (Lee, I 2018)

This concept of sharing has certainly created a shift of how people communicate, create and consume information that is publicly available. Specially with the high availability and low cost of internet along with the wide spread of mobile technologies, all those factors were the primary force behind the rise of Social media (Stieglitz et al. 2014)

2.2　Social Network Analysis & Big Data

The enormous usage of social media platforms resulted in an increasing accumulation of data generated by real life users which is now considered a rich source of data for academic researchers. The massive data produced by and about people, things and interactions among them is representing large scale datasets that is easily extracted from popular platforms such as Facebook, Twitter..etc. creating the new term "Big Social Data" where scholars are showing huge interest to use it in their researches.

Stieglitz et al. (2014) study talked about an emerging need to monitor, collect, analyze and summarize information from Social contents which will be of a great help in many domains such as public administration, politics, B2B and even decision making. However, these tasks do not appear to be simple tasks due to large number of social media platforms available and the vast amount and complexity of social media data and meta data associated with it such as views, like, shares, mentions.. etc; those attention-related data creates another level of information that needs to be addressed and considered as an additional source of information when studying social media data. This introduced a need to create a new field of research called "Social Media Analysis SMA" with an objective to establish a scientific base for mining, analyzing and modeling large scale Social media data and simplifying it for academic use, an additional goal of SMA is to establish and adapt methodologies that help to transform the social media data into a more useful meaningful information can be used for the benefit of the field under study.

Zeng et al. (2010) defines SMA as a field of research that is more about evolving and assessing technologies and informatics tools that are made for collecting, scanning, analyzing and outlining social media data to fulfill requirements from a specific context or application.

Their study is also encouraging the usage of Artificial Intelligence tools into Social Medial analysis to introduce the concept of social media intelligence that leverage the mature AI tools and frameworks in the social media data to derive beneficial information from the spectrum of social media and develop decision-aiding frameworks .

2.3   Usage of Social Media Data

Due to the explosive growth in the data generated and shared via social media platforms the popularity of social media data has massively increased, hence social media platforms are now adopted as a new channel of communication and considered as an important source for diffusing and procuring information in many different contexts such as: business, politics, government, crisis management and others.

The enormous number of users, and high volume of data in the social media platforms, has created a drive to many businesses and fields to know who is creating the contents and which profiles are creating high influence among the crowd. Such information will provide organization insights and help them improve mass communication with the public to serve their field. For example: Businesses have started publishing their marketing materials over social media applications allowing their products to reach millions of viewers by being just a click away and removing all the limitation of location and time. Governments also have used those platforms to make announcements or alert in case of emergencies as the concept of sharing

allows the news to spread quickly. In a political context, social media platforms have potentials to encourage voters and enhance political participation. (Stieglitz et al. 2018)

Lee, I (2018) and Stieglitz et al. (2018) have both mentioned many examples of using social media data, it has been agreed in both studies that social media is an important unit for business intelligence platforms that is emerging in the next generation. Furthermore, both researches agree that social media is not only a substantial source of information but can also be used for marketing, enhancing product design and strengthen consumer relationship.

From politics point of view, social media is an ideal platform to gain public opinion and to promote political candidates running elections for example.

Fan, W. and Gordon, M.D.(2014) have demonstrated more details around how can social media change consideration on the products and services through a products (or services) life cycle from design to disposal. The study classifies the product lifecycle in four main stages: Design-development, production, utilization and disposal. They also explored ways to assist companies to gain increased completive advantage by understanding the business environment, suppliers and customers. For example, trend analysis techniques and tools can help identifying changes in sentiments and customers' behavior which can impact the product design & development. Whereas, in production stage a business can use social media analytics tools to learn about competitor products and avoid their manufacturing mistakes, social media can also be used to closely monitor suppliers and assure flow of essential manufacturing materials is smooth and will not be disturbed by any political or natural situations.

Usage of social media data have also reached tourism sector, in the study done by Shao, H., Zhang, Y. and Li, W. (2017) they have used quantitative methodology to consume social media data generated by tourists to extract most common touristic districts in China—Huangshan, they have proposed a framework for analyzing touristic data and suggested future work to further enhance it to be used in urban development.

In a marketing context, trends shown by Misirlis, N. and Vlachopoulou, M. (2018) that tourism industry, Facebook and Twitter as well as consumer-centric marketing to be the dominant categories, platforms and concepts behind social media marketing strategies in the recent years.

From the above, Social Media Analysis can be used in the broad spectrum of data analysis and can fit many business and personal needs. Hence SMA has been placed in the focus of research considering its publicity and easily accessible data.

2.4   Challenges of Social Network Analysis:

Stieglitz et al. (2014) discussed that SMA still doesn't have a research methodology base that can be referred to since it is still emerging as a new field of research, hence a well-defined structure is needed to establish research agenda that can simplify the work for future studies.
 From research methodology point of view, the same study categorizes challenges faced in this field into two major difficulties: first, issues related to the social media data nature and the collection of this data; second is the analysis techniques and mining methods. Social media data is known to have complex nature, dynamic and generated in very large volumes. Hence standard analysis tools such as Database management systems, desktop visualization and statistics applications cannot be used.

Social media data consist of structured data such as user profile details, attention related data (such as likes, shared, follower ...etc) as well as unstructured data which is the user generated content that can vary from a simple textual status to complex materials that might contain audio-visual content and even geo-location information. Hence extracting meaningful information from such large scale complex source of data can be challenging for data mining and demand computing capabilities that can adapt to this high level of processing requirements.

Another challenge of social media data is the collection, social media platforms don't have a standard ways that allow their data to be collected as some of them like Facebook offers easy access methods such as application programming interface "APIs" others don't and in these cases manual parsing or scripting techniques are be required. When using APIs that are offered by some of the social media platforms user should always mind the limitations as those API are under limitations and restriction of data access and be aware of the standards as they are always subject to platform changes.

Once data is collected and before it can be used for further analysis pre-processing is required and it represents another challenge considering the Meta Data associated with social media information which has multiple formats in nature. Furthermore, the concept of social media platforms are fast changing in nature introducing new concepts and designs over the time in a rapid evolving way which introduce new information that need to be considered in the analysis such as retweets or hashtags.

Stieglitz et al. (2014) have also pointed to another challenge when using social media as a source of data, the reliability, considering the information derived from internet can be incomplete and inconsistence . This can increase the uncertainty and might lead to misinterpretations.

Lastly, privacy is always mentioned when collecting data from social media sources and analyses it then report results and findings. Some people always question how ethical it is to use social media data even when it is public.

