

Migration in the internet age

The 21st century: When refugees went online

Oskar Lif

Abstract

This paper provides a new take on how to explain irregular migration. The results of this study takes a leap forward in scientific thinking of how large irregular migration flows can be understood in a modern age where migrants are using internet technology to their advantage. By using existing migration theory and big data search patterns from google a completely new way of understanding irregular migration is developed. This longitudinal quantitative cross-country study gives new analytical insight in how irregular migration can be explained by analysing potential and actual migrant's internet behaviour. This paper is the first of its kind to use Google search patterns as a determinant for where and when migrants journey by combining it with time series of number of asylum seekers yielding a successful way to predict large irregular migration. Data from eight different receiving countries and Europe as whole is used in combination with data from two transit countries and five sender countries. It's argued that to understand the 'Age of migration' we must see how it relates to the 'Information age' and that these two global phenomena must be seen in connection and not as separate from each other if we are to understand modern migration.

Key words: Irregular-migration, Internet, Search-patterns, Google, Network

Words: 9996

Table of contents

1	Introduction.....	1
1.1	A Short Story.....	1
1.2	Question & Aim	2
2	Definitions.....	3
3	Theory	4
3.1	Irregular Migration.....	4
3.2	Literature Review	5
4	Theoretical Approach.....	7
5	Methodology	8
5.1	Choice of Methodology & Research Design.....	8
5.1.1	Explaining the Internet Variable	9
5.1.2	Google Trends & Correlate & Trends-Question	9
5.1.3	Operationalizing of Variables	11
5.2	Empirical Overview: Data Collection & Procedure.....	13
5.3	Statistical Analysis: Data Collection & Procedure	14
5.3.1	Data Collection.....	14
5.3.2	Sender & Receiving Countries	15
5.3.3	Other Procedure.....	15
6	Hypotheses.....	16
7	Results & Empirical Overview	17
7.1	Report.....	17
7.2	Empirical overview: Eastern Mediterranean Route	18
7.2.1	Google Correlate: Turkey.....	18
7.2.2	Google Trends: Turkey	19
7.3	Empirical overview: Syria.....	21
7.3.1	Google Trends: Syria	21
7.4	Statistical Test Results	21
7.4.1	Syria	22
7.4.2	Syria-Sweden	25
7.4.3	Afghanistan	27
7.4.4	Nigeria and Senegal	27
8	Analysis.....	28
8.1	Analysis: Development of theory.....	29
8.1.1	Nigeria & Senegal	32
8.1.2	Generalizability & Validity.....	32

9	Conclusion	33
10	References.....	34
11	Annex	39

1 Introduction

The refugee crisis of 2015 and its large amount of migrants travelling to Europe (BBC, 2016a; Nordland, 2015) is by far the largest case of irregular migration in Europe since world war II (Frontex 2016; UNHCR, 2016). Yet this is only a smaller piece of the puzzle in the larger picture, in which migration flows are becoming more and more complex. Particularly irregular migration flows are picking up speed in the new ‘Age of migration’ (King 2002; King et al. 2010; Martiniello and Rath 2012). This paper develops a new way of describing and explaining large irregular migration flows and by studying the role internet behavior has on irregular migrants unprecedented evidence is generated for how to predict and explain irregular migration. This is primarily done with a longitudinal quantitative cross-country method which uses linear regression as a statistical tool.

1.1 A Short Story

In November 2015 during the peak of Syrian asylum seekers in Sweden and the culmination of the refugee crisis in Europe, I went on a train. It departed from the southern Swedish city of Malmö, the city which most migrants first arrive too in Sweden just at the border to Denmark. The train was headed to the capital of Sweden, Stockholm. I went onboard in Lund 10 minutes north of Malmö and next to my assigned seat sat a woman. She smiled at me and asked with a broken English if I could help her connect her phone-charger to the outlet. I helped her and afterwards I asked where she came from. She replied she’d arrived in Sweden less than two hours ago and that she’d fled from Syria many weeks ago. Our conversation continued and I helped her connect her smartphone to the Wi-Fi onboard the train. I asked her if she would call her family or friends now but she shook her head and said: No! First I have to update Facebook and say I’m finally in Sweden!

1.2 Question & Aim

The research question is:

- How to explain and describe irregular migration by using internet behavior?

An important thing to recognize when deciding how to answer this question scientifically is to have an awareness of what type of question it is. Only then can important methodological choices of collecting evidence begin. The question fulfils what Teorell and Svensson defines as the three principles for a scientific question. 1: Scientific *cumulative*. 2: Scientific *relevance*. 3: Societal *relevance* (Teorell and Svensson, 2007). The question builds upon previous research and much developed theory on the subject of migration and irregular migration (Hooghe et al, 2008; King, 2012). Even though this field has well developed theory the incorporation of modern technology such as internet is seemingly nonexistent, this makes the question relevant study. Lastly, the world is seeing the largest migration flows since the Second World War II. Together this satisfies the three principle and portrays how relevant and needed increased knowledge about the question is.

The principle aim is to develop theoretical understanding about forces that drives irregular migration. The research question is operating under the preconceived notion that migration can be explained to some extent with internet technology and therefore an answer to the question needs an explanatory mechanism for how this relationship looks (Teorell and Svensson, 2007). It's a question that's freed from the boundaries of time and space so a valid answer to it has generalizable implications (Teorell and Svensson, 2007). Much is included in this question since its abstraction level is high. This is a deliberate choice which increases the burden of proof needed to answer the question adequately. In this thesis extra attention is given to methodology because of two reasons. Firstly, because this paper is breaking new ground it's important to provide full insight in how and why the research design has been devised as it has. Secondly, since the used methodology has in comparison to more established methodology fewer clearly constructed approaches and therefore the need to explain certain aspects in more depth is important.

2 Definitions

Phrases like economic migrants, refugees, guest workers etc. are all names for international migrants. But it's clear that the circumstances and the reasons underlying the causes for why a person emigrates are vastly different, which makes the need for precision and clarity necessary in order to avoid vagueness and conceptual stretching (Teorell and Svensson, 2007). Migration is defined as *“a person who is outside the territory of the State of which they are nationals or citizens and who has resided in a foreign country for more than one year irrespective of the causes, voluntary or involuntary, and the means, regular or irregular, used to migrate”* (EMN 2014). What is then *irregular* migration? Migration can be divided into two categories, regular and irregular. This definition is relevant since the distinction between legal and illegal immigration questionable (Bommes and Sciortino, 2013; Gambino, 2015). Regular migrants are defined as those migrants who emigrate and immigrate in accordance with the regulation of a state such as guest workers, quota refugees, exchange students etc. (EMN, 2014). Irregular migrant is defined as:

“A person who, owing to irregular entry, breach of a condition of entry or the expiry of their legal basis for entering and residing, lacks legal status in a transit or host country. In the EU context, a third-country national present on the territory of a Schengen State who does not fulfil, or no longer fulfils, the conditions of entry as set out in the Schengen Borders Code, or other conditions for entry, stay or residence in that Member State” (EMN, 2014, 172)

The term irregular migration is used to describe the flows and behavior of irregular migrants in this paper. Irregular migration shouldn't be confused with the term illegal migration even though they are often used synonymously (Jandl, 2007). Irregular migration is not by default illegal (Triandafyllidou and Marouki, 2012) even though illegal migrants often are irregular (EMN, 2014).

3 Theory

3.1 Irregular Migration

Regular migration is controlled inflows of migrants that behaves more or less in accordance with policy-regulation. Irregular migration is in this paper considered to be cases of migration that falls outside of such controlled migration and this approach shows clearly how irregular migration is linked to regulation and policy making (EMN, 2014). Irregular migration can be understood as migration flows where the regulative *policy-output* from countries doesn't match the *outcome* of the policy. The difference between output and outcome is known as a 'policy-gap' (Hollifield, Martin and Orrenius, 2014). Policy output is defined as desired policy consequence and policy outcome are the actual consequences. Thus irregular migration is migration which isn't managed as intended or migration which there aren't any clear policies for (Jordan and Düvell, 2002). The 'policy-gap' concerning immigration policies and border controls is in general much larger than in other political areas (Hollifield, Martin and Orrenius, 2014). Even though western European countries since the 1970s have adopted restrictive legislation to control migration flows this hasn't prevented an increased amount of asylum seekers (Zlotnik, 1998).

A growing problem since Hollifield et al. (2014) has shown that the output-outcome gap in this policy field is currently increasing. The reasons why the 'gap' is larger in this policy area than others are complex and array from a plethora of different reasons and this is outside the scope of this thesis. However, one reason why outcome doesn't match output is because policy drafters don't have an adequate understanding of the underlying forces of what drives irregular migration and this article could play an important role to mitigate this. Some authors such as Bhagwati (2004) has urged that better policies in order to better manage immigration are badly needed, something that relates to the gap problem. Others have stressed the importance of understanding the role of sender, transit and receiving country as to be the key in order to effectively handle irregular migration flows (Jandl, 2007). The reason for why knowledge about irregular migration is needed is because it affects how effective policies political institutions can draft. Increased understanding about the topic would reduce the uncertainty and increase the ability to handle the complexity for academics and practitioners (Frontex, 2016).

3.2 Literature Review

Two of the most dominant theories in the field of migration are ‘push-pull’-theory and migration network theory. Generally, these different theoretical approaches don’t contradict each other and can be useful to complement each other (Hooghe et al, 2008; Massey et al, 1998). They both share one problem though which is that the literature in this field poorly explains why migrants chose to settle in particular countries and why they avoid others (Venturini, 2004).

Push-pull theory is influenced by economic neoclassical thinking (King, 2012) and is probably the most know approach to international migration (Tomanek, 2011) and it has been applied with reasonable success (King, 2012). The essence of the theory uses different factors, mostly related to economics, and states that these have either a pulling or pushing effect on migrants (King 2012). The driving causes of migration are in a broad sense well known as push factors and is ‘forces’ like violence, poverty, human rights violation and so on. These push factors works in combination with pulling factors which are forces like economic opportunity and safety (Crisp, 1999; Hatton and Williamson, 1994; 1998; King, 2012; Massey et al, 1998). What’s less known in push-pull is why migrants choose country A over country B (Hooghe et al, 2008). Push-pull has been successfully used from a macro-perspective describing larger trends over time (Hooghe et al, 2008) and the theory can to some degree explain why people leave regions undergoing rapid social and economic change (Castles and Miller, 1994). From a macro level the theory should be understood as a supply and demand theory, and from a micro perspective it describes how individuals work out cost-benefit analyses to calculate the “cost” of migrating versus the reward (Borjas 1995; Hooghe et al, 2008). In push-pull theory personal factors and “intervening obstacles” also plays a relevant role, these obstacles can be geographic, information deficits and regulating policies like border controls (King, 2012; Hatton, 2004).

