Journal of Geographical Systems (2019) 21:161–187
https://doi.org/10.1007/s10109-018-0278-1

ORIGINAL ARTICLE

Exodus 2.0: crowdsourcing geographical and social trails of mass migration

Troy Curry[1] · **Arie Croitoru**[1,2] · **Andrew Crooks**[1,2,3] · **Anthony Stefanidis**[1,2,4]

Received: 4 October 2017 / Accepted: 29 August 2018 / Published online: 21 September 2018

Abstract

The exodus of displaced populations is a recurring historical phenomenon, and the ongoing Syrian humanitarian crisis is its latest incarnation. During such mass migration events, information is an essential commodity. Of particular importance is geographical (e.g., pathways and refugee camps) and social (e.g., refugee activities and networking) information. Traditionally, such information had been produced and disseminated by authorities, but a new paradigm is emerging: Web 2.0 and mobile computing technologies enable the involved stakeholder communities to produce, access, and consume migration-related information. The purpose of this article is to put forward a new typology for understanding the factors around migration and to examine the potential of crowd-generated data—especially open data and volunteered geographic information—to study such events. Using the recent wave of migration to Europe from the Middle East and northern Africa as a case study, we examine how migration-related information can be dynamically mined and analyzed to study the migrants' pathways from their home countries to their destination sites, as well as the conditions and activities that evolve during the migration process. These new data sources can provide a deeper and more fine-grained understanding of the migration process, often in real-time, and often through the eyes of the communities affected by it. Nevertheless, this also raises significant methodological and technical challenges for their future use associated with potential biases, data quality issues, and data processing.

Keywords Refugees · Forced migration · Humanitarian crisis · Volunteered geographic information · Crowdsourcing · Social media · GIS · Web 2.0

JEL Classification C800 (Data Collection and Data Estimation Methodology; Computer Programs: General)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10109-018-0278-1) contains supplementary material, which is available to authorized users.

✉ Arie Croitoru
acroitor@gmu.edu

Extended author information available on the last page of the article

1 Introduction

From ancient times, human history has been sporadically marked by mass migration events, often triggered by poverty and economic hardship, armed conflict, or political and social turmoil. The combined effect of globalization and geopolitical instability of the first two decades of the current millennium has triggered notable migration trends across a fluid and turbulent global socioeconomic landscape. The number of forced migrants has reached 65.6 million, of which there are currently 22.5 million registered refugees (UNHCR 2017a).

During such mass migration events, information is an essential commodity. Whereas in the past such information was sparse, and its dissemination and manipulation were the sole purview of authoritative agents (primarily governments and international organizations), we are now entering a new paradigm of mass migration in the digital information age. In particular, the emergence of the Web 2.0 concept in the early 2000s (O'Reilly 2005) has fostered digital participation, with individuals creating, curating, and disseminating rich digital content for a multitude of issues, bypassing the traditional top-down information dissemination channels (Stefanidis et al. 2013a).

As a result, while forced migration is a recurring historical phenomenon, one can argue that today's migration events are unique in the sense that they evolve in an information-rich and often participatory environment. Whether it is by the refugees themselves who use various apps in their cellphones to communicate with each other and record their travails (Dekker and Engbersen 2014; Maitland and Xu 2015), or by the communities that interact with them (Andrade and Doolin 2016), we now have access to large amounts of web-accessible information that offer snapshots of the migration process. Deciphering these digital trails can offer unique insight into this complex process. In that respect, the study of migration can now transition from being information sparse to information rich.

Here, we adopt the term Exodus 2.0 to refer to this new migration paradigm in the digital age, whereby information is a commodity in the migration process. Examples of this new paradigm can already be observed of refugees not only carrying their possessions while seeking refuge but posting content, connecting with family and friends, and using crowdsourced geographic information in mobile applications for navigating (Charmarkeh 2013; Ram 2015; Sebti 2016). Migrants are also increasingly using social networks for preserving social ties and addressing weak ties that are identified during the planning process (Komito 2011). Such social networks are also being utilized for the establishment of new infrastructure of previously latent ties and can be a source for insider knowledge (Charmarkeh 2013; Dekker and Engbersen 2014). For example, a recent study of Eritrean asylum seekers in Europe has found social media to be a dominant source of information and networking, especially while en route (Brekke and Brochmann 2015). In the same vein, Alencar (2017) found that social media networking sites are of particular importance for refugees' assimilation and the creation of social capital.

Given the nature of migration processes, it is possible to explore them across two key dimensions: geographical and situational. The *geographical dimension* is

associated with the physical migration pathways migrants take from a country of origin to a destination site [often through a number of intermediate "stop" sites (Hübl et al. 2017)]. The *situational dimension* is associated with the social connectivity of moving migrant populations, the conditions on the ground, and the activities that take place as part of migration efforts. Accordingly, in this article we study the potential of crowd-generated data in the form of volunteered geographic information (VGI) and social media content to provide information about migration processes and subsequently to offer a new lens to study such phenomena.