From another perspective, Stieglitz et al. (2018) research was keen to have a specific classification of the challenges and adopted a theory from big data literature and used the concept of four V's : Volume,
Velocity, Variety, and Veracity. Where the volume challenge is when the collected data consumes high physical spaces on memory and hence reduces the performance and speed of any algorithms can be run on it. Velocity is about the speed of creating social media data and how quickly it can be analyzed on real-time or offline.   Variety of the social media data format is another challenge that was also addressed by Stieglitz et al. (2018) in similar way Stieglitz et al. (2014) discussed it with an added concerns to its real-time nature and rapidly changing format of the data which directly influence architectural choices. Lastly, veracity is concerns on the reliability and genuinely of the collected data as some researchers argue that social media data can be morally questionable since users cannot be surely linked to actual members of the society. In their study Stieglitz et al. (2018) used a qualitative approach to collect challenges and difficulties faced by researches in each phase of the study and finally proposed a solution for each identified challenge.

2.5   Steps of SMA

Researchers have made progress in structuring the process of social media analysis into defined frameworks that can be a reference for future studies to follow as they tend to categorize the needed work into defined steps.

Stieglitz et al. (2018) have adopted a framework proposed by them in earlier study published in 2014 which they claim it is highly used in SMA in context of IS literature , this framework breaks the SMA analysis process into three main phases , tracking, preparation & analysis. The study proposed additional preliminary step to the original framework which is Discovery to allow researchers to combine their pre analysis work such as topic discovery in one phase .
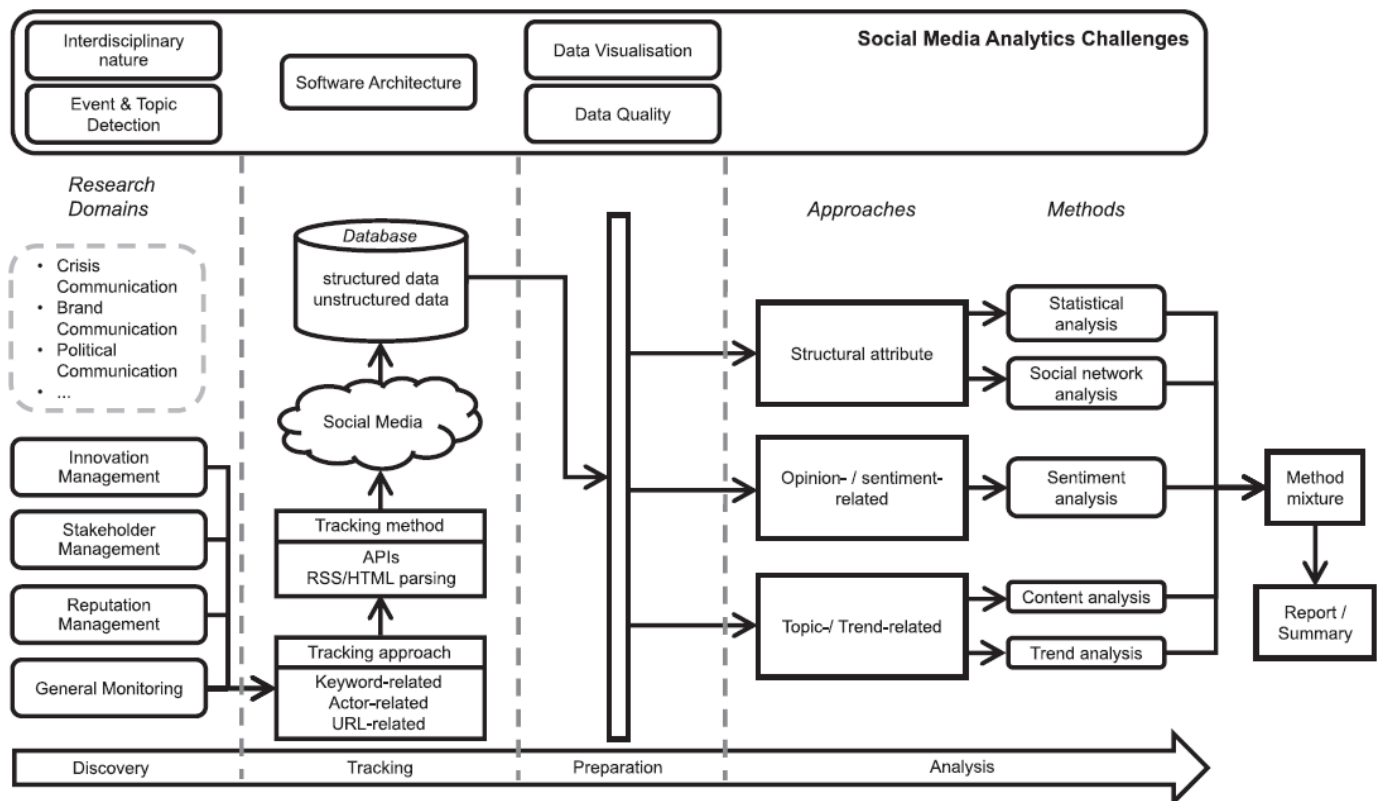
Figure 1: Suggested SMA Framework (Stieglitz et al. 2018, P. 165)

This Framework propose four main phases for SMA as below :

- Discovery: encapsulate the pre work such as topic selection and revealing patterns

- Tracking: this step outline required work to decide the source of social media data i.e. which platform to be adopted, and the methods of collecting data along with the formats of outputs.

- Preparation: Most researchers defined this step as data pre-process step which include removal of duplicated data, un wanted or irrelevant data. This step should prepare the collected data to be inputted into the analysis methods or algorithms.

- Analysis: depending on the nature of study, multiple SMA approached and methods can be adopted such as opinion mining or trend analysis.

Chapter 3: Data Retrieval & Selection

3.1   Social Media Platform & Tool

The study requires an access to historical and real-time data in order to fetch a reasonable amount of social media information which can be used to conduct a research.

We have used Facebook as the social media data source; Facebook APIs were adopted to fetch information on accounts, users and posts. For that we have used Facepager (Jünger, J and Keyling, T 2018) as a tool for retrieving data from Facebook.

Facebook offers many APIs to retrieve the data stored as Objects, most commonly used are Graph, public feed and Keyword Insight APIs. To fetch an object along with all properties the object ID should be available. To have this ID a Search API which is part of the Graph API can be used, in this API we can set the 'Query' input parameter to search values such as: user, page, place, group ..etc. the result of the search will return the unique ID of the object. Additionally, Facebook APIs will require an access token in the input; this can be obtained by having a valid account on Facebook. (Batrinca, B. and Treleaven 2015)

Facepager was developed mainly to collect information from social network sites through accessing and consuming the publicly available APIs, after getting the feeds Facepager have the ability to save the data and make it easy to navigate through a user friendly UI screens. Additionally, Facepager provides and extract option in many formats such as XLS or CVS sheets.