The theory has been criticized for its simplicity but is generally considered to be useful (Massey et al, 1998). The large weakness of push-pull theory is that it has failed to explain why countries with the same structural economic conditions have such varying out-migration and it’s been criticized for not considering social factors, socio-cultural factors and the political reality migrants face enough (Arango, 2004; Massey et al, 1998).

The second theory reviewed is social/migration network theory. It’s a systemic approach and has been hailed as providing a comprehensive framework for studying migration (King, 2012). The core of this theory states that people in the sender origin country are connected with current and previous emigrants and these connections can be described as social networks that transcend borders and contribute to mediate migration by reducing the costs and risks by providing social capital, which lowers uncertainty for migrants (Massey et al, 1998). Massey et al, defines the networks migrants form as “sets of interpersonal ties that connect migrants, former migrants and non-migrants in origin and destination areas through ties of kinship, friendship and shared community origin” (Massey et. al. 1998). Ties can be ‘loose’ or ‘close’, close ties are when relationship are more intimate, such as family bonds (King, 2012).

Transnational networks create social capital which facilitates additional movement within the network (De Haas, 2010) something that over time, through cumulative causation, causes *chain migration* (Hooghes et al, 2008; Massey et al, 1998). The theory has been vital in developing the understanding of the phenomena of chain migration and how migration flows over time are self-feeding and self-regulating (Faist, 1997). Network theory explains how migration patterns adjust to 'shocks' in the system (King, 2012). Why there's a 'cumulative causation' and the reason for why the networks with time perpetuate more migration is because networks expand (Massey et al., 1993) and have a motivational effect on potential migrants (Wegge, 1998). This can cause expectations to rise for prospective migrants because "the success story tends to be further magnified" by the diaspora (Ghosh, 1998, 67). These Networks facilitate information acquiring about policies and institutions in the destination country (Munshi, 2003) and moreover, the positive effect networks have for the irregular migrant's prospects are large (Blaschke, 1998).

Migration networks help establish and facilitate organizational support, both by NGOs, private organizations and smugglers. This can be understood as 'organizational infrastructure' (Castles and Miller, 1994:25; Massey et al, 1998; Massey and Van hear 1998, Ghosh 1998; Castell, 1998). Moreover, smuggling networks are fostered by stricter border policies (Triandafyllidou and Maroukis, 2012). On the point of irregular migrants there's evidence that the behavior of asylum seeker and/or refugees can be explained well with network theory (Crisp, 1999) and some authors claims that the effects from networks are crucial for understanding irregular migration, especially when the possibility of legal entry is removed (Ghosh, 1998).

4 Theoretical Approach

The purpose for this paper is to develop existing theory to include internet technology as a factor. I take an interdisciplinary approach because the shortfall of limiting oneself in this field of study has held back development of theoretical understanding (Favell 2008; Brettell and Hollifield 2008). Before the theoretical approach is presented epistemological and ontological assumptions ought to be disclosed. This paper is written from a point of view of critical realism (Hollis, 2002) and ontological individualism (Teorell & Svensson, 2007). I write from a background of political science and build upon research from this field and from economics, international relations and sociology.

The theoretical approach is to test hypotheses derived from push-pull theory and migration-network theory. Hence a hypothetico-deductive method lies at the foundation for methodological choices (Teorell & Svensson, 2007). Hence the research design will be adjusted in order to enable hypotheses testing of the ‘internet variable’ (IV).

It’s clear by reviewing past literature that there’s a considerable lack of theoretical understanding in how internet usage affects migrants. I don’t mean to disregard past literature, I mean to do exactly the opposite. In this paper I use the well-developed theoretical mechanism’s that’s been used to explain migration in my attempt to develop theory. This is made by integrating the IV *within* these existing theoretical frameworks. By using the logic of e.g. migration network-theory it’s possible to create hypotheses that can confirm or reject my new methodology with existing theoretical understanding. If it’s rejected it implies that my method isn’t useful but if the method can generate results which confirms the ‘chain of relations’ already established it implies that this new method can develop existing theory.

A brief comment is needed on why two theories are used since this could hinder depth and development of one theory. I recognize that this is true but if the new variable is relevant for explaining migration it should add explanatory power to theories that do so already. Thus by not being bound to one theory the results from my new method are partially safeguarded towards the objection that the variable only works within a certain framework. By showing that it works with both of them it implies the broader impact this new approach has and its wider potential for developing theory.

If the IV is judged to have been proven it’s important that the research design is well adapt to fulfill what Teorell and Svensson defines at the four causal criteria (2007):

- i. Mechanism. The chain of relations which explains the connection between variables
- ii. Order in time. Does cause precede effect?
- iii. Counterfactual difference. If the independent variable isn’t present, the variation in the dependent variable would differ.
- iv. Isolation. All else being equal the causal effect from the independent variable can be isolated from alternative explaining factors.

5 Methodology

5.1 Choice of Methodology & Research Design

Syria is the main country of study for this paper and it's because the large irregular migration from Syria provides a suitable case for testing. If my method fails to explain such a critical and large-scale case, seemingly, its potential to explain any other type of irregular migration is severely reduced. The methodology adapts an intensive-extensive approach by combining qualitative depth with the wider scope of quantitative methodology, an approach which gives the research design a strong foundation for generalizable results (Teorell and Svensson, 2007; Bogdan and Biklen, 2006). The study of Syria alone can almost be described as a case study and in this sense the study is *intensive*. But the enlargement of the study by including five sender countries, nine receiver countries/regions and two transit states captures the benefits of a quantitative study which enhance the possibility to make generalizable claims (Teorell and Svensson, 2007).

I conduct a cross country longitudinal quantitative statistical analysis, by using linear regression; a bivariate linear model is built with an independent variable of 'Google queries' that predicts the variation in the dependent variable. The dependent variable is the total amount of asylum seekers for a specific period of time in a receiving country. The independent variable is Google searches made in the sending country about the receiving country. Before the quantitative analysis a comprehensive collection of empirical evidence is gathered in order to give an empirical overview of how internet behavior relates to migration. This is because a valid explanation needs to have an empirically correct description of the world, *especially* for a quantitative study (Teorell and Svensson, 2007).

In addition to triangulate and validate the asylum data, Frontex data is used to confirm the direction and variations in the flows and this enables us to control if similar patterns can be seen with another measurement as well to validate the operationalizing of the IV.

The research design draws from the logic of a most different design (Teorell and Svensson, 2007). *Receiving* countries/Europe was chosen on the basis if they'd received more than 10,000 asylum seekers during 2015 from respective sender state. The most different design seeks variation in 4 criteria for the *sender* countries.

- i. The country had to have 15,000 asylum seekers or more during 2015
- ii. The country needed to be among the top 20 sender countries of asylum seekers to Europe during 2014-2015.
- iii. Internet usage.
- iv. Geographic location.

Variation in these factors enabled me to see if search-data could predict irregular migration flows at different volumes. While at the same time maintaining geographic diversity. Why for example Senegal was chosen instead of Iraq. The countries had to have large variations in their number of internet users per capita since variation in this variable could be assumed to have an effect on the validity and precision of the IV used in the quantitative analysis. I.e. a higher amount of internet users per capita could plausible effect the strength of the predictions in the linear regression. The countries were geographically differently located since this affect what obstacles and ‘organizational infrastructure’ they’re exposed to. The potential for generalizable conclusions is helped by the fact that if similar phenomena are observed from different countries with different conditions this indicates that potential results are applicable to irregular migration flows in general (Teorell and Svensson, 2007). Thus generalizability can be claimed with higher certainty.

Furthermore, economic conditions, Syrian Diaspora and conflict levels are displayed in section 7 since these factors can affect migration flows.

5.1.1 Explaining the Internet Variable

The research question operates under a preconceived notion that migrants make use of internet technology. I make the assumption that internet is used to access information and that this influence migrants behavior. The claim is that these two are related and that the effects of using internet has either a moderating and/or mediating impact irregular on migration flows. I act under the assumption that it’s conceivable that internet is used to communicate with one’s social network and that this could reinforce the expansion of migrants’ networks.

Most importantly I assume that google searches in a particular country express interest/intention about the searched for country. The dependent variable is an irregular migrant and because of the connection-chain I hypothesize that the migrant will have its behavior affected in some way by the IV.

5.1.2 Google Trends & Correlate & Trends-Question

Now Google-Trends and –Correlate will be explained. Even though this is a new method in comparison to other data colleting tools and type of measurements that’s been used in migration theory, G-Trends and G-Correlate aren’t new tools in social science. The methodology used in this paper is built upon earlier usage of it in political science and economics (Askitas, 2015; (Trevisan, 2014; Mellon, 2013; Choi and Varian, 2009) and guides on how to use Google data (Varian and Stephens-Davidowitz, 2015). These data sources have in general been accepted as very useful for social science (Mellon, 2013; Varian and Stephens-Davidowitz, 2015; Trevisan, 2014) and G-Trends has successfully measured people’s intentions (Stephens-Davidowitz, 2012).

What Google trends do is that it “provides an index of search activity on specific terms and categories of terms across time and geography “(Varian and Stephens-Davidowitz, 2015, 1). It makes it possible to see how popular a search-term is over time by looking at how

many people search for it (Google, 2016). The index measures the *fraction* of search queries that includes a specific search-term. This is done either on a national or regional level. The maximum of the index is 100 and the minimum is 0 (Varian and Stephens-Davidowitz, 2015; Google, 2011) E.g. if the word “refugee” is the search-term and searches made from Syria for the time period 2015.01 to 2015.12 the highest data point that satisfies the condition of the query during that time could be indexed at 100 in June. If the index number then is 50 in August this means that the *fraction* of searches for “refugee” of all queries that month made from Syria was half of what it was in June. This raises some issues. If a smaller region has a higher index for a term in comparison with a more populated region this means that the term is more popular as a percentage of the total amount of searches in each region but not that the term in absolute terms is more popular (Varian and Stephens-Davidowitz, 2015). Generalizability is helped by the fact that this method includes the full group of Google searches and not individual samples.