Often, the geographical and the social dimensions intertwine, as much of the social media that is created by users encompasses a geographical component, either through the direct sharing of geographical information or by providing such information implicitly. The former form of contributions has been termed volunteered geographic information (VGI, Goodchild 2007), while the latter has been termed ambient geographic information (AGI, Stefanidis et al. 2013b). Together, these forms of geographical information have already been proven valuable with respect to collecting, analyzing, and modeling human activity, especially in the context of disasters. For example, social media and citizen participation have been found to effectively disseminate early warning information to populations after natural disaster events (Chatfield et al. 2013), mapping the extent of damage after an earthquake (Zook et al. 2010), providing information on crisis-relevant information (Herfort et al. 2014), and understanding digital "data shadows" that link the virtual to the physical environment (Shelton et al. 2014). Similarly, the influence of social media networks has been shown to affect evacuation decisions during disaster events (Widener et al. 2013). VGI platforms such as Ushahidi (2017) have created a technology system that enables crowdsourcing by activists to collect, monitor, and distribute local observations via mobile phone for applications such as election monitoring, crisis response, advocacy, and human rights. Despite these early successes, the exploration of the potential utility of such new data sources for exploring migration—and in particular forced migration—has remained largely untapped. This is the focus of this paper. Using the Syrian migration crisis as a case study, we explore how the combination of more traditional top-down migration-related data sources can be augmented with bottom-up crowdsourced data to study forced migration (Sect. 2). In order to study how these two data sources can be integrated into a single framework, we then propose an information typology that specifies the relations between the different sources and data types (Sect. 3) and showcase its implementation in the context of our case study. Following this, we explore how the combination of traditional and emerging migration-related data sources can be used to inform migration modeling (Sect. 4). Finally, we discuss key challenges, opportunities, and future research directions that emerge from our typology.

2 Background

In recent years, forced migration has experienced a surge in activity due to the civil conflicts and political instability in the Middle East and Africa. Through such activities, refugees forge migration paths as they move through land, water, and air (or a

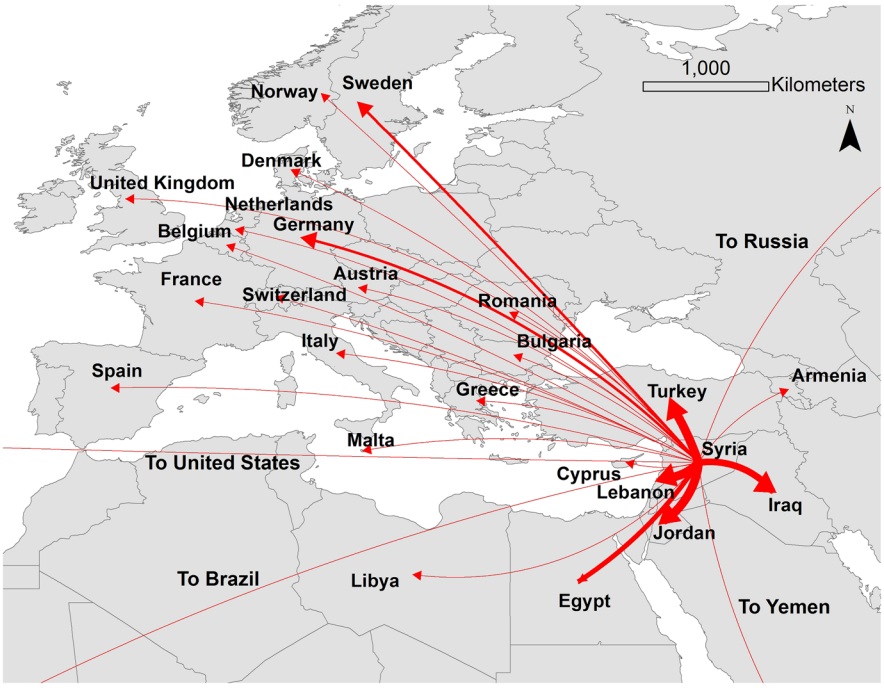

Fig. 1 Cumulative flow (2011–2015) illustrating Syrian forced migration to neighboring countries and other destination countries. Line thickness indicates increasing number of persons migrating (data from: UNHCR 2017a)

combination thereof), from their place of origin toward their perceived safe havens. Much of this surge can be attributed to the Syrian conflict and the resulting refugee crisis. At the peak of the Syrian crisis, in 2015, the International Organization for Migration (IOM) estimated that over 1 million migrants used sea routes to move from the Middle East to the west, primarily Greece and Italy (IOM 2016a). However, this has been a long process. Over the past 6 years there has been a mass forced migration event that has displaced the Syrian population both within the country, as internally displaced persons (IDP), or outside the country as refugees. Based on UNHCR data (2017a), Syria's neighboring countries have been receiving a large proportion of the refugees: the top hosting countries for refugees include: Turkey (2.5 million), Lebanon (1.1 million), and Jordan (0.7 million) (UNHCR 2016). Many of these refugees subsequently sought a passage to the European Union (EU), as illustrated in Fig. 1.