To start data retrieval the user should have a valid account at the social platform and then select the required API with the right configurations and parameters that will decide which and how the data should be fetched.

Facepager makes it easier to manage Object IDs and grant token access, as the user is advised to log in first in order to secure the token access for consuming APIs successfully. And that is easily managed in the user friendly screens. .

Usages of Facepager certainly have limitations, some of them are enforced by the social sites APIs and others are as per the software. For example Facebook APIs allow certain limit for accessing the amount of data retrieved, once these limits are exceeded the used account shall be blocked from getting API responses for certain amount of time. This is used to reduce the load on the social sites resulted from un-controlled hits from the same source that may be caused by scripts or spams. Furthermore, the size of the data retrieved on FacePager have limited number of pages to access, for example we have noticed that retrieval do not exceed 12,500 post per account.

The following steps were followed in Facepager to produce the resulted data set:

- Creation of database: for easier management and control we have created a separate Facepager Database for each country.
- Creation of seed nodes: Seed nodes are the main Facebook accounts that are the subject of study. In the case of our subject, those seed accounts are the country's 5 selected local news pages that have been nominated for each country as per the criteria in section 4.2.2.
- Filtering the required columns: as Facepager can retrieve even the media in each request we can customize the specific fields in order to reduce the volume of data used to fit into

our need. In the case of this study, we have discarded the content of the posts/comments as they are of no benefit. However, details of comments and users posted each comment is of great interest.

- Parameters set up: Facepager offers sets of parameters in the main UI that allow the user to change before hitting the APIs, in the case of our study we have used the default settings and changes only parameter for the number of pages the API should retrieve data from, as Facebook arrange the posts in pages format. This is set to 10 for this study. Additionally resource of data needs to be set and for post retrieval this is set to "page's posts"

Facepager assigns a unique identifier for each object retrieved, it also formulates the data in hierarchal format and defines objects in a parent/child relationship as per below illustration:
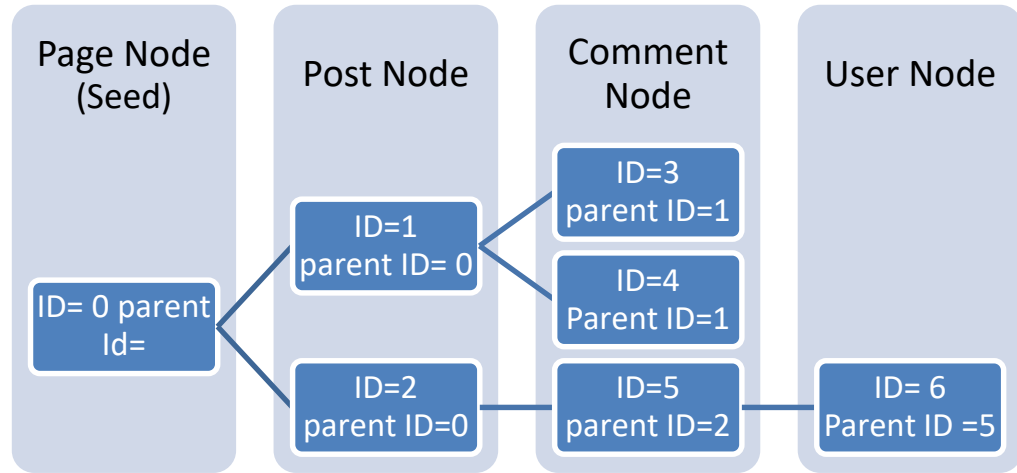
Figure 2. Data structure in Facepager.

As shown in Figure 2, the user found in the last column is the user who left comment id =6 on a post id=2 which was published in a Facebook page id=0 .

Once a node is retrieved it will be saved in Facepager database in Jason format, the search can go one level deeper to get further data about the selected node i.e. if a post node is selected we can retrieve related data about this post depending on the type of object selected in the parameters of the API. We can retrieve data about the author of the post, comments of the post or likes/shares of the post.

3.2   Countries

Since we need a perspective to start with, we have to narrow the selection of countries, hence we searched nationalities of UAE residence since the study focuses on expats in the UAE only.

Official authorities which publish formal statistics on UAE demographic such as Federal Competitiveness and Statistic Authority, Statistic Center Abu Dhabi and Dubai Statistics Center categorize UAE populations broadly to "Local" & " Non-Locals" or "GCC" and "non-GCC" only without any details regarding specific nationalities of UAE populations.

Hence we relied on sources such as articles and blogs that demonstrated UAE expatriate numbers.

Country	Estimated residence in UAE
India	2,600,000
Pakistan	1,200,000
Bangladesh	700,000
Philippines	525,530
IRAN	450,000
Egypt	400,000
Nepal	300,000
Srilanka	300,000
Syria	250,000
China	200,000
Jordan	200,000
Afghanistan	150000

Palestine	150000
United Kingdom	120000
South Africa	100000
Lebanon*	100000
Ethiopia	90000
Yemen	90000
Indonesia	85000
Sudan	75000
Saudi Arabia	70000
Somalia	70000
Iraq	52000
United States	50000

Table2. Number of UAE residence by country (Froilan T. et al. 2013)(GMI 2018) & (CIA 218)

From results in Table 2 we have selected top countries from Asia, Europe while focusing more on Arab countries. Hence 6 top Arab countries were considered along with top 4 countries in Asia. To add more variety to the data US and UK were added to represents western citizens.

The below table represents the final set of countries:

Region	Country
Asia	India
	Pakistan
	Philippines
	IRAN
Arab	Egypt

20

	Syria
	Jordan
	Lebanon
	Iraq
	Yemen
Western	United Kingdom
	United States

Table 3. Selected countries Vs Region

3.3   News Accounts on Facebook

As the topic of the study is to analyze UAE expats' relation to their home countries through interaction on social media, we have selected News accounts as those will have the major updates and highlights on each country where an expats will show interest to look at and interact with showing higher level of engagement to what is happening in their home countries.

After deciding on the list of countries that will be the base of the research data, we needed to define Facebook accounts that will be used for retrieving the subject data of study.

We relied on the following methodology to identify important Facebook accounts with respect to News :

- Using Facebook search tool box enter keywords combination of : Country, News, Local, Newspaper

- Filter results to see Pages only.