The advantages Google Trends provides are that it’s possible to see if a search-term is becoming more popular in a region and then to know that this isn’t because the region’s absolute numbers of searches are rising and that is what has caused the perceived increased interest. Instead the indexation is adjusted for absolute volumes in order to make comparisons over time possible and to compare relative popularity (Google, 2011; Varian and Stephens-Davidowitz, 2015). The disadvantages with this indexation is that we can’t say that even if the index for a search-term is higher in country A than in Country B it means the search-term is more popular in country A. The methodology has been adjusted to this limitation and this is the reason for why comparison between receiving country A, B and C between sender countries X, Y, Z etc. aren’t made.

Google Correlate is used to check the operationalization’s validity. Correlate works like Trends but in reverse. In Trends you type a query and receive its time-series frequency. In Correlate you type in a query then it shows how other query’s time-series correlates with it I.e. they have similar patterns over time. Google generates these correlations using Pearson’s Correlation Coefficient (Google, 2011). I use Google Correlate as a complementary method for several reasons. Firstly, it indicates if the search-terms are relevant, thus it can validate or reject the operationalizations of the IV. It does so by showing what other search words are used and clustered with the search-term. The reason only Turkey is chosen isn’t by my deliberation but has to do with limitations set by Google and data for other countries that are of interest isn’t available (Google, 2011). This prevents me from using the method in Libya, Syria, Nigeria, Senegal, Somalia and Afghanistan. This isn’t a substantial disadvantage however; Turkey is an excellent testing ground due to its properties. 1: there is a large group of displaced Syrians in Turkey, approximately 2,5 to 3 million in 2015 (UNCHR, 2016). In Turkey the language of Arabic isn’t spoken by the majority population. This means that a large proportion of all Google searches that are made in Arabic will be made by Syrians or other Arabic speakers like Iraqis whom are a significant refugee group in Turkey (UNHCR, 2016).

Practically Trends works in the following way: if one word is used as a search-term Google Trends will include all Google searches that includes that term, even if it’s used in a longer sentence. It’s possible to combine two or more terms by inserting Term1+Term2 and then it aggregates the data in its indexation for when either one of these was searched for.

This has the advantage that when the term ‘Sweden’ is used for operationalizing informational gathering for Sweden almost all searches that was intended as such are likely included. However, there are ambiguity problems because it’s not known if the search-term was used to collect information about Sweden or for something else just related to Sweden (Varian and Stephens-Davidowitz, 2015). How large this ambiguity issue is difficult to say but the results suggest it has a low impact.

The search-terms validity is also cross-checked using Trends-Question. This method controls the accuracy of the operationalizations. With it the top questions that’ve been asked by people as they have searched for search-term X in Google are provided by Trends. This means that for the collective of people in Syria that e.g. googles Sweden in Arabic, google generates the top questions they asked in connection and relation to that search (Google, 2016). Google questions are only generated from Syria. For results on operationalization accuracy see section 7.

5.1.3 Operationalizing of Variables

An irregular migrant is operationalized as an asylum seeker in the statistical analysis. The main reason for this indicator is because there exists reliable and comparable data for all Europeans states on the amount of asylum applications per month (Eurostat; 2016). This means that the dependent variable is a time series of the total amount of asylum seekers from origin X for one month, or one week in the case of Sweden. In the empirical overview an irregular migrant is operationalized as a person detected by Frontex to have illegally crossed a border. I.e. people crossing without necessary documentation. I’m well aware that this is a different measurement of irregular migrants but the Frontex data isn’t chosen to measure the same way but because it’s a different indicator of irregular migration flows. Which enables us to see if similar patterns can be detected in transit countries with my new method? The reason for why only Eastern Mediterranean Route (EMR) which is a sea route was selected is because a larger proportion of the migrant flow is believed to be detected there then in comparison with other routes and because this was the route with the highest number of Syrians detected (Frontex 2016; 2015).

The IV was operationalized by using a term that was assumed to be queried to indicate interest for target countries in the quantitative analysis. Accordingly, the most straightforward way was to use the name of the country and this was done for Syria, and partially for Nigeria. However, for Senegal, Afghanistan, and Somalia the selection process couldn’t be done similarly because Trends data on individual target countries wasn’t available because of too few searches. Instead only search-data on Europe as a whole was available and consequentially the amount of asylum seekers from respective country was gathered for Europe as a whole and not by the countries who had 10000 or more asylum seekers. Ideally these countries would’ve had the same selection as Syria. But it’s not a major problem that the term ‘Europe’ is used instead since the logic is the same as for Syria, but it’s probable that the IV will be less accurate in these sender countries, but even so the longitudinal variation in the flow is still captured.

In Syria terms was translated to Arabic. Nigeria, Senegal and Afghanistan are different from Syria because they don't have one clear predominant language as in Syria so for them all major minority languages was also included. That is the same term but translated to the different languages. All translations were first done via google translate. Because of ambiguity problems with automatic translation all Arabic words were checked by a native Arabic speaker. All translations were test-googled to see check spelling and if they gendered relevant or irrelevant results, e.g. by checking how prominent sites like Al Jazeera spelled. All terms checked against how the country had been spelled on it's to the Country-Wikipedia page. Now there are some issues with this way of operationalizing and using Trends concerning under-covering and over-covering of the target population of irregular migrants. It covers too little because we don't know that migration network make use of these to a significant degree. Mainly it covers too much. I.e. many people might google without actually leaving. This is partially why the results for Syria are triangulated with the other countries.

Operationalizations of search-terms in Google

Table 1. Syria

Syria Language: Arabic	Aus	Bel	Ger	Hun	NL	Nor	Swe
Term	الانتمسا	ب ل ج يكا	ألمانيا	الأمجر+هنغاريا	هولندا	النرويج	السويد

Ambiguity on how to spell for Hungary resulted in a combination term of two different spellings. This is more likely to catch the full scope of google searches for Hungary.

Table 2. Nigeria

Languages: English+Hausa+Yoruba+Igbo. The spelling is the same in Yoruba, Igbo and English.

Nigeria	Italy	Europe
term	Italy + italiya	Europe + turai

Table 3. Senegal. For other minority languages no translation was found

Senegal: Languages: French+Wolof+Arabic	Europe
term	Europe + Tugal + أوروبا

Table 4. Afghanistan. Dari spelling couldn't be found so Persian was used instead.

Afghanistan: Languages: Pashto + Persian	Europe
term	اروپا + د اروپا

The operationalizing of using 'Europe' as a term to indicate interest is questionable in comparison to the terms used for Syria. Nevertheless this negative impact on the validity the terms works reasonably well, see section 7.4.4

Operationalizations for the search terms in the Empirical Overview

Table 5. Google-Correlate

Country: Turkey	Language: Arabic	
terms	Sweden: السويد	Germany: ألمانيا

Table 6. Trends-Question

Country: Syria	Language: Arabic		
Term	Sweden: السويد	Germany: ألمانيا	Immigration: الهجرة

Table 7. Frontex-data

EMR	Turkey; regions of Canakkale and Izmir in turkey
Language: Arabic	
Term	Map + sea + smuggler + broker + asylum + Europe + Sweden + Germany ألمانيا + السويد + أوروبا + بحر + خريطة + المهربين + وسطاء + اللجوء حق

5.2 Empirical Overview: Data Collection & Procedure

Data in section 7.2.2 shows the Eastern Mediterranean Route (EMR) as it's defined by Frontex and the measures are number of illegal border crossings (BCP) made by Syrians per quarter. Turkey to Greece are the two main transit states that most Syrians pass on this route, it's also the route a majority of Syrians takes and thus the data are collected from the regions of Izmir and Canakkale in Turkey (2015; 2016). It would've been preferable to have the Frontex data sorted accordingly to months as a unit of analysis but no such data was available, why no statistical testing was done with this dataset. Google correlates are as mentioned only used in Turkey. All correlations with non-Arabic search-terms was removed since these most likely are spurious which means that the big data generated from turkey are generally not by Turkish people which removes noise in the big data. The seven highest correlate words that were deemed relevant was picket out from the top 20 results.

5.3 Statistical Analysis: Data Collection & Procedure

5.3.1 Data Collection

The data used for the linear regression consists of two different kinds of time series. Monthly data of the number of asylum seekers are collected from Eurostat. In the case of Sweden weekly data of Syrian asylum seekers is also used and this data is from the Swedish migration agency and was obtained via personal request. It provides the advantage of drastically increasing the N in the linear regression. Though this is the same data-source which Eurostat use so reliability isn't negatively impacted.

Google Trends and asylum data was gathered for the period of 2011.01.01 until 2016.02.31. But both asylum data and Trends data for 2011 and 2012 was removed from the statistical testing in all cases except for Sweden and Germany due to very small numbers. I.e. the indexation for each operationalized search-term for each individual country was gathered. E.g. for Sweden to the search-term: السويد was used in order produce the Trends time-series index from Syria. That means that the tables in section 7.4 only says 2015. Trends-data data has been gathered for that year and that country *only*, this in order to adjust for variations over time. If it says 2013-2015 that means Trends-data was gathered for that whole period for that particular country. Trends data was rid for noise in three instances where distinct indexed spikes could be witnessed that was judged not relevant and that was for 1, the world cup in Football 2014 and 2, the European championship in football 2012. 3, for a football friendly between Nigeria and Italy. Data from the receiving countries that had national teams competing in these tournaments was removed for that time period.

2014 teams: Germany, Belgium, Italy, Netherlands (UEFA, 2016a).

2013: Nigeria played a friendly against Italy (Nairaland, 2013)

2012 teams: Sweden and Germany (UEFA, 2016b). Other countries were in the competition but only Sweden and Germany are tested 2012.

5.3.2 Sender & Receiving Countries

Table 8. Criteria for selection

	Individuals using the Internet %	Geographic location	GDP per capita	Peace index	Asylum seekers in	Rank	Primary Route
Afghanistan	6,4	Central Asia	\$2,000	160	238 935	2	EMR
Nigeria	42,7	West Africa	\$6,400	151	53 590	8	CMR
Senegal	17,7	West Africa	\$2,500	49	16 120	20	CMR
Somalia	1,6	East Africa	\$400	157	40 985	12	CMR
Syria	28,1	Middle East	\$5,100	162	511 630	1	EMR

Internet usage data is gathered from the Worldbank (2016). GDP per capita is based on the most recent estimate (CIA, 2016). Peace index data ranks how non-peaceful a country is and the range is from 1-162 with 1 being the most peaceful (Institute of economics and peace, 2015). Asylum seekers data is from Eurostat (2016). Rank is based on 2014-2015 where 1 is the rank of the country from which Europe received the most asylum seekers (Eurostat, 2016). Primary route is based on 2015-2014 Frontex detections (Frontex 2015; 2016).