Several factors have been identified in the literature as potential causes of refugee production and mass movement of refugee populations. The factors can be categorized into (1) root conditions, (2) proximate conditions, (3) intervening factors, and a (4) triggering event (Clark 1989). *Root conditions* include poverty, hunger, overpopulation, ethnic animosity, and resource scarcity (Zottarelli 1998). *Proximate conditions* include human rights violations, generalized violence, and political conditions (Schmeidl 1997; Zottarelli 1998). *Intervening factors* either facilitate or inhibit the crossing of

international borders. Examples include transportation and communications systems such as access to flights, road conditions, expected country reception, or seasonal factors (Clark 1989). A final factor in the production of refugees is the *trigger event* which is the catalyst for the forced migration event. The trigger event (e.g., erupting violent civil conflict, political upheaval, change in border control) could be a change in a push factor or intervening factor that is the final variable that convinces a population to migrate versus staying in the same place (Clark 1989). As can be seen from the description of these factors, the mass movement of refugees has strong spatial and temporal components since the migration process itself occurs over space and time and is often influenced by location-driven conditions.

Data related to these factors have been traditionally collected and disseminated by local, regional, or global authorities. However, as noted earlier, new opportunities for collecting migration-related data are emerging. Specifically, VGI and AGI offer a new data source to study forced migration at a variety of spatial, temporal, and thematic scales, complementing traditional data sources. The rapid growth of social media, with Facebook, Instagram, and Twitter reaching 1.9 billion, 600 million, and 317 million active monthly users, respectively (Hutchinson 2017), has presented the general public with a new avenue to generate, disseminate, and consume information. Much of this information, as discussed in Sect. 3, has both a spatial and a temporal dimension. Specifically, these contributions are increasingly being geotagged, as a result of the proliferation of real-time geolocation technology (e.g., GPS-enabled smart phones).

Beyond the use of VGI, personal networks can now sustain the adverse effects of distance along a migration route (Brekke and Brochmann 2015). When refugees were arriving on the shores of Europe, their most precious possession was arguably a smartphone. It served as a portal (potentially, the *only* portal) to connect with their family and friends who were left behind, or who were concurrently traversing different stages of the routes toward safe havens. This enabled refugees to share information about their status and experiences within their social networks, and the world at large document ordeals, navigate using crowdsourced maps, find legal and social services (Ram 2015; Sebti 2016, Weiss-Meyer 2017), and connect with communities in country destinations (Charmarkeh 2013). Refugees, for example, used social media to post online their arrival to shores of a destination country as shown in Fig. 2a or during their migration trajectories (Fig. 2b). In this respect, the traditional view of the data-poor isolated migrant is substituted in the twenty-first century by that of a data-rich networker. These changes highlight the need to study how such new information sources can augment and enhance the more traditional information sources, what potential added value can be gained toward a better understanding of mass migration, and what issues should be considered when implementing such an approach. The typology of mass migration data proposed in Sect. 3 aims to support this goal.

3 A typology for mass migration data

The development of the proposed typology stems from the conceptual link between the process of migration and the designation of a person as a refugee. However, defining who is a refugee is not always easy. For example, Miranda (1990) noted

Fig. 2 Refugees using smartphone technologies during migration trajectories harvested from Flickr: **a** refugees arriving along a coast and documenting this with a smartphone (Flickr 2015a); **b** refugees charging phones at a makeshift charging station (Flickr 2015b)

that the definition of a refugee at the state level may vary due to the legal implications of the term under international and local law, and that as a result, other terms are used. Similarly, Shacknove (1985) noted that while there are dozens of definitions for the term refugee, some of which are jurisdiction dependent, many of these definitions are based on the constructs of the United Nations (UN) convention. Consequently, we adopt the UN definition of a refugee in the context of this paper, as we describe below.

The UN formal definition of a refugee was originally set through the 1951 Refugee Convention, Article 1, stating that a refugee is "A person who owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it" (United Nations Conference 1951).