- Select top 5 Facebook pages accounts in terms of followers or likes

With the above criteria, each country is now represented by 5 Facebook Profiles, average number of likes for each country's 5 selected profiles is as per below figure :
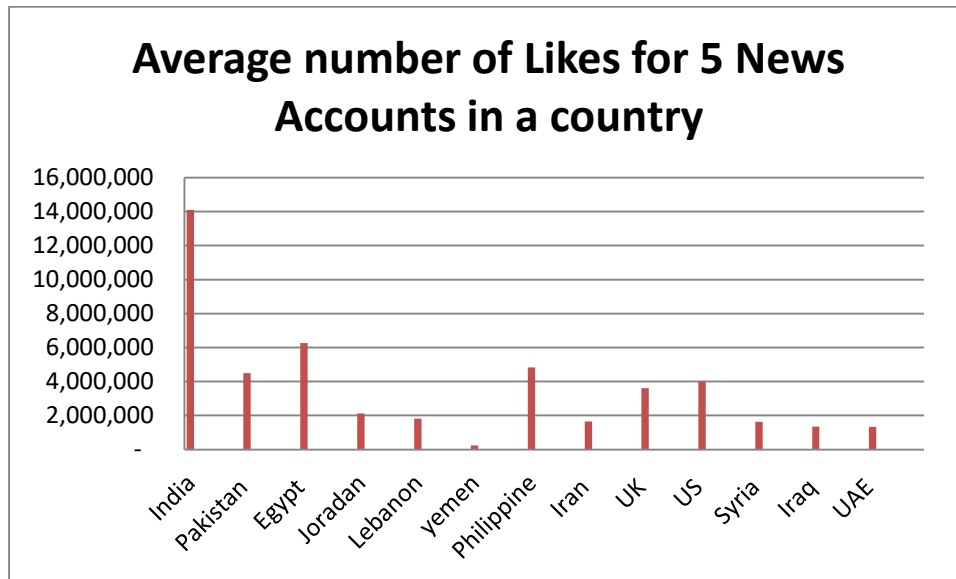
Figure 3. Average number of likes for each country's 5 Local News accounts

3.4   Posts

Posts were collected in two weeks span (22 Jan 2018 – 8 Fev2018), In order to produce valid results, we have introduced a constraint to fix the date of posts retrieval in this timespan in order to have results in the same time span across all countries and all News accounts.

It was noticed that Facepager do not retrieve more than 2600 posts per account resulting approximately 12,500 Posts for each country, this takes place when setting the page parameter on Facepager to 10 .

Retrieved post were created between 2014-03-27 & 2017-11-08
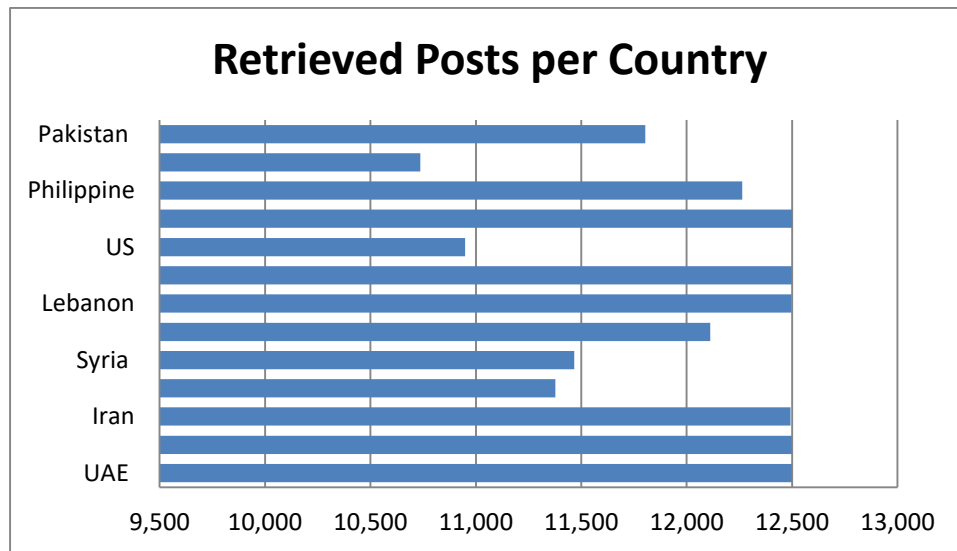
Figure 4. Total Posts retrieved for each country.

3.5   Comments & Users

The objective of this study is to define how expats are related to their home countries through their interaction with their local news accounts/pages on social media. We have made the below assumptions to achieve this objective:

- UAE expat: A user who interacted with news accounts of UAE local news accounts and another county's new accounts.

  This definition considers that UAE expat must be interested on the news happening on the country of residency and on news happening on their home countries as well.

- Interactions: social media platforms offers various ways of showin intrest, Facebook allow users to follow a page, like a post, leave a comment, share information ..etc. for the purpose of this study we consider a Facebook user is "Interacting" with a news account/page once the user leaves a comment on any post of that account.

Retrieved posts from Facepager as described in section 1.2.4 will provide the posts for each Facebook account going back for 10 pages of posts as per the defined parameter. However, in order to identify the users, posts alone are of no use if we don't retrieve their corresponding comments with details on users who have created those comments.

Following steps was done to retrieve comments for each post, results as per Figure 5:

- Change the resource parameter to "posts' comments"
- Selection of required posts that need to retrieve their corresponding comments.
- Define number of threads to be used to run and store this response, as the tool was running on personal desktop threads were varying from 5-10 per hit.

- Define the required fields to be retrieved in the column sections of the UI, for the subject of this study we have removed all content of the comments/posts as they are of no use to