Table 9. Syrian Diasporas

Austria	Belgium	Germany	Hungary	Netherlands	Norway	Sweden
1,865	1,809	32,878	-*	10,659	<2000**	22,357

*The diaspora in Hungary isn't 0 but it was smaller than 1 of 100,000.

**Data for Norway isn't available from Rand and is an approximation (ReliefWeb; 2013; Eurostat, 2016; SSB, 2001)

Table 9 depicts the Syrian diaspora as of 2011 in the seven receiving countries. Data is from the RAND institute (Taylor et al, 2014).

5.3.3 Other Procedure

All questions, queries and correlate results was made in incognito surf mode and from a VPN based in Syria to avoid personalized search results (Physioc, 2015). Somalia was after timeline review deemed unreliable due to what appeared as complete randomized results. In comparison it was possible to measure results from Afghanistan which shows a limitation of google data and suggest that there is a critical point between 1,6%-6,4% internet usage in a population to generate relevant results from Trends.

6 Hypotheses

In order to be able to explain a question like this a relationship of causation is needed between what is affected and what is affecting. This can be expressed as having a *mechanism*. The mechanism is what is used to describe the connection between different variables. (Teorell & Svensson, 2007). This means that the *independent* variables IV are linked with the *dependent* variable. If they are in fact linked and the operationalizing of the independent IV and the depended variable are relatively accurate the quantified indicator will have testable empirical consequences (Teorell and Svensson, 2007).

In the linear regression analysis, the independent variable X is the index of google searches per time unit. For example, 100 in September 2013 for Syria to Sweden. The dependent variable Y is the total amount of asylum applications made by people from country *a* in country *b*. For example, 3090 asylum applications made by Syrians in Sweden during December 2015.

Hypotheses

Migration network:

- i. If there is a chain of connection between the two variables, there will be a significant correlation between internet searches and number of irregular migrants.
- ii. If Search patterns in sender countries increase or decrease before the migration flow increase or decrease internet behavior can predict when and where migrants journey.
- iii. Migration-networks adjust to ‘shocks’ to the system if news about a receiving country are followed by a decrease or increase in queries for the country.

Push-pull:

- i. If increases or decreases in search patterns adjusts (correlates) with number of asylum seekers in the receiving country this indicates queries adapts to the changing nature of ‘intervening obstacles’ like information deficit and border controls.
- ii. If increases and decrease in search patterns corresponds with number of asylum seekers in receiving countries this implies that the information gathered by queries affects the cost-benefit analyses for potential migrants.

7 Results & Empirical Overview

7.1 Report

A brief report that is not an attempt to give a full overview of what sites etc. a migrant use, rather it aims to illustrate for the reader what a migrant can access via the internet.

Table 10.

Border controls and travel path	How to get asylum	Organizational infrastructure	'Loose tie' connectors	'Close tie' connectors
Tickets and routes (Parkinson, 2016; Alarabiya, 2014)	NGOs (Open Borders Immigration, 2014)	How to find smugglers (Parkinson, 2016)	Internet Forums	Skype
Guides for how to take journey to Europe (Alarabiya, 2014)	Migration agencies (Migrationsverket, 2016)	Where NGOs are located and help is to be found in transit countries (Médecins Sans Frontières, 2016)	Facebook Groups	WhatsApp
			Instagram	Facebook

Notable information from the table is that Syrians in Turkey use Facebook groups for example (Reuters, 2015) and there are Facebook groups called Syrian Refugees in Europe with >10000 members (Facebook, 2016). That migrants use internet was confirmed via personal communication with two volunteers at refugee welcome centers in Sweden and from one person who works for medicines sans frontiers in Greece. It was confirmed that open Wi-Fi access existed in some refugee camps in Greece. All three had directly observed migrants using internet via computers or smartphones at different refugee centers (see Annex for reference).

7.2 Empirical overview: Eastern Mediterranean Route

7.2.1 Google Correlate: Turkey

r = Pearson correlation coefficient and the Correlate results in table 11 is from the time period 2013.01.02-2015.12.26 for Sweden and 2014.08.29-2016.01.24 for Germany¹.

Table 11.

r	Sweden: السويد	r	Germany: ألمانيا
0,90	Maps	0,92	Great Britain
0,87	Cyprus	0,91	Immigration
0,87	Denmark	0,91	Syrians
0,86	Maps	0,90	The sea
0,86	Asylum	0,88	Syrian refugees
0,85	World map	0,88	Migrant
0,84	Asylum in Sweden	0,87	Seek refugee

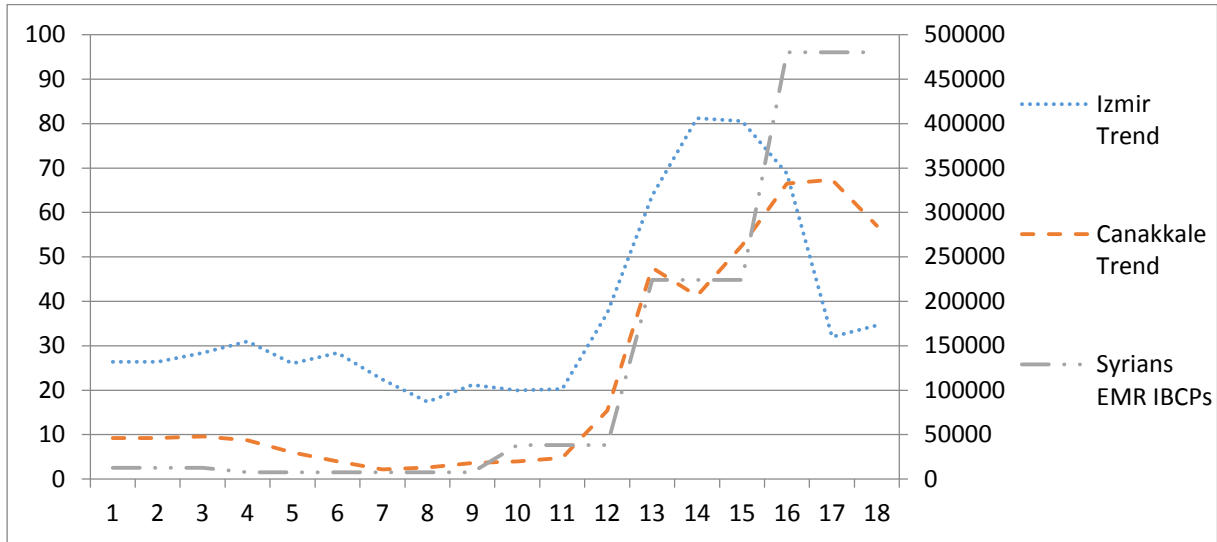
Table 11 reveals that migrants in Turkey have access to internet and use it for information gathering whilst in a transit country. It validates the operationalizing-terms since it shows the search-terms are used for migration related information gathering. The results show search-pattern behavior by Arabic speakers. E.g. the correlation with searches for Sweden in Arabic has a very high correlation with other search-terms in Google like ‘Asylum in Sweden’.

It’s important to highlight that this section is not meant to provide an empirical and generalizable view of internet technology by itself. The purpose of the empirical overview is to complement the results from the quantitative analysis and to validate the operationalizations.

¹ To rid for world cup correlations

7.2.2 Google Trends: Turkey

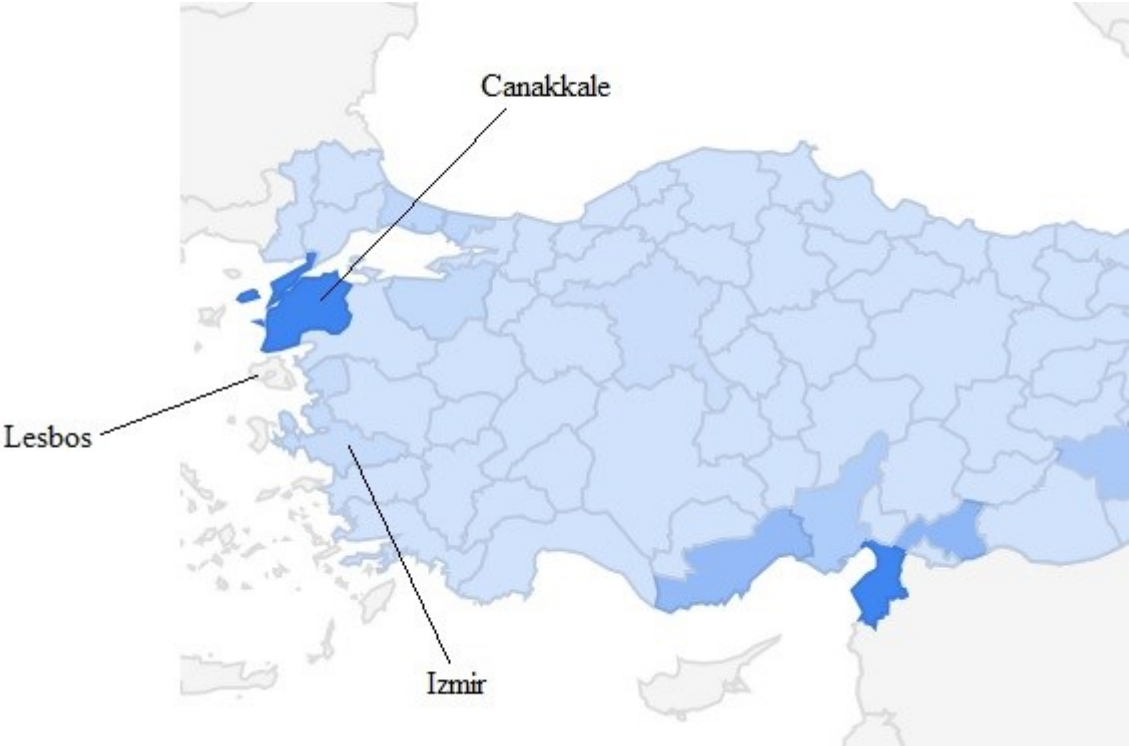
Graph 1. Turkey regions Izmir and Canakkale and number of Syrians detected by Frontex as illegal BCPs in Greek islands. Time period is 2014.08-2015.12².