This definition, which still holds today, encompasses elements that relate to the geographical and the situational context in which a person is situated in order to be considered a refugee. Specifically, it relates to the physical location of a person with respect to a country of origin as well as to the conditions that increase (or mitigate) the risk to the safety and well-being of that person. These elements can, in principle, be operationalized by the systematic collection of four key types of observables which form the proposed typology: *sites and pathways* that relate to the geographical locations of a person, and *conditions and activities* that relate to risk factors to which a person is exposed (and their perception) or activities in which a person is engaged.

In the context of human migration, sites and pathways can be defined based on the basic building blocks of movement data proposed by Andrienko et al. (2008) and the taxonomy of movement patterns proposed by Dodge et al. (2008): a *site*

is a two- or three-dimensional delineated space that has one or more affordances relevant to a moving entity in which the entity is located for a period of time; and a *pathway* is the sequence of time-ordered geographical locations of a moving entity over time. Here, a geographical location along a pathway becomes a site (e.g., a refugee camp) when the time during which an entity is located at a specific location is considerably longer than the time the entity spends between consecutive locations along the pathway. Generally, the term site is used here in relation to geographical space, while the term place is used in relation to the accumulated particular meaning of that space at the individual or group level (Cresswell 2014). It is also important to note that the definition of a pathway is based on the assumption that refugee pathways are likely to be inherently discontinuous, i.e., they are comprised of movement between sites, rather than being a continuous movement path forming a trajectory. In the same vein, we define conditions and activities following Galton (2000): a *condition* is a given state or a property of a location or an entity, and an *activity* is in essence a change in the state or property of a location of an entity. Accordingly, a key difference between a condition and an activity is therefore the occurrence (or the lack) of change. In the context of human migration, it is reasonable to assume that the term activity is often associated with intentional human action.

The typology proposed here is inherently linked to the scale and granularity at which observations are being made. While the United Nations High Commissioner for Refugees (UNHCR) definition of a refugee focuses on the individual person, mass migration is based on the widespread adoption of one or more movement patterns (Dodge et al. 2008). At the same time, the detection and understanding of such patterns depend on the spatial and temporal scale(s) at which data related to movement, conditions, and activities are collected. For instance, in the case of pathways, it is possible to identify scales of organism movement, from movement steps, to movement phases, and lifetime tracks (Nathan et al. 2008).

While efforts to collect data related to the types of observables defined above have traditionally been carried out by authoritative organizations, these efforts are often met with difficulties. Consider, for example, the effort to count refugee populations. According to UNHCR (2013), the collection of data regarding the size and characteristics of refugees and other related populations depends largely on UNHCR field offices, government agencies, and non-government organizations (NGOs) since in many non-industrialized countries, refugee registration is not centralized. This results not only in data fragmentation, but also in an increased risk of entirely missing or misrepresenting parts of the refugee population. Such difficulties are further exacerbated when attempting to collect data regarding conditions and activities, which are often based on limited field surveys and reports, or when the migration is comprised of multiple, distinct sub-flows (Crawley et al. 2016).

The emergence of crowdsourced data provides an opportunity to address these data challenges and mitigate their impact on the ability to better understand mass migration processes. Today, there are already several crowdsourced data sources that can be used to gather information regarding sites and pathways as well as the conditions and activities that take place on the ground. Since much of the crowdsourced data are not produced specifically for migration analysis, they can be both *explicitly* derived (i.e., a VGI data collection or a report of the number of refugees

Table 1 A typology of observables and data sources related to massive migration events

	Sites and pathways		Conditions and activities	
	Traditional	Crowdsourced	Traditional	Crowdsourced
Explicit	UNHCR	OSM refugee locations	Eurostat	IOM
Implicit	The World Bank	Refugee blogs	Relief Web	Twitter, Flickr, Instagram

in a site) and *implicitly* derived (i.e., an evaluation of the safety conditions along a pathway through additional processing, such as content analysis or natural language processing). A key premise of our proposed approach is that rather than viewing crowdsourced data as an alternative source of migration-related information, they are regarded as a *complementary* data source, which can offer several advantages in the study of mass migration. Table 1 provides a summary view of this complementary approach, and Sects. 3.1 through 3.4 discuss how such data could augment traditional data sources.

3.1 Traditional data sources of refugee sites and pathways

There are several sources of traditional authoritative data for refugee sites and pathways, which provide both explicit and implicit data. Examples of such data sources include the UNHCR, U.S. Department of State, Humanitarian Information Unit, and the World Bank. The UNHCR collects and disseminates explicit information on the number of refugees that is based on registration systems, NGO information, and government reporting systems. This organization has existed since 1950 and has a footprint in 126 countries with a mission to assist and protect refugees globally. Refugee application numbers are available on an information sharing portal with an approximate lag time of 2 months. The data are aggregated to a larger administrative scale but can yield insights into past trends with consistent definitions of a refugee. Using such data, the Humanitarian Information Unit at the U.S. Department of State has created a data set of UNHCR refugee camps that contain information on the location, status, and additional descriptive information on the refugee camps for the countries neighboring Syria including Turkey, Iraq, and Jordan. The camp locations provide insight into the physical refugee infrastructure outside the boundaries of Syria.