Retrieved Comments & Users Per Country

	UAE	Egypt	Iran	Yemen	Syria	Iraq	Lebanon	Jordan	US	UK	Philippine	India	Pakistan
comments	202,677	252,442	409,788	67,585	537,419	992,278	721,487	702,629	990,206	995,034	993,176	990,497	993,186
users	127,185	160,064	208,472	33,903	284,770	449,679	373,623	375,938	515,914	703,225	548,856	423,546	461,599
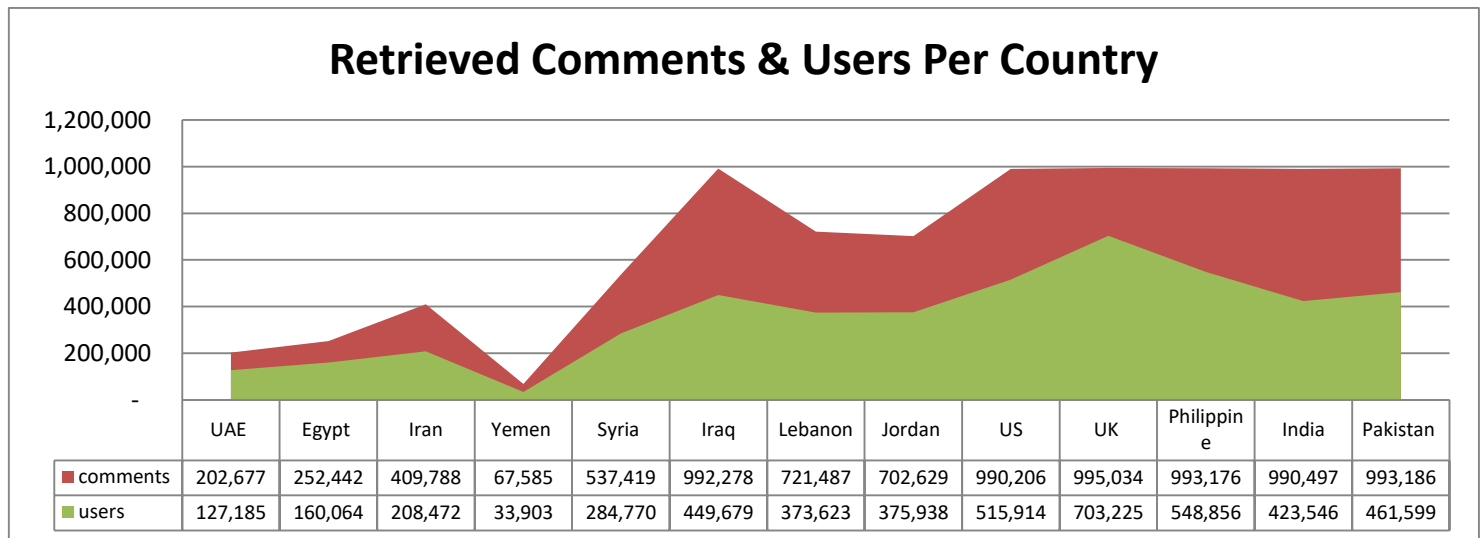
this study however all details of the creator of the comments are needed.

Figure 5. Number of retrieved Comments & users per Country (From 5 News Accounts/Pages)

Chapter 4: Data Preparations & Preprocessing

Social media data are naturally unstructured textual data, and hence can be very noisy, hence pre-processing is an important step in social network analysis. The process of data cleansing can range from removing typographic errors to removal of unwanted data. (Batrinca & Treleaven2015).

In this study we have followed the below steps for performing data cleaning & prepration to have a data set more suitable for analysis:

4.1   Raw data

The retrieved data using Facepager can be extracted in an excel sheet, we have extracted each country's data in a separate sheet to allow data manipulation and change since the size of a country's data can reach 6GB which makes it very difficult to display/amend and save changes on the sheets due to memory restriction on the machine used. So the raw data set had 13 spreadsheet one for each country.

4.2   Removal of unwanted data

Each retrieved object from Facepager which can be a post, comment or a user is expected to consume one row in the spreadsheet, while the properties of that object are spread over the columns. Unwanted line breaks were extensively noticed which caused each object to overrun on multiple rows which created difficulty in separating the objects and their properties.

4.3   Removal of duplicates

During data collection though Facepager and as explained in section 3.1, we can retrieve deeper data about an object upon selecting the node representing that object and hitting the search API.

If any node is mistakenly selected multiple times as this is a manual approach it will cause duplications on the final data set. However, this duplication can be identified as each retrieved object in a Facepager database is uniquely identified by an object id; hence duplications were filtered to distinguish unique rows using Excel formulas. A Total of 4335 duplicate posts were removed after this step.

4.4   Data Import into Access DB

After data is prepared as per the above, it was possible to import the retrieved data set into Access DB to easily extract information and gather aggregated outputs.

13 tables were imported into a single DB, each table represents a country, SQL scripts created for identifying interested facts and aggregated data from the collected dataset.

As explained in section 3.1, the hierarchal structure and the parent/child relationship in the dataset as defined by Facepager made it clear to identify common relationships among the data since every object has a unique identifier "object_id" we can additionally track the object relation to the parents through the "parent_id". With such framework of the dataset, we were able to recognize the interactions made by each users on the entire dataset.

4.5   Common interaction definition

The goal is to identify each country users' common interaction with the other countries news accounts, an interaction is defined by comments. So we would count a common interaction between country A & country B when a user X leaves a comment on a news account from country A & users X leaves another comment on a news account from country B.

This is illustrated as below in Figure 6.

Figure 6. An illustration of common interaction definition

Chapter 5: Results and Findings

5.1   Common Interactions

As defined in section 4.5 an interaction of a user toward a post is defined by comments i.e. an interaction is counted when the user leaves a comment on the post.

Access DB was used where SQL scripts created to cater for this information, below table shows the total common interaction among the countries in the dataset.

	UAE	Egypt	Iran	Yemen	Syria	Iraq	Lebanon	Jordan	US	UK	Philippine	India	Pakistan
UAE		904	745	190	1,063	1,058	2,128	2,346	612	1,274	2,279	876	4,244
Egypt			99	125	952	1,027	1,923	1,691	113	219	15	15	56
Iran				44	470	612	580	185	2,405	2,156	286	715	5,810
Yemen					402	102	146	264	23	45	-	4	31
Syria						2,145	9,379	4,029	268	515	119	50	393
Iraq							14,918	5,435	423	1,002	4	2	60
Lebanon								9,505	527	1,732	73	31	199
Jordan									248	592	30	15	132
US										10,796	3,453	555	1,879
UK											3,649	560	2,384

Philippine		31	272
India			2,181
Pakistan			

Table 4. Common interactions i.e. number of users commented on news accounts in both countries.

Each cell in Table 4 represents the number of common interaction between the 2 countries (Row X Column), the bottom of the table is not filled as the interactions between two countries would be the same either ways.

It is noticed from table 4 that Pakistani users in Facebook have the highest interest in news accounts from both UAE & Pakistan, while Yeminis are the least in interacting with UAE news accounts.

The data in Table 4 is showing the relation among the 13 countries in the study and interestingly users showed highest interest in both Lebanese & Iraqi news accounts. In contradiction, it is noticed that no common interest is shown between Philippine & Yemen news accounts.

The data was visualized in network format using Cytoscape software, which allows network data visualization and formatting. Cytoscape is a Java based desktop application available as an open source. Originally developed to be used on a biological network but is currently used for any context for complex network visualization and analysis purposes. (Shannon P et al. 2013).

Figure 7. Common Interaction Network

TOP Interactions		
Iran	Pakistan	5,810
Syria	Lebanon	9,379
Iraq	Lebanon	14,918
Iraq	Jordan	5,435
Lebanon	Jordan	9,505
US	UK	10,796

Table 5. Top 5 Interactions

Figure 7shows a network diagram that represents data from table 4, where a node is a country

and the connection is the common interactions between two countries. The thickness of the links

define the strength of the relationship i.e. the higher number common comments in two countries

the thicker the link would be.