The left axis corresponds to the Canakkale and Izmir Trends over time. The right axis is attached to Syrians detected by Frontex as illegal BCPs in Greek islands over time. The reason this timeline jumps each 3 months is because this data is per quarter. The chart shows a clear connection between the Google-Trends time-series and irregular migrants trends. That's more people journey across the EMR simultaneously with the increase of the relative popularity of search-terms connected to migration made by Arabic speakers in these two Turkish regions (see figure 1), which confirms the two variables are connected.

² To avoid noise from the World Cup of football.

Figure 1: Map of Turkey



Map generated by Trends

Geographic location of Izmir and Canakkale in relation to Greece, and the island of Lesbos which received more refugees on the EMR than any other island 2015 (Leadbeater, 2016).

7.3 Empirical overview: Syria

7.3.1 Google Trends: Syria

Table 12. Trends-Question

Country	Syria		
Question nr.	Sweden: السويد	Germany: ألمانيا	Immigration: الهجرة
1	Photos of Sweden	Map of Germany	Immigration to Canada
2	Immigration to Sweden	Flag of Germany	Immigration to Germany
3	Asylum in Sweden	-*	Immigration to Sweden
4	Resorting to Sweden	-	Immigration to Australia
5	Sweden's capital	-	Syrian Immigration and Passport**

*No more questions generated by Trends.

**This was the 7th and not the 5th most frequent asked question

This table confirms the findings from table 11 and further validates the operationalization of the search-terms with the Internet variable and this indicates an internet behavior by potential migrants in Syria that the search-terms was used to ask migration related questions.

7.4 Statistical Test Results

For all tables in this section 7.4.X the measurements in the tables are the value of the Beta coefficients. To be clear what this means is that each cell in all the tables are the Beta coefficient from a linear regression analysis where the independent variable, the index, i.e. relative popularity of that specific google search-term is used to predict the variation in the dependent variable, the total amount of asylum seekers for the googled receiving country. t+x means there is a prediction effect that's been tested for if someone in Syria googles 2015.06 and this correlates with number of asylum seekers in receiving county 2015.07. That means that it's assumed they google in sender country *a* and then they apply for asylum in country *b* one month later. t-x means the opposite, that there is a time lag effect that's been tested. In the case of Nigeria that means that if we test the asylum data in Italy from I.e. 2015.06 there is a strong correlation with google searches made in Nigeria for Italy 6 months later in 2015.12.

The reader should not that the main purpose of results from countries other than Syria is to triangulate the results from Syria.

7.4.1 Syria

Table 13.

Months*	Aus	Bel	Ger	Hun	NL	Nor	Swe
2015	0,75	0,67	0,03	0,92 ⁰	0,75	0,88 ⁰	0,40
2014	0,77	0,07	0,48	0,64	0,58	0,82	0,83
2013	0,63	0,32	0,18	0,42	0,90	0,69	0,84 ⁰
2012			0,28				0,96 ⁰
2011			0,64 ⁰				0,44
2015, t+1	0,96	0,94 ⁰⁰	0,40	0,47	0,96 ⁰⁰⁰	0,94 ⁰⁰⁰	0,80 ⁰
2014, t+1	0,96 ⁰⁰	0,26	0,12	0,53	0,86 ⁰	0,76	0,73
2013, t+1	0,84 ⁰	0,44	0,41	0,11	0,76	0,54	0,69 ⁰
2015, t+2	0,82	0,87 ⁰	0,60	0,07	0,81	0,59	0,94 ⁰⁰⁰
2014, t+2	0,83	0,35	0,13	0,01	0,88	0,48	0,31
2013, t+2	0,78	0,26	0,51	0,25	0,65	0,30	0,37

* 2015 includes the first two months of 2016

⁰ Significant at 90%

⁰⁰ Significant at 95%

⁰⁰⁰ Significant at 99%

Every cell in table 13³ above represents one linear regression analysis for that particular time period and country. E.g. for the top middle one, Hungary for the time period 2015.01-2016.02 regression analysis has been conducted with N=14, the independent variable is google searches for Hungary from Syria indexed by month. The dependent variable is the amount of Syrian asylum applications in Hungary per month.

Table 14.

Months*	Aus	Bel	Ger	Hun	NL	Nor	Swe	Ger**	Swe***
2013-2015*	0,90 ⁰⁰⁰	0,81 ⁰⁰⁰	0,52	0,93 ⁰⁰⁰	0,75 ⁰⁰	0,92 ⁰⁰⁰	0,21	0,97 ⁰⁰⁰	0,73
2013-2015, t+1	0,74	0,58	0,44	0,81 ⁰⁰⁰	0,55	0,69 ⁰⁰	0,49 ⁰⁰	0,82	0,78
2013-2015, t+2	0,61	0,34	0,37	0,38	0,35	0,34	0,24	0,72	0,48

* 2015 include the first two months of 2016 as well

⁰ Significant at 90%

⁰⁰ Significant at 95%

⁰⁰⁰ Significant at 99%

³ All cells in tables 13-18 have been calculated likewise.

Table 15.

Months*	N_WC_C	WC_C	N_WC_C ***	WC_C **
2013-2015*	0,91 ⁰⁰⁰	0,58	0,86 ⁰⁰⁰	0,94 ⁰⁰⁰
2013-2015, t+1	0,77 ⁰⁰	0,75 ⁰⁰	0,66 ⁰⁰	0,95 ⁰⁰⁰
2013-2015, t+2	0,55	0,85 ⁰⁰⁰	0,36	0,84 ⁰⁰⁰

* 2015 include the first two months of 2016 as well

⁰ Significant at 90%

⁰⁰ Significant at 95%

⁰⁰⁰ Significant at 99%

** This is measures where 5 outliers have been removed for Germany. Which are the five months Germany received the most Syrians.

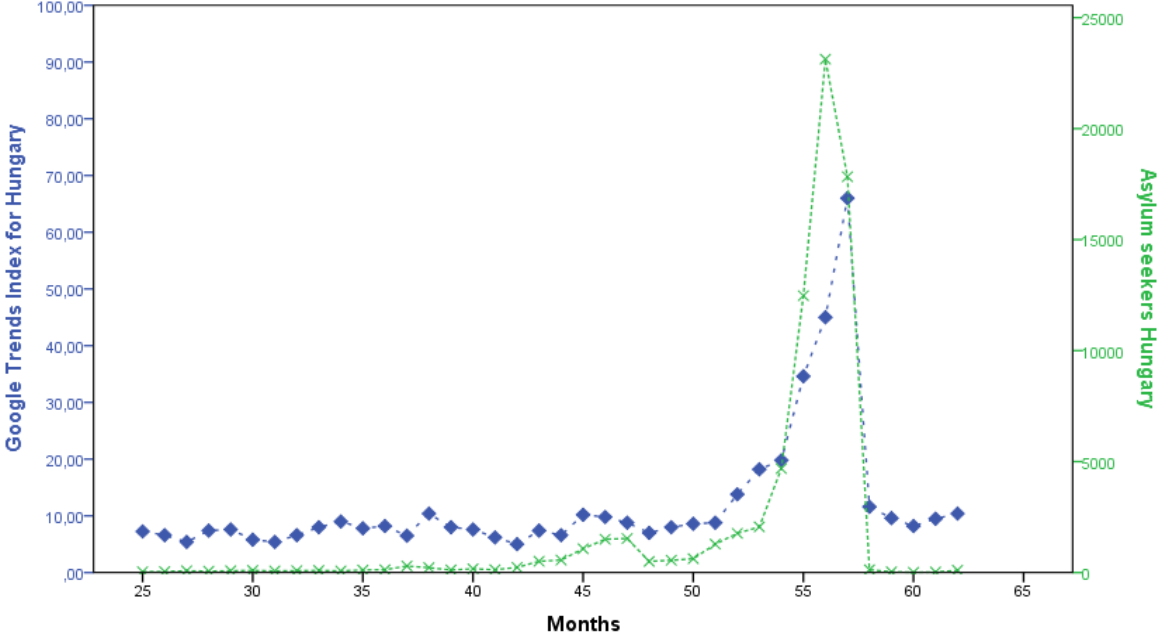
*** This is measures where 4 outliers have been removed for Sweden. Which are the four months Sweden received the most Syrians.

- N_WC_C: Countries that didn't participate in the world cup of football 2014: Swe, Nor, Hun and Aus.

- WC_C: Countries that did participate in the world cup of football 2014: Ger, Bel and NL.

The very high measure in the Beta coefficients shows that the independent variable strongly can predict the variation in the dependent variable. For example, we can some extremely high predictions made by the independent variable if we look at 2015 t+2 for Sweden. This show a Beta coefficient of 0,94 at 99% significance level which is very close to 1 which would indicate at least in principle perfect cause and effect. It's remarkable that for single years there are significant result even though the N is only 12. That many Beta measures are low and not significant for 2013 and 2014 isn't strange since the flow was lower than compared to 2015. This is actually a validation of the operationalization of search interest since it shows that it's more likely to correlate with asylum flows if the asylum flows are large enough.