The World Bank is an example of a source that provides both implicit and explicit data. Focusing on the implicit data components, the World Bank produces migration-related reports and distributes them via an open data portal from which information has to be mined. Examples of such implicit data include reports that have been published by The World Bank such as the Forcibly Displaced: Toward a Development Approach Supporting Refugees, the Internally Displaced, and Their Hosts (2017a, b), The Welfare of Syrian Refugees: Evidence from Jordan and Lebanon (2016), and The Kurdistan Region of Iraq: Assessing the Economic and Social Impact of the Syrian Conflict and ISIS (2015).

The ability to utilize implicit data sources becomes particularly important in the context of refugee sites, since such sources may provide additional information about sites that have not been designated as refugee camps by the UNHCR. During the past two decades, there has been an increasing trend among refugees to live in sites outside designated refugee camps, which could include individual accommodations, self-settled camps, or within the outskirts of urban areas. This issue has been recognized by the UNHCR: according to its report on global trends in forced displacement in 2015 (UNHCR 2016), there are an estimated 52 thousand refugees in Europe and Central Asia living in camp sites while 4 million refugees are living in other sites. A similar trend in the Middle East and North Africa has been observed, where 269 thousand refugees are in camp sites and an estimated 3 million are living in other sites (UNHCR 2016). This trend has created a new challenge in monitoring the magnitude and extent of the current Syrian refugee crisis as refugees have become physically dispersed well beyond the boundaries of refugee camps.

3.2 Traditional data sources of conditions and activities

Open data sources that are available through the Eurostat database portal provide explicit data that are available from a country level to a finer administrative area at a Nomenclature of Territorial Units for Statistics (NUTS) 3 level of 1342 regions (Eurostat 2017). This database provides a range of thematic and statistical machine-readable data sets for population, social conditions, health, asylum migration, and enforcement of immigration (e.g., migrant integration indicators such as living conditions, risk of poverty or social exclusion, education).

An example of implicit open data and information about conditions and activities is the ReliefWeb portal, which provides synthesized and aggregated information for humanitarian support that can be used by NGOs, UN agencies, and governments for numerous types of crises. In the case of the Syrian refugee crisis, ReliefWeb offers direct links to humanitarian data sources (ReliefWeb 2017) as well as reports that provide overviews, humanitarian response plans, and other implicit information about humanitarian efforts documenting the Syrian refugee crisis.

3.3 Crowdsourced data of refugee sites and pathways

Perhaps the most prominent example for crowdsourced efforts to collect explicit information related to refugee camps sites is OpenStreetMap (OSM). OSM provides a platform for collecting and editing geospatial data on a global scale. The data are collected in the format of points (nodes), polylines (ways), polygons (enclosed ways), and relations (Neis and Zipf 2012; Rehrl and Gröchenig 2016). OSM has made digitizing mainstream with an easy-to-use interface that utilizes aerial or satellite imagery along with the ability for users to upload data sets derived from other OSM-compliant open data sources. OSM data are continuously updated by users and edited for accuracy of content by a user base and have in many cases been found to be a suitable alternative to government and commercial mapping products (Haklay 2010; Dorn et al. 2015). As a result, OSM data have been used extensively in

disaster and humanitarian applications and have been populated by stakeholders that include humanitarian organizations, governments, and private citizens (Mooney and Minghini 2017).

Several humanitarian organizations and agencies (e.g., REACH, UNOSAT, Humanitarian OpenStreetMap Team (HOT), Missing Maps project) have adopted the use of OSM for mapping refugee camp areas with common tags and feature attributes for the purposes of sharing information with a common geographical reference system, coordination of effort and documentation, and a "low-tech, low-investment option for humanitarian actors" (OpenStreetMap Wiki 2017). Resources are listed on the OSM wiki for contributors to reference for the mapping of displaced populations to ensure a common framework can be achieved (Baker et al. 2015). The ability to search OSM data using tags revealed over 400 refugee point features that were distributed worldwide. Humanitarian organizations, government agencies, and refugees themselves are using such applications for tracking of supplies and services, resources, and coordination activities with other relief organizations and government organizations (Hanson 2016).