Top interactions were filtered in Table 5 to demonstrate the highest 5 common interactions among the 13 countries. It is noticed that Facebook users showed highest interest in commenting on News account from Lebanon and Iraq, the second top common interactions is shown between US and UK news accounts.

5.2 Directed Common Interactions

	UAE	Egypt	Iran	Yemen	Syria	Iraq	Lebanon	Jordan	US	UK	Philippine	India	Pakistan
UAE		904	745	190	1,063	1,058	2,218	2,346	612	1,274	2,279	876	4,244
Egypt	904		99	125	952	1,027	1,923	1,691	113	219	15	15	56
Iran	745	99		44	470	612	580	185	2,405	2,156	286	715	5,810
Yemen	190	125	44		402	102	146	264	23	45	-	4	31
Syria	1,063	952	470	402		2,145	9,379	4,029	268	515	119	50	393
Iraq	1,058	1,027	612	102	2,145		14,918	5,435	423	1,002	4	2	60
Lebanon	2,218	1,923	580	146	9,379	14,918		9,505	527	1,732	73	31	199
Jordan	2,346	1,691	185	264	4,029	5,435	9,505		248	592	30	15	132
US	612	113	2,405	23	268	423	527	248		10,796	3,453	555	1,879
UK	1,274	219	2,156	45	515	1,002	1,732	592	10,796		3,649	560	2,384
Philippine	2,279	15	286	-	119	4	73	30	3,453	3,649		31	272
India	876	15	715	4	50	2	31	15	555	560	31		2,181
Pakistan	4,244	56	5,810	31	393	60	199	132	1,879	2,384	272	2,181	

Table 6. Directed Common Interactions

A directed notion has been defined by filling the interactions form both ends, it is noticed that the values in the table varies in scale of ten thousands and a normalization on the values was needed in order to have a clearer view on the numbers .

5.3 Normalized Common Interactions

	UAE	Egypt	Iran	Yemen	Syria	Iraq	Lebanon	Jordan	US	UK	Philippine	India	Pakistan
UAE		0.213	0.176	0.045	0.250	0.249	0.523	0.553	0.144	0.300	0.537	0.206	1.000
Egypt	0.470	-	0.051	0.065	0.495	0.534	1.000	0.879	0.059	0.114	0.008	0.008	0.029
Iran	0.128	0.017	-	0.008	0.081	0.105	0.100	0.032	0.414	0.371	0.049	0.123	1.000
Yemen	0.473	0.311	0.109	-	1.000	0.254	0.363	0.657	0.057	0.112	-	0.010	0.077
Syria	0.113	0.102	0.050	0.043	-	0.229	1.000	0.430	0.029	0.055	0.013	0.005	0.042
Iraq	0.071	0.069	0.041	0.007	0.144	-	1.000	0.364	0.028	0.067	0.000	0.000	0.004
Lebanon	0.149	0.129	0.039	0.010	0.629	1.000	-	0.637	0.035	0.116	0.005	0.002	0.013
Jordan	0.247	0.178	0.019	0.028	0.424	0.572	1.000	-	0.026	0.062	0.003	0.002	0.014
US	0.057	0.010	0.223	0.002	0.025	0.039	0.049	0.023	-	1.000	0.320	0.051	0.174
UK	0.118	0.020	0.200	0.004	0.048	0.093	0.160	0.055	1.000	-	0.338	0.052	0.221
Philippine	0.625	0.004	0.078	-	0.033	0.001	0.020	0.008	0.946	1.000	-	0.008	0.075
India	0.402	0.007	0.328	0.002	0.023	0.001	0.014	0.007	0.254	0.257	0.014	-	1.000
Pakistan	0.730	0.010	1.000	0.005	0.068	0.010	0.034	0.023	0.323	0.410	0.047	0.375	-

Table 7. Directed Common Interactions

A normalized version of Table 6 is shown in Table 7, this is done by identifying the maximum value of each country (row) and dividing the values of that row into this maximum. With this approach we constructed a normalized version of the interactions where the values are following the same scale (1 is the maximum and 0 is the minimum).

Table 7 was visualized in Cystoscope as a network; Figure 8 is showing the resulted diagram after Normalization and defining a direction notion:
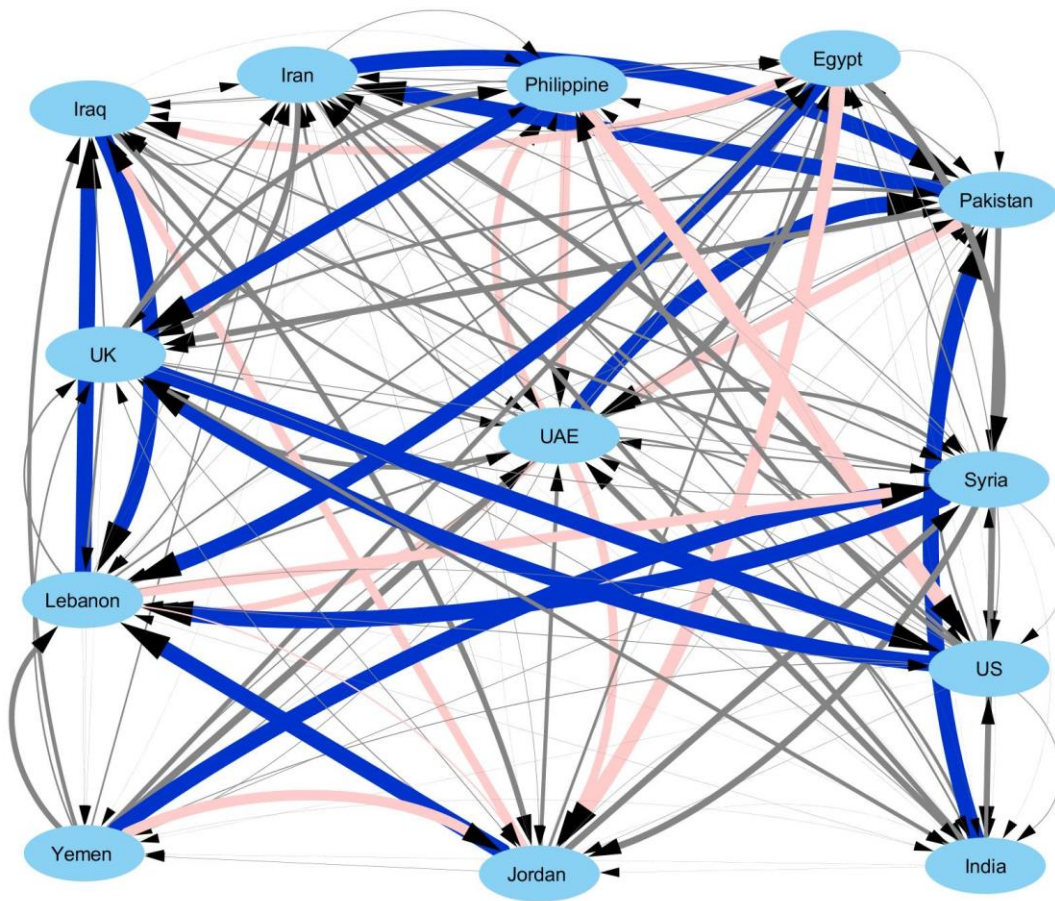

Figure 8- Normalized & Directed Common Interactions.