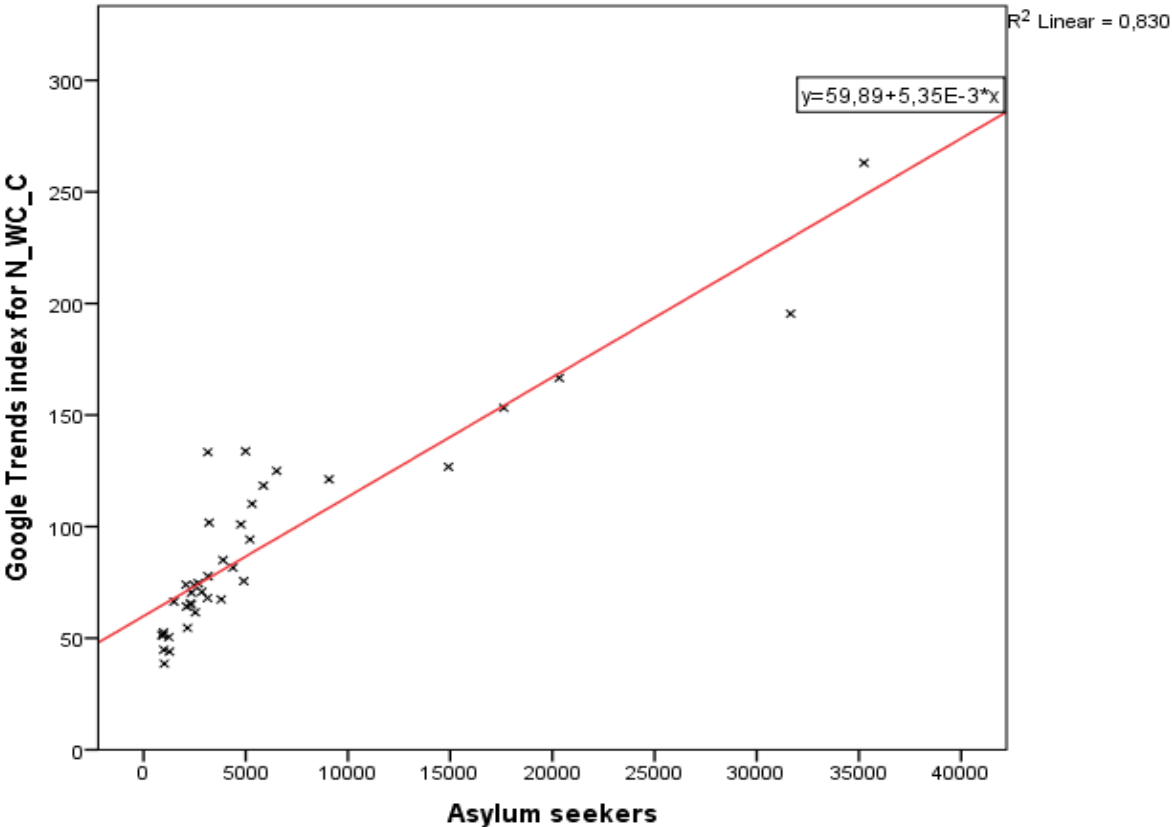
Graph 2. Hungary



Diamonds are Google-Trend index and Crosses are asylum applications.

This graph shows a dual axis timeline for 2013.01-2016.02, the left axis show how search interest for Hungary rose and fell in in Syria. The right axis shows the number of asylum applications made by Syrians in Hungary. The timelines follow each other clearly when the amount of asylum seekers increase.

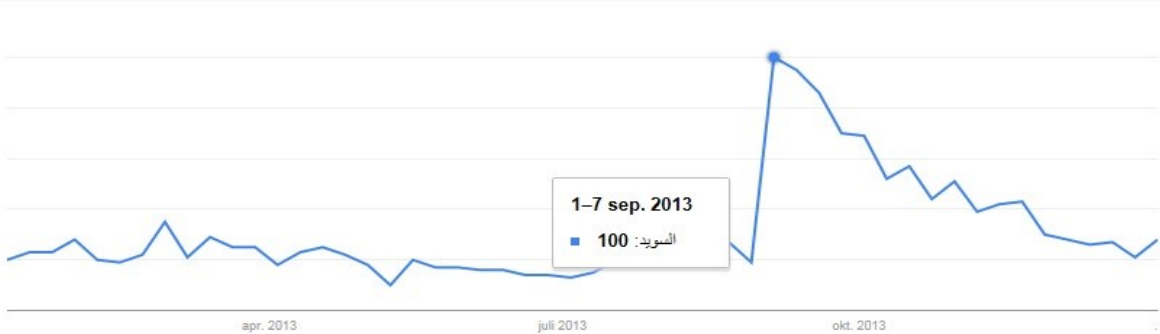
Graph 3. Scatter plot with N_WC_C countries with a fit line.



Note that R^2 is a different statistical measure than Beta and that this is the same equation as in corresponding to the numbers in table 15. Graph 3 illustrates with a scatterplot that the connection between the independent and dependent variable can be described as linear with high precision.

7.4.2 Syria-Sweden

Graph 4. Trends-timeline



This is a time series of 2013. The spike of increased search interest for Sweden in September occurs simultaneously with the announcement that Sweden becomes the first country in Europe to declare permanent residency rights for Syrian asylum seekers on the 3rd of September 2013 (DN, 2013). This shows that news migration related news from Sweden travels almost instantaneously and affects the internet behavior of people in Syria.

Table 16. Syria to Sweden. Regression analysis with weekly data.

Weeks*	Sweden
2014	0,71 ⁰⁰
2014, t+1w	0,64 ⁰⁰⁰
2014, t+2w	0,63 ⁰⁰⁰
2014, t+3w	0,61 ⁰⁰⁰
2014, t+4w	0,58 ⁰⁰⁰
2014, t+5w	0,48 ⁰⁰⁰
2014, t+6w	0,39 ⁰⁰⁰
2015	0,25
2015, t+1w	0,38
2015, t+2w	0,46 ⁰
2015, t+3w	0,55 ⁰⁰
2015, t+4w	0,63 ⁰⁰⁰
2015, t+5w	0,72 ⁰⁰⁰
2015, t+6w	0,81 ⁰⁰⁰
2015, t+7w	0,82 ⁰⁰⁰
2015, t+8w	0,83 ⁰⁰⁰
2015, t+9w	0,80 ⁰⁰⁰
2015, t+10w	0,77 ⁰⁰⁰

* 2015 include the first 11 weeks of 2016 as well
⁰ Significant at 90%
⁰⁰ Significant at 95%
⁰⁰⁰ Significant at 99%

It’s astonishing how detailed and significant the results are at a weekly level for Syrian asylum seekers in Sweden and Google-searches for Sweden made in Syria. This shows a high precision in predicting variations in asylum flows.

7.4.3 Afghanistan

Table 17. Afghanistan-Europe

Months*	Afghanistan
2015	0,74
2014	0,16
2013	0,11
2015, t+1	0,90 ⁰⁰
2014, t+1	0,17
2013, t+1	0,26
2015, t+2	0,17
2014, t+2	0,12
2013, t+2	0,76
2013 - 2015	0,83 ⁰⁰⁰
2013 - 2015, t+1	0,73 ⁰⁰⁰
2013 - 2015, t+2	0,64 ⁰⁰

Even though Europe was the only testable term it's remarkable that the results are significant and that significant Beta coefficients are very high.

* 2015 includes the first two months of 2016 as well

⁰ Significant at 90%

⁰⁰ Significant at 95%

⁰⁰⁰ Significant at 99%

7.4.4 Nigeria and Senegal

Table 18.

Months*	Nigeria - Europe	Senegal - Europe	Nigeria - Italy
2015	0,39	0,72 ⁰⁰⁰	0,26
2014	0,12	0,52	0,27
2013	0,02	0,42	0,42
2013 - 2015	0,30 ⁰	0,35 ⁰⁰	0,29
2014 - 2015	0,37	0,45 ⁰⁰	0,08
2014 - 2015, t-1	0,22	0,35 ⁰	0,33
2014 - 2015, t-2	0,19	0,25	0,57
2014 - 2015, t-3	0,22	0,17	0,76 ⁰
2014 - 2015, t-4	0,25	0,30	0,86 ⁰⁰⁰
2014 - 2015, t-5	0,24	0,51 ⁰⁰⁰	0,85 ⁰⁰⁰
2014 - 2015, t-6	0,25	0,42 ⁰⁰	0,82 ⁰⁰
2014 - 2015, t+1	0,36	0,50 ⁰⁰	0,06
2014 - 2015, t+2	0,31	0,28	0,13
2014 - 2015, t+3	0,06	0,01	0,06

* 2015 includes the first two months of 2016.

⁰ Significant at 90%

⁰⁰ Significant at 95%

⁰⁰⁰ Significant at 99%

A time lag different from the previous results seems to appear here. I.e. 4-6 months after there is an increase of Nigerian and Senegalese asylum applications correlates with an increase in queries in the sender country.

8 Analysis

First the main findings made in this paper will be discussed and secondly its theoretical implications. The purpose of this paper was to prove the existence of a relationship between irregular migration and internet behavior. This has been done to a large extent. The results in this paper gives evidence for a strong relationship between the two and thus I have developed a new methodological and theoretical understanding about irregular migration. These findings are unprecedented in two ways. First, the high levels of correlated variance shown by the close to 1 Beta coefficients and the findings of the empirical overview give quantitative depth and detail with how migrants internet behavior is connected to when and where they go. Secondly, just the fact that the result shows that there is a relationship between the variable asylum seekers and the variable internet behavior is the first of its kind and is the main finding in this paper.

The increased understanding for driving forces behind irregular migration can be used in concern on how to regulate and manage migration. Since this method effectively can predict if large flows will increase or decrease depending on the search behavior of potential migrants in sender states. This gives policy makers and politicians a new way to grapple and understand large cases of irregular migration, like the migration crisis of 2015. This method can be beneficial for people working with business intelligence and global/environmental scanning and analysis (like migration agencies). These completely new methodological toolkits unveil a new way to look at the phenomena of large migration flows. And if nothing else these are remarkable findings in themselves despite any other analysis that is done in this paper. Hopefully this could reduce the gap between policy outcome and output and lead to better management preventing human loss and suffering.

All of this is of course very intriguing and generates more questions than answers about what this means and how this can be used to explain migration, something that shows how these results take a leap forward for theoretical development in this field and creates the potential for future groundbreaking research. Hence I have exceeded the goals to develop theory and instead I believe a new door has opened in how to understand migration with the help of big data.

However, the burden of proof is larger than this due to how the research question was asked. Some comments are needed before the analysis proceed. Because the findings open a door to uncharted territory there are few footsteps for me to follow and I am well aware of the uncertainty and ambiguity in how the results can be interpreted with existing theory but I ask the reader to be patient and to understand that there is just so much one single researcher can do with limited time and resources. There are several interesting takes on how these results can develop existing theory. But I would like to first caveat the analysis that follows below with stating that: Since this relationship was first discovered in this paper by using a new way of operationalizing a new factor the need for complementing research that

can confirm these findings is much needed! Hence the attempt to develop theory based on these unprecedented results should be seen in light of this and much caution on how to interpret them should be taken.

8.1 Analysis: Development of theory

I will in this section develop and discuss the reasons for *how internet behavior can explain irregular migration*. The overarching theme here is quite simple, if the IV can show to have a mediating and/or moderating effect on already proven theoretical factors it can be relatively well reasoned that the usage of internet by migrants have a causational effect on the outcome of irregular migration over time. I urge the reader to keep in mind that if there are good reasons for why these findings are related to existing theory that is all that is necessary in order to explain and prove, at least, partially the causational relationship.