An example of mapping a refugee camp site is that of Za'atari, which is located in Jordan, near the Syrian border. This refugee camp did not exist prior to the Syrian conflict as can be seen from aerial imagery in Fig. 3a and by the lack of features in OSM prior to 2011 as seen in Fig. 3b. Since its creation in July 2012, the Za'atari refugee camp has grown in extent (Fig. 3c) with a population of over 79,000 and has had 69,000 edits in OSM as can be seen in Fig. 3d (UNHCR 2017b; OSM 2017). On closer inspection of what has been mapped with respect to the Za'atari refugee camp, one can see in Fig. 4, medical facilities, mosques, humanitarian services, schools, playgrounds, and the street network. In this example, the information shown was contributed by various stakeholders, including humanitarian organizations on the ground and remote users from around the world, thus illustrating the multifaceted nature of the crowdsourcing effort. In Fig. 3d, OSM edits originated from humanitarian organizations in Jordan and remote users from countries including France, Germany, Austria, Russia, and Indonesia.[1]

While OSM provides an online mapping platform, there are a number of mobile OSM editor applications (i.e., Vespucci, StreetComplete, OsmAnd)[2] including the MAPS.ME (MAPS.ME 2017) application which is a popular OSM editor that enables downloads of OSM data that can be used offline without a wireless internet connection. Additionally, the user can create new OSM edits that can be uploaded to the OSM database for use by the online community. It is worth noting that while there are many OSM editors, these editors may vary in their functionality. For example, the edits by MAPS.ME users are nodes tagged with an identification that attributes edits by these users to the created geometric features. Features often created with MAPS.ME include telephone locations, community centers, cafés, convenience

[1] For users interested in finding information on OSM user contributions, please visit https://hdyc.neis-one.org.

[2] For more information about mobile OSM editors, see https://wiki.openstreetmap.org/wiki/Comparison_of_editors.

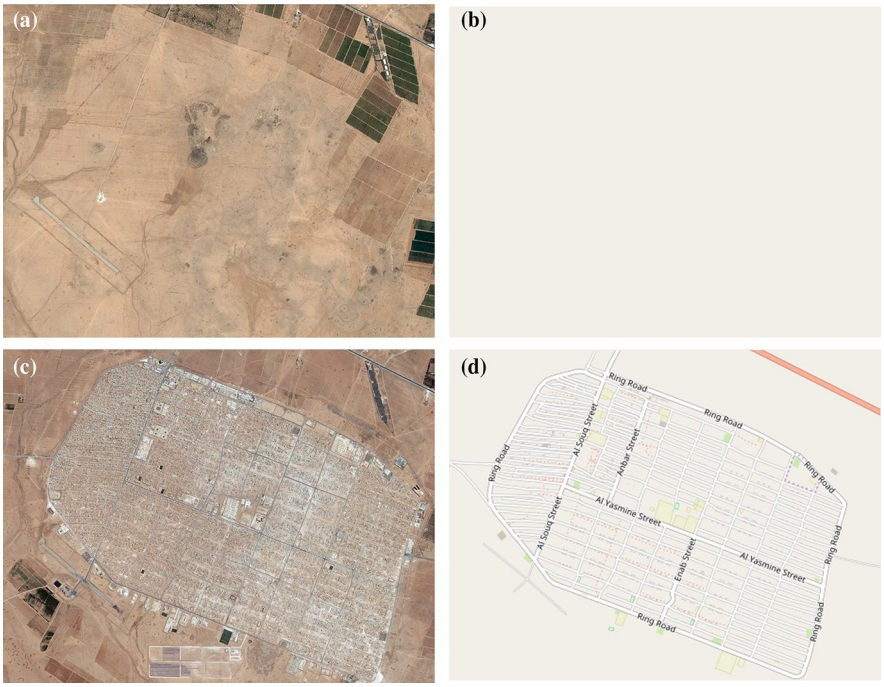

Fig. 3 OSM data and satellite imagery for the Za'atari Refugee Camp in Jordan: **a** image courtesy of Google, DigitalGlobe from September 2011; **b** ©OpenStreetMap contributors from 2011 (no features were present); **c** image courtesy of Google, DigitalGlobe from August 2017; **d** ©OpenStreetMap contributors from August 2017. OSM is available under the Open Database License (ODbL)

shops, electronics shops, schools, hospitals, camp sites, guesthouses, places of worship, caravan sites, mobile phone shops, ATMs, and clinics.

Other crowdsourced data sources that show the explicit refugee camp sites include Wikipedia and Wikimapia. In these sources, the crowdsourced entries for refugee camp extent and composition have been digitized and attributes have been provided such as descriptive links for further information, coordinates of the refugee camp, satellite images, and ground photographs. Additional information on the individual edits for the entries is detailed to show the history and timeline of digitizing the extent, streets, photographs, and user-uploaded descriptions of the place or point of interest.