Each node represents a country, where the links are the common interactions between two countries. The strength/weight of the common interaction was represented by the thickness of the link. Top interactions were identified using blue arrows and those are the countries that scored 1 after applying normalization as per table 7. The countries which scored a common interaction with weights above 0.5 and less than 1 are represented with a pink arrow; Table 8 below shows a list of those interactions. The remaining links are in grey and they show the interactions less than 0.5.

Top Interactions		
Weight =1		
Pakistan	Iran	1.000
Yemen	Syria	1.000
Lebanon	Iraq	1.000
Egypt	Lebanon	1.000
Syria	Lebanon	1.000
Iraq	Lebanon	1.000
Jordan	Lebanon	1.000
UK	US	1.000
US	UK	1.000
Philippine	UK	1.000
UAE	Pakistan	1.000
Iran	Pakistan	1.000
India	Pakistan	1.000

Table 8. Top 10 common interactions with weight =1

The blue links from figure 8 representing the highest interactions that scored 1 after normalizations, table 8 show the countries in that list. It is noticed from this list that Lebanon (top interactions with 4 countries) and Pakistan (top interaction with 3 countries) are the highest countries which top interactions with other countries.

Top Interactions	
Weight Greater 0.5 and not 1	

Philippine	UAE	0.625
Pakistan	UAE	0.730
Lebanon	Syria	0.629
Egypt	Iraq	0.534
Jordan	Iraq	0.572
UAE	Lebanon	0.523
UAE	Jordan	0.553
Egypt	Jordan	0.879
Yemen	Jordan	0.657
Lebanon	Jordan	0.637
Philippine	US	0.946
UAE	Philippine	0.537

Table 9. Top common interactions with weight >0.5 and not = 1

Table 9 shows the second top interactions, those are the countries scored more than 0.5 and less than 1. Jordan is noticeable with high interactions with 4 countries.

5.4 Discussion

- UAE

  o Among the expats in UAE, Pakistan residences are the highest interacting expats to their home country news.as shown in Table 8.

  o In contrast, the least common interaction recognized is with Yemen News, this might be due to the current political situations between the two countries. Generally, we can notice that interactions with Yemen news accounts are fairly poor compared to the rest of the countries in study, difficulties in gaining internet access due to the current war can be one of the reasons.

  o From Table 9 & Figure 8 we can notice that Philippine, Lebanon & Jordan shows high interactions with UAE as well, a person who is seeking news in UAE & news in another country is assumed to be a foreign resident that is also interested in news about their home country.

- Interactions between countries with normalized weight "1" shows strong users' interest in news about both countries. Examples are as below:

  o Figure 8 shows 4 countries: Egypt, Syria, Iraq & Jordan have top interactions with Lebanon. We can refer this to the fact that Lebanon, Jordan and Egypt are countries hosting many refugees coming from surrounding countries with difficult living conditions due to wars or political struggles such as Syria and Iraq. Furthermore, many Syrian have moved to Egypt considering it was one of the countries Syrians could enter without a travelling visa hence it was an easy relocation option for those who wanted to keep their families/businesses away

from the current wars. Hence it is meaningful that a user who is commenting on a Lebanese news account is found to be also interested in interacting with news accounts of their home country.

o Similar high interaction is shown between (India, Iran, Pakistan and UAE) this can be referred to the fact that Pakistani represents the second highest number of residents in UAE's population as shown in section 3.2 . On the other side India is the neighbor country with political struggles and complications and this can explain the high level of interest between users for news accounts in both countries.

o UK & US represents two major countries in the world and news about both of them would be an common interest for a user seeking news around the world.

o Syria & Yemen are countries suffering from complicated wars resulting humanitarian crisis. Hence it would be logical that users who would show interest in news about one of them would also care about news in the other country.

5.5   Challenges

They below section show the challenges encountered in this study throughout the steps of data discovery, collection and preparations in order to achieve a successful social network analysis.

- Facebook API limitation

As discussed in section 3.1 Facebook Graph API was used to retrieve Facebook data; Graph API require a token to grant a permission to access the data. A token is automatically generated when a user logs in into a valid account. However each token has a lifetime controlled by Facebook as they apply limits on the application hitting the APIs, for example one of the limits is not to exceed 200 API calls per an hour for all application users , once a user or application hitting the PAI exceeds the limit the calls fails and returns errors. (Reuter, and Scholl 2014).

This limitation caused a lot of delays in data retrieval and sometimes caused the data collection process to start all over again as the data needs to be collected in a very close same time span window.

- Size of data

With 65 News account on Facebook, we have collected over 13 million records representing objects from posts, comments, and users, each with over 50 associated properties. Although we have broken down the exported data from Facepager per country, some countries had data exceeding 8 GB in one excel sheet which created a load on the machine used especially when data cleaning was required and preprocessing step and many changes/filtrations where done on the data

- Data Format

Retrieved data had textual format and as explained in section 4.2 & 4.3 many steps were needed for illuminating noise, duplications and correcting values in order to bring the data set into a shape where it is possible to import it in a DB for further analysis.

Chapter 6: Conclusion and Future Work

6.1 Conclusion

In this study we have used social network analysis to analyses expat community in the UAE in terms of their relationship to home countries.

The research revealed that Facebook users who have interest in UAE news have also high interest in Pakistani news which can be interpreted as UAE's Pakistani expat community are the most active social media users who shows interest in their home country news. This answers the research questions of how UAE's expats address their relation to home countries through social media and also shows that social media trends and usage can reveal real life situations and reflect current events. Furthermore the common interaction network demonstrated stronger relationship between countries suffering from consequences of political challenges and internal wars such as Lebanon and Iraq.