The remarkably high Beta coefficients and the results from the empirical overview in general shows that information flows via internet adjust quickly to changes in circumstances and that information instantaneously is reached by potential migrants in Syria. The results indicate that this methodology is very precise in predicting and gives a precise macro-overview of the direction and volume flows of asylum seekers and that Google Trends is useful not only in econometrics (Choi and Varian, 2009). The results demonstrate that the search patterns are quick to adjust to ‘shocks’ in the system (King, 2012) like border control news of Sweden and Hungary, see graph 2 and 4 for illustration. This increases the plausibility that this method can be incorporated in social network theory. In part I disprove the notion that is expressed by Frontex (and many academics) that irregular migration is a complex phenomenon that changes quickly and unexpectedly (Frontex 2016). Because the results show that even though unexpected shocks happen and influence migration the search patterns follow quickly. This indicates that information in the migration network travels via the internet.

There is evidence to prove all of the hypotheses; moreover the results seem to confirm the logic of self-feeding chain migration. The reason for this is because that from Syria to target country X we can see in 2013 and 2014 how most of the significant results correlate asylum flows with search interest in the origin country simultaneously and not with a time delay. This indicates that movement of the ‘first-movers’ generates increased interest for the target country in the sender country. Then in 2015 when large numbers of Syrians begin to move it shows that they know where they want to go and seek out information about their target country 1-2 months before they arrive. It can also mean that they follow in the footsteps of the diaspora and google before they leave Syria in order gather information. A third interpretation is that the results indicates that this have a ‘motivational effect’ on potential Syrian migrants in 2015 because they then actively seek out information before they leave about receiving states and this decreased information deficit and increase social capital could raise motivation to depart.

In Syria its observable how variation in interest for a destination country over time predicts with high certainty how large numbers of Syrians migrate. The results portray how Google-queries can be viewed as a predictor for if; irregular migration increase or decrease. It also shows that it act as a determinant because by analyzing the search patterns its possible to explain why Syrians go to country A instead of country B Something which develop a gap in previous theory (Venturini, 2004. Thus I claim to have developed further understanding in both migration-network theory and push-pull theory. That's, if a country is being more searched for than before it can reasonably well be known that increased migration flows follow.

What the results prove is that the Syrian flows follow the Diasporas to some extent I.e. the network ties in general. It also seems like the effects of the information spread within the network is intensified by the linking made possible with internet access. Because if one considers earlier findings that to a large extent are pre-internet, migration network theory show how networks reduces the cost and risks of migration via close and loose social ties. There are very strong reasons to believe that these social ties, both loose and strong (Massey et al, 1998; Crisp 1999), are expanded by sheer access of seeing and interacting with people in huge social media groups, internet forums etc. Even though my findings directly to this phenomenon are miniscule, indirectly there are good reasons to claim this. Because my findings directly show that migrants use internet to a large extent when deciding when and where to go. This implies that the internet is used to a large extent by these migrants thus it would be reasonable that this affects the systemic network factors that cause migrants to move in the first place. However, I cannot beyond doubt fully prove counter factuality, but the results strongly imply that if internet access is removed it would have an effect on the variation and direction of the flow of migrants.

Isolation of the variable isn't either proved beyond doubt but nevertheless the fact that results from Somalia was random and that data for more precise search-terms couldn't be generated for Afghanistan, Nigeria and Senegal strengthens the reasons for that this variable is isolated. Furthermore, I don't claim it's *completely* isolated only that is has an effect on important factors *proven by others*.

The findings show how information regarding receiving countries is picked up swiftly and we can with these results with a high degree explain that Syrians in Syria know when to *not* continue the chain-migration along the lines of their network ties. The reason for this is simple. Information is quickly accessed by potential migrants in Syria and elsewhere. It also seems to be the case that receiving countries that had larger diasporas of Syrians 2011 has received more asylum seekers and my results doesn't contradict why it would be so. But this doesn't explain why all of sudden Syrians started to seek asylum in countries such as Hungary, something that my methodology can do. But when they do go to Hungary or other small diaspora countries there seems like the chain migration phenomena immediately kick in. I.e. that once initial first movers have reached a destination the costs and risks for other within their network to go there is reduced (Massey et al, 1998), migration becomes self-feeding (De Haas, 2010; Faist 1997). And given that interest about the destination increases almost directly in Syria about Hungary, Sweden etc. this shows that information about them had reached migrants via media or their social networks. I argue that this gives some evidence for that the chain-migration self-feeding effect of how migrants behave are intensified by the

internet. Similar effects can be seen in all countries receiving Syrians and Afghanis. Given this and given that networks reduces uncertainty by lowering the hurdle for potential migrants (Massey et al, 1998) it shouldn't come as a surprise that this causes more migration to happen and that this self-feeding effect is probably increased due to internet access. This gives insight in how future migration flows are likely to be more rapid and possibly harder to predict, at least with methods of analysis that doesn't include this new factor.

The results show that this methodology is successful in acting as a predictor of irregular flows. Especially when the effect of the networks facilitating information spread has been in effect for a while and the larger the flows are. The weekly data measured for Syria-Sweden can be viewed as an accurate measurement of how long it takes from the point migrants in Syria begin their information gathering behavior about destination country to the point when they reach the destination country. Even though the reasons why they leave are probably due to external push forces such as violence it gives a detailed description of how well queries can predict destination country on a week level. Something that develops the theoretical understanding of why a migrant choose a particular country over others and that push-pull theory can together with migration network theory explain irregular migration with a new type of empirical evidence (Venturini, 2004; Hooghes et al, 2008).

The reason why the results are so accurate as they are can be understood by applying the logic of a threshold perspective on access to internet happens. Because of language and cultural similarities and of social ties in a migrant's social network it's likely he or she will interact with migrants in the same social network and then e.g. even if a young Syrian man don't own a smartphone he is likely to either travel with, or make contact with Syrians who has access or had internet access in the origin country and thus the benefits internet access provides extends to him as well. So despite the obvious fact that all migrants don't have direct access to internet via smartphones etc. this shows how they could gain advantageous from it anyhow.

Summary of the analysis:

1. Internet access increases the size of the migrant's network.
2. Internet access makes existing social ties stronger.
3. Internet access boosts pull-factors such as information about opportunity or perceived opportunity and lowers the risk and costs of migration.
4. Internet access mitigates the effect of 'intervening obstacles'.
5. Internet access provides information about travel routes and connecting migrants with 'organizational infrastructure'.
6. Search patterns are able to precisely predict the increases and decreases of migration flows and insights in how migrants chose target country.

8.1.1 Nigeria & Senegal

The results from Nigeria and Senegal should be commented because the results are different for mainly Nigeria than for the other countries. What's clear in the case of Nigeria is that there is a strong relationship between the independent and dependent variable. Hence much of the reasoning in the previous section applies to Nigeria too. But instead of a strong prediction correlation there's instead a time lag of about 6 months. I would like to stress again the need for further research about the plethora of ways my big data approach can be used and therefore I don't wish to spend too much time on this time-lag effect. However, what it could indicate if push-pull theory and network theory is used to explain this is that it takes a longer amount of time from the departure from Nigeria to reach their destination and/or once in Italy/Europe it takes longer time for the people in the migrant's network to react and spark interest in Nigeria. The 'order in time' dilemma that could arise from this is thus not incompatible with existing theory. However, the relationship might be spurious given the low N and more research is needed in order to better understand these findings.

8.1.2 Generalizability & Validity

The results and analysis I claim are generalizable unto the larger abstract defined population of migrants that live in countries with above 'critical levels' of internet usage. Because the definition of the target population studied are freed from the constraints of time and space (Teorell & Svensson, 2007). The research designs solid foundation provides good ground for this statement (Teorell and Svensson, 2007). However, what still concerns the soundness and validity of the logic is the very broad operationalizing made for the IV and I do admit that there are objections that can be made against it. But the operationalizing seems to be valid at least for a macro perspective and provides a quantitative measurement of interest to some extent that information gathering by migrants with search engines takes place. This seems almost irrefutable in the very narrow sense of it. Though, in a wider perspective there are ambiguity concerns of specifically to what extent this can be applied and be seen in relation to how information travels via social networks and ergo how this affect variation in the dependent variable. The specific causal chain of events on *how* this happen I make no claim to have fully uncovered, only that there are strong indications that the variables are connected.

9 Conclusion

In this paper a large-scale overview of how irregular migration can be described and understood with big data. The analysis shows that these findings are applicable over time for four different countries in varying degrees. And when the ‘internet-variable’ is used *within* the logical framework of push-pull theory and migration network theory it can explain how migration is affected by it. The findings remarkably high precision was attained with quite simple tools but the findings leaves much room for further research to develop and confirm these results. It’s quite clear that internet behavior should be included as a factor for understanding modern migration. Even though concerns surrounding validity can be made the results provides remarkably significant measures and makes it possible to reconstruct how irregular migrants behave with the use of big data search patterns.

The fact that internet usage and smartphone penetration worldwide is steadily growing shows that the implications from this paper reveals both current but increasingly so the future importance it can have for the field of politics and academia. On an abstract notion this research can be seen as providing evidence for that the ‘Information age’ have fundamental political impacts on society (Castell, 1998).