Blogs contribute a wealth of implicit information pertaining to sites used by refugees through content generated by a diverse online community comprising journalists, NGOs, humanitarian organizations, activists, and individuals that want to share experiences and information. Such user content is in the form of text, digital photographs or images, and possibly video, and news that is not reported in the mainstream press. Textual information that is mined from blogs can include place names, movement trajectory information, the sentiment of the host population, and the conditions at a potential destination.

Fig. 4 OpenStreetMap data as of August 2017 for Za'atari refugee camp, Jordan

While data about the location and the characteristics of sites, such as refugee camps or intermediate stops along the migration route, can be derived from several sources, crowdsourced data (and in particular VGI) about migration pathways are more scarce. This can be attributed, at least in part, to concerns related to safety and privacy, as well as to data quality issues. In terms of safety and privacy, it is important to recognize that vulnerable populations such as refugees are often reluctant to share detailed location data due to concerns that such data can be used later on to increase their vulnerability or compromise their safety (Latonero and Kift 2018). From a data quality perspective, tracking human movement using crowdsourcing or social media platforms may result in inaccurate results due to the potential inconsistency between a platform user and an individual person, and due to the relative low percentage of GPS-based geotagged data in some platforms (Stefanidis et al. 2017). Notably, Hübl et al. (2017) recently proposed a method for tracking migration patterns as a sequence of movements between discrete geographical regions (Dodge et al. 2008).

3.4 Crowdsourced data on conditions and activities

By its very nature, successfully migrating to a new site (e.g., a new country) requires an understanding of the conditions refugees might encounter, as well as the types of activities refugees should engage in or avoid (Valtonen 2004). Often, this entails

gathering information about the conditions on the ground with respect to safety along pathways and intermediate sites, the availability of resources and amenities at such sites, or the collection of documents for crossing international borders. At the same time, migrants are often required to engage in activities such as establishing connections with potential support providers (e.g., relatives and friends), finding temporary accommodation, or the establishment of a path to citizenship. In the context of these conditions and activities, geographical space has emerged as key to the movement and settlement of refugees as well as to their well-being. For instance, in studying youth with refugee backgrounds, Sampson and Gifford (2010) have identified four different types of places that provide conditions for supporting the well-being of young refugees: places of opportunity, places of restoration, places of sociality, and places of safety. In addition to such places, migration-related activities and processes are often supported by underlying communication and social networks. Considering social networks as an example, Komito (2011) suggested that while Web 2.0 technologies such as social media can support the migration process and provide the means for maintaining relations with other migrants, they can also change and even hinder the integration and participation of migrants in their new environment. Monitoring the confluence of *places*, *networks*, and *people* is therefore important for monitoring and exploring the migration of refugees.

In the case of the Syrian refugee crisis, there are various explicit crowdsourced data sources that can shed light on the conditions and activities that can be useful in the context of migration. An example of such a source that focuses on places is The Migrants' Files (2014), which was an effort to quantify the number of migrants that die while on their migration paths across the Mediterranean that was led by journalists. This effort included the compilation of data from open source published accounts and documents regarding migrant deaths over the past 16 years. Attribute information contained in the data set includes the event location, number of dead or missing migrants, cause of death, source of information, description of the event, and source URL for fact checking the death information. Figure 5 shows an example of this type of information, in which death cases along the Syria–Turkey border and in the Mediterranean Sea (particularly along the coastlines of Libya and Tunisia) are dominant. Information such as this can reveal areas where risk to migration activities is particularly high. Since this effort has ceased in June 2015, there has been another effort by the International Organization for Migration (IOM) to provide explicit migrant-related information. Specifically, IOM's focus is on providing reliable migrant death documentation that is based on open sources including traditional media, migrant interviews, and social media reports (IOM 2016b).

Similarly, information about networks, such as social networks of migrants, can also be derived from crowdsourced data. In past studies on migration, social networks have been found to impact all stages of migration (Boyd 1989). While in the pre-internet era, social networks did provide the means to spread information, preserve community ties, and provide information that guided others in their decision to migrate (Mabogunje 1970), current Web 2.0 technologies provide a medium for delivering implicit information at an increasing rate and wider dispersal. For example, the utilization of AGI derived from social media platforms such as Twitter were shown to provide opportunities to quantify demographic