From the above, we can answer the research questions as below:

- Which UAE's residents shows strong attachment to their home countries through interaction with News accounts in Facebook? As shown in Figure 8 & Table 8 Pakistani are the top UAE residents who shows interest in news from their home country & their resident country. Followed by Jordanian & Lebanese who shows high activity through Facebook news accounts. .

- Can social network analysis be used to reflect political events between the studied countries? Yes, as discussed in section 5.4 the social network revealed many relationship which reflects real life political situations such as the string interactions between Lebanon, Jordan & Egypt who have high refugees from nearby countries suffering from

political events and wars such as Syria & Iraq. This finding is supported by UN published statistics which confirms that Lebanon & Jordan are both countries with highest residence for refugees from Syria & Iraq. (UNCHR Statistics 2018)

- Will social networks analysis reveal any interconnections between countries? Yes, as shown in section 5.4 it was noticed that countries with similar interest are scoring the highest levels of interactions such as: Yemen – Syria due to the internal wars occurring concurrently. Another example is the strong interaction noticed between UK-US who is global leaders with common interests.

6.2 Future work

Through this study we have found interesting results when social media data were analyzed, the below considerations can enhance the quality of work and can be a guide for future enhancements:

Selection of Countries

If the study is made on all published countries of expats in UAE there might be chances to find out further insights. The study is limited to a total of 13 countries

Selection of newspapers/accounts used to retrieve data

The study is assuming that an expat will be connected to their home country through following local news pages and interact with them through comments, however there might be other types of interests that shows further relation to home country such as following entertainment pages, celebrities..etc.

Way of interaction is comments only

As we have identified a user through their interaction with posts through comments only, there might different results if we decided to consider Likes, shares and other ways of interactions in Facebook or even combinations of them.

Time span of comments/posts

If retrieved data fallen into a different timespan there might be chances of identifying different results, especially if some events took place that results users to highly interact with user media such as events in Cinema, Festivals, or even politics changes.

Selection of Social Media Platform

We have relied only on Facebook, but further insights can be found when using more than one social media, cross analysis can be made then to compare usage of the different social media platforms.

References

Al-Jenaibi, B. (2011). Use of social media in the United Arab Emirates: An initial study. *European Journal of Social Sciences*, 23(1), pp.84-97.

Alsudais, K., & Corso, A. (2015). GIS, big data, and a tweet corpus operationalized via

natural language processing. *AMCIS 2015 proceedings*.

Batrinca, B. and Treleaven, P.C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*(1), pp.89-116.

Central Intelligence agency (CIA). (2018), *The World Fact Book, Middle East, United Arab*

Emirates[online]. [accessed online 16 Feb 2018] available at:
https://www.cia.gov/Library/publications/the-world-factbook/geos/ae.html

Dennis, E.E., Martin, J.D. and Wood, R. (2017). Middle use in the Middle East, 2017: A six-

nation survey [online]. *Northwestern University in Qatar*. [Accessed online 1 Sep 2018]

www.mediaeastmedia.org/survey/2017.

Fan, W. and Gordon, M.D.(2014). The power of social media analytics. Communications of

the ACM, 57(6), pp.74-81.

Federal Competitiveness and Statistic Authority (FCSA). (2018*). Statistics By Subject:*

Population Estimates 2006 – 2010 [online]. [Accessed 10 Feb 2018] . Available at:

http://fcsa.gov.ae/en-us/Pages/Statistics/Statistics-by-Subject.aspx

Forstenlechner, I. and Rutledge, E.J. (2011). The GCC's "demographic imbalance":

Perceptions, realities and policy options. Middle East Policy, 18(4), pp.25-43.

Froilan T. Malit Jr. and Al Youha.A. (2013). Labor Migration in the United Arab Emirates: Challenges and Responses. Migration Policy Institute (MPI) [online] 18 September [Accessed 15 Feb 2018]. Available at: https://www.migrationpolicy.org/article/labor-migration-united-arab-emirates-challenges-and-responses

Gardner, A.M. (2015). Migration, labor and business in the worlding cities of the Arabian Peninsula. *Institute of Developing Economies*, pp.1-11.

Global Media Insights (2018). United Arab Emirates Population Statistics [online], 11 January.[Accessed 15 Feb 2018]. Available at:

https://www.globalmediainsight.com/blog/uae-population-statistics/

Griffiths, M., & McLean, R. (2015). Unleashing corporate communications via social media: A UK study of brand management and conversations with customers. Journal of Customer Behaviour, 14(2), 147–162

Haßler, J., Maurer, M. and Holbach, T. (2014), Advancement through technology? The analysis of journalistic online-content by using automated tools1.*SCM 2/2014*

Jabeur, N., Nait-Sidi-Moh, A. and Zeadally, S. (2018). Crowd social media computing: Applying crowd computing techniques to social media. *Applied Soft Computing*, *66*, pp.495-505.

Jünger, J and Keyling, T (2018). Facepager. An application for generic data retrieval through APIs [online].[accessed online Feb 1 2018] available

at: https://github.com/strohne/Facepager/.

Lee, I. (2018). Social media analytics for enterprises: Typology, methods, and processes. *Business Horizons*, *61*(2), pp.199-210.

Misirlis, N. and Vlachopoulou, M. (2018). Social media metrics and analytics in marketing– S3M: A mapping literature review. *International Journal of Information Management,* 38(1), pp.270-276.

Radcliffe, D. and Lam, A. (2018). Social Media in the Middle East: The Story of 2017.

Reuter, C. and Scholl, S. (2014), September. Technical Limitations for Designing Applications for Social Media. *In Mensch & Computer Workshopband* (pp. 131-139).

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape (2013). A Software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov; 13(11):2498-504

Shao, H., Zhang, Y. and Li, W., 2017. Extraction and analysis of city's tourism districts based on social media data. *Computers, Environment and Urban Systems*, 65, pp.66-78.

Statistic Center Abu Dhabi (SCAD). (2018). *Population & Demographic Statistics*[online][Accessed 11 Feb 2018]. Available at: https://www.scad.gov.abudhabi/en/pages/statistics.aspx?topicid=24

Stieglitz, S., Dang-Xuan, L., Bruns, A. and Neuberger, C. (2014). Social media analytics. *Wirtschaftsinformatik*, *56*(2), pp.101-109.

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. (2018). Social media analytics– Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, pp.156-168.,

The United Nations High Commissioner for Refugees (UNHCR) (2018). *Refugee Statistics* [online] [Accessed 1 Sep 2018] Available at https://www.unrefugees.org/refugee-facts/statistics/#

Van Dijck, J. and Poell, T. (2013). Understanding social media logic. *Media and Communication (2013)* V1, P 2–14

Zeng, D., Chen, H., Lusch, R. and Li, S.H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, *25*(6), pp.13-16.