10 References

- Alarabiya.net. (2014). [online] Available at: <http://www.alarabiya.net/ar/arab-and-world/syria/2014/07/09/%D8%A8%D8%A7%D9%84%D8%B5%D9%88%D8%B1-%D8%B1%D8%AD%D9%84%D8%A9-%D8%B3%D9%88%D8%B1%D9%8A-%D9%85%D8%B9-%D8%A7%D9%84%D9%85%D9%87%D8%B1%D8%A8%D9%8A%D9%86-%D9%81%D9%8A-%D8%A7%D9%84%D8%A8%D8%AD%D8%B1.html> [Accessed 19 May 2016].
- Askitas, N. (2015). Google search activity data and breaking trends. 1st ed. [ebook] IZA World of Labor. Available at: <http://wol.iza.org/articles/google-search-activity-data-and-breaking-trends.pdf> [Accessed 17 May 2016].
- BBC News. (2016a). Why is EU struggling with migrants and asylum? - BBC News. [online] Available at: <http://www.bbc.com/news/world-europe-24583286> [Accessed 16 May 2016].
- Bhagwati, J. (2004). In defense of globalization. New York: Oxford University Press.
- Bimal Ghosh, (1998). Huddled Masses and Uncertain Shores: Insights Into Irregular Migration, IOM and Martinus Nijhoff, The Hague, , p. 67.
- Blaschke, Jochen, (1998). ‘Addressing the employment of migrants in an irregular situation: the case of Germany’, paper presented to the Technical Symposium on International Migration and Development, The Hague, 29 June - 3 July, p. 19.
- Bogdan, R. C. & Biklen, S. K. (2006). Qualitative research in education: An introduction to theory and methods. Allyn & Bacon
- Brettell, C. B., & Hollifield, J. F. (eds.) (2008). Migration Theory. Talking Across Disciplines. New York: Routledge. Foggy Social Structures: Irregular Migration , European Labour Markets and the Welfare State , edited by Michael Bommers and Giuseppe Sciortino. Amsterdam, NL: Amsterdam University Press, 2012. 228pp.
- Brunwasser, M. (2015). A 21st-Century Migrants Essentials: Food, Shelter, Smartphone. [online] Nytimes.com. Available at: <http://www.nytimes.com/2015/08/26/world/europe/a-21st-century-migrants-checklist-water-shelter-smartphone.html> [Accessed 1 Apr. 2016].
- Hollis, Martin, (2002). The Philosophy of the Social Science. Cambridge University Press.
- Castles, S., and M. Miller (1994). The Age of Migration. 1st edn. Basingstoke, UK: Macmillan.
- Data.worldbank.org. (2016). Internet users (per 100 people) | Data | Table. [online] Available at: <http://data.worldbank.org/indicator/IT.NET.USER.P2> [Accessed 19 May 2016].

- De Haas, H. (2010). Migration and Development: A Theoretical Perspective 1. *International Migration Review*, 44(1), pp.227-264.
- Douglas Massey, quoted in Nicholas Van Hear, (1998). *New Diasporas: the Mass Exodus, Dispersal and Regrouping of Migrant Communities*, UCL Press, London, p. 259.
- Facebook.com. (2016). Syrian refugees. [online] Available at: <https://www.facebook.com/syriangreats/> [Accessed 19 May 2016].
- Faist, T. (1997). The Crucial Meso-Level. In T. Hammar, G. Brochmann, K. Tamas & T. Faist (eds.), *International Migration, Immobility and Development. Multidisciplinary Perspectives*. Oxford: Berg, 187-218.
- Favell, A. (2008). Re-booting Migration Theory: Interdisciplinarity, Globality and Postdisciplinarity in Migration Studies, in C. B. Brettell, & J. F. Hollifield (eds.), *Migration Theory. Talking Across Disciplines*. New York: Routledge, 259-278.
- Gambino, L. (2015). 'No human being is illegal': linguists argue against mislabeling of immigrants. [online] *the Guardian*. Available at: <http://www.theguardian.com/us-news/2015/dec/06/illegal-immigrant-label-offensive-wrong-activists-say> [Accessed 16 May 2016].
- Hatton, T. (2004). Seeking asylum in Europe. *Economic Policy*, 19(38), pp.6-62. Hollifield, J., Martin, P. and Orrenius, P. (2014). *Controlling immigration*. 3rd ed. Palo Alto: Stanford University Press.
- Hooghe et al, (2004). Migration to European Countries: A Structural Explanation of Patterns, 1980-2004 Author(s): Marc Hooghe, Ann Trappers, Bart Meuleman and Tim Reeskens Source: *The International Migration Review*, Vol. 42, No. 2 (Summer, 2008), pp. 476-504 Published by: Center for Migration Studies of New York, Inc. Stable URL: <http://www.jstor.org/stable/27645259>
- Itu.int. (2016). ITU releases 2015 ICT figures. [online] Available at: http://www.itu.int/net/pressoffice/press_releases/2015/17.aspx#.VzpAXvmLTIU [Accessed 16 May 2016].
- Jandl, Michael. (2007). "Irregular Migration, Human Smuggling, and the Eastern Enlargement of the European Union". *The International Migration Review* 41 (2). [Center for Migration Studies of New York, Inc., Wiley]: 291–315. <http://www.jstor.org.ludwig.lub.lu.se/stable/27645667>.
- King, R. (2012). *Theories and Typologies of Migration: an overview and a primer*. 1st ed. Malmö: Malmö Institute for Studies of Migration, Diversity and Welfare (MIM).
- Leadbeater, C. (2016). Which Greek islands are affected by the refugee crisis?. [online] *The Telegraph*. Available at: <http://www.telegraph.co.uk/travel/destinations/europe/greece/articles/greek-islands-affected-by-refugee-crisis/> [Accessed 19 May 2016].
- Massey, D. et al. (1998). *Worlds in Motion. Understanding International Migration at the End of the Millennium*. Oxford: Clarendon Press.
- Marcellin, F. (2016). Europe's refugee crisis: How one Wi-Fi network is offering a lifeline for Lesbos migrants | ZDNet. [online] ZDNet. Available at:

- <http://www.zdnet.com/article/europes-refugee-crisis-how-one-wi-fi-network-is-offering-a-lifeline-for-lesbos-migrants/> [Accessed 16 May 2016].
- Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, J. (1993). *Theories of International Migration: A Review and Appraisal*. *Population and Development Review*, 19(3), p.431.
- Manuel Castells, (1998). "A New Globe in the Making: Manuel Castells on the Information Age". Review of *The Information Age: Economy, Society and Culture*. *Acta Sociologica* 41 (3). Sage Publications, Ltd.: 269–76.
<http://www.jstor.org/stable/4201086>. – ÄNDRA TILL URSPRUNGS KÄLLNA
- Médecins Sans Frontières (MSF) International. (2016). Greece. [online] Available at: <http://www.msf.org/en/greece> [Accessed 19 May 2016].
- Migrationsverket.se. (2016). - Migrationsverket. [online] Available at: <http://www.migrationsverket.se/Other-languages/-alrbyt-.html> [Accessed 19 May 2016].
- Munshi, K., (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics* 118 (2), 549–599
- Nairaland.com. (2013). Nigeria Vs Italy - International Friendly: (2 - 2) On 18th November 2013 - Sports - Nigeria. [online] Available at: <http://www.nairaland.com/1508675/nigeria-vs-italy-international-friendly> [Accessed 19 May 2016].
- Nordland, R. (2015). A Mass Migration Crisis, and It May Yet Get Worse. [online] *Nytimes.com*. Available at: <http://www.nytimes.com/2015/11/01/world/europe/a-mass-migration-crisis-and-it-may-yet-get-worse.html?hp&action=click&pgtype=Homepage&module=second-column-region®ion=top-news&WT.nav=top-news> [Accessed 16 May 2016].
- Open Borders Immigration. (2014). How to Get Refugee Asylum in Sweden. [online] Available at: <https://openbordersimmigration.com/how-to-get-refugee-asylum-in-sweden/> [Accessed 19 May 2016].
- Parkinson, D. (2016). Agadez Traffickers Profit From Movement Through Niger to Libya. [online] *WSJ*. Available at: <http://www.wsj.com/articles/agadez-traffickers-profit-from-movement-through-niger-to-libya-1437002559> [Accessed 19 May 2016].
- Portes, A., and R. Rumbaut, (1996). *Immigrant America: A Portrait*. 2nd edn. Berkeley, CA: University of California Press.
- Physioc, H. (2015). The Complex Web of Personalized Search. [online] *Tentacle Inbound*. Available at: <http://www.tentacleinbound.com/articles/personalized-search> [Accessed 20 May 2016].
- ReliefWeb. (2013). Norway to take in 1000 Syrian refugees. [online] Available at: <http://reliefweb.int/report/norway/norway-take-1000-syrian-refugees> [Accessed 20 May 2016].
- Reuters. (2015). For many refugees, journey to Europe begins on Facebook. [online] Available at: <http://www.reuters.com/article/us-europe-migrants-facebook-idUSKCN0R14CX20150901> [Accessed 20 May 2016].

- Ssb.no. (2001). Tabell 5 Innvandrarbefolkninga og personar med annan innvandringsbakgrunn, etter innvandringskategori, kjønn og landbakgrunn. 1. januar 2001. [online] Available at: <http://www.ssb.no/a/kortnavn/innvbef/arkiv/tab-2001-10-29-05.html> [Accessed 19 May 2016].
- Taylor et al. (2014). Mapping Diasporas in the European Union and United States. (2014). 1st ed. [ebook] RAND. Available at: http://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR671/RAND_RR671.pdf [Accessed 20 May 2016].
- Teorell, J. and Svensson, T. (2007). Att fråga och att svara. Stockholm: Liber.
- Tomanek, A. (2011). Understanding Migration: International Migration Theories. [online] Understandingmigration.blogspot.se. Available at: <http://understandingmigration.blogspot.se/2011/03/international-migration-theories.html> [Accessed 16 May 2016].
- Triandafyllidou, A. Maroukis, T. (2012). Migrant smuggling. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- UEFA.com. (2016a). FIFA World Cup - Teams – UEFA.com. [online] Available at: <http://www.uefa.com/worldcup/season=2014/teams/index.html> [Accessed 19 May 2016].
- UEFA.com. (2016b). UEFA EURO 2012 - History – UEFA.com. [online] Available at: <http://www.uefa.com/uefaeuro/season=2012/index.html> [Accessed 19 May 2016].
- UNHCR. (2016a). UNHCR Syria Regional Refugee Response. [online] UNHCR Syria Regional Refugee Response. Available at: <http://data.unhcr.org/syrianrefugees/regional.php> [Accessed 19 May 2016].
- UNHCR. (2016b). World Refugee Day: Global forced displacement tops 50 million for first time in post-World War II era. [online] Available at: <http://www.unhcr.org/53a155bc6.html> [Accessed 16 May 2016].
- Varian, H. and Stephens-Davidowitz, S. (2015). A Hands-on Guide to Google Data. 1st ed. [ebook] Google, Inc. Available at: <http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf> [Accessed 17 May 2016].
- Venturini, A. (2004). Postwar Migration in Southern Europe, 1950?2000: An Economic Analysis. Cambridge: Cambridge University Press.
- Wegge, Simone A. (1998). "Chain Migration and Information Networks: Evidence from Nineteenth-century Hesse-cassel". The Journal of Economic History 58 (4). [Economic History Association, Cambridge University Press]: 957–86. <http://www.jstor.org/stable/2566846>.
- Zlotnik, H. 1998 "International Migration 1965-96: An Overview." Population and Development Review 24(3):429-468.

11 Annex

For contact info from personal communication please contact me on
Oskar.Lif@hotmail.se