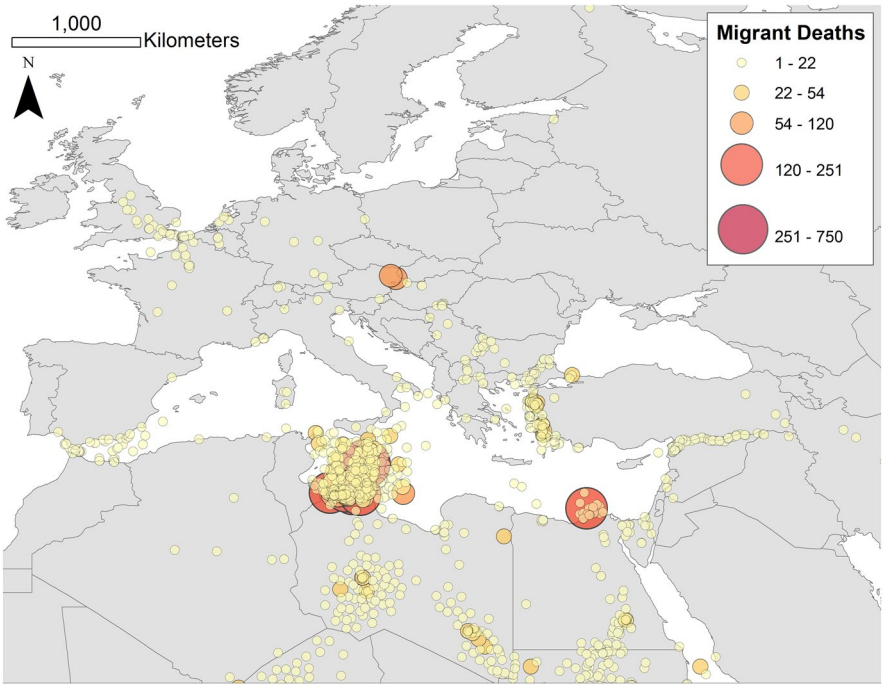

Fig. 5 Migrant deaths that have been documented and crowdsourced by journalists and international relief organizations since 2000 (data from: The Migrants' Files 2014; IOM 2016a)

information such as gender, race/ethnicity, and age (Mislove et al. 2011; Schwartz et al. 2013; Sloan et al. 2015) in migrant populations. In addition, analyzing social media content can shed light on what is the users' sentiment at different locations, which trends emerge, and how information propagates across locations (Stefanidis et al. 2013b; Zhong et al. 2014). An example of the latter is shown in Fig. 6, which depicts a network of retweets mentioning the term "refugee" in multiple languages between May and August 2017. This network spans from North America to Europe, the Middle East, Africa, southeast Asia, and Australia.

Like Twitter, Instagram is another social media platform in which users share migration-related information. Instagram, which was first released in 2010, enables its users to share photographs, videos and text (Seltzer et al. 2015). The potential of Instagram as a source for implicit refugee-related information is demonstrated in Fig. 7, which illustrates user-provided content (10,029 geolocated posts out of a total of 60,301 posts) that was collected between February 2011 to March 2016 using the keyword "refugee." Focusing on Europe, the Middle East, and Africa, it is particularly interesting to note the concentration of Instagram images along borders (e.g., the Syria–Turkey border, the Syria–Lebanon border, and the Spain–France border), as well as along some coastlines (e.g., the areas around the Aegean Sea, and the southern coast of Italy). The figure also shows that in terms of Instagram post density, the highest densities are found in Greece, Turkey, and Spain, followed by Jordan, Hungary, Italy, Germany, and the UK.

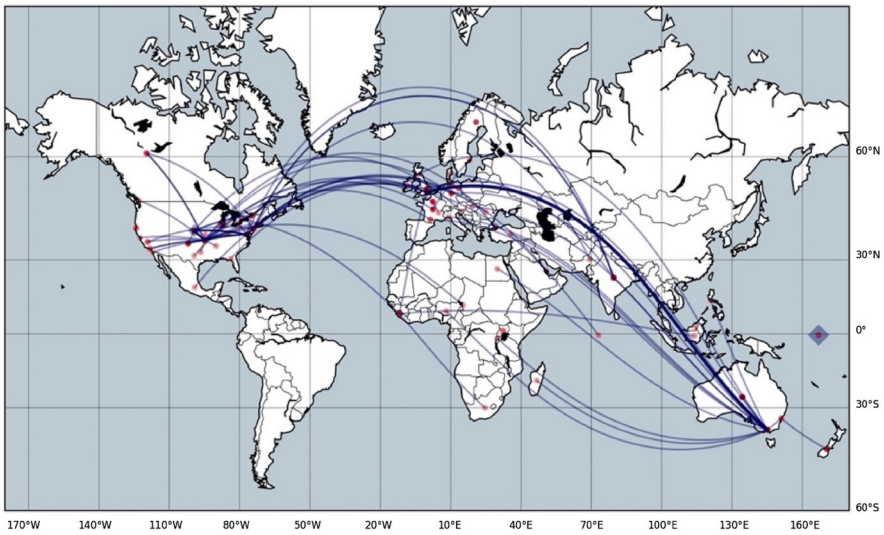

Fig. 6 Retweet network of geolocated Twitter microblogs that are discussing opinions, news and retweeting information related to "refugee" in multiple languages from May to August 2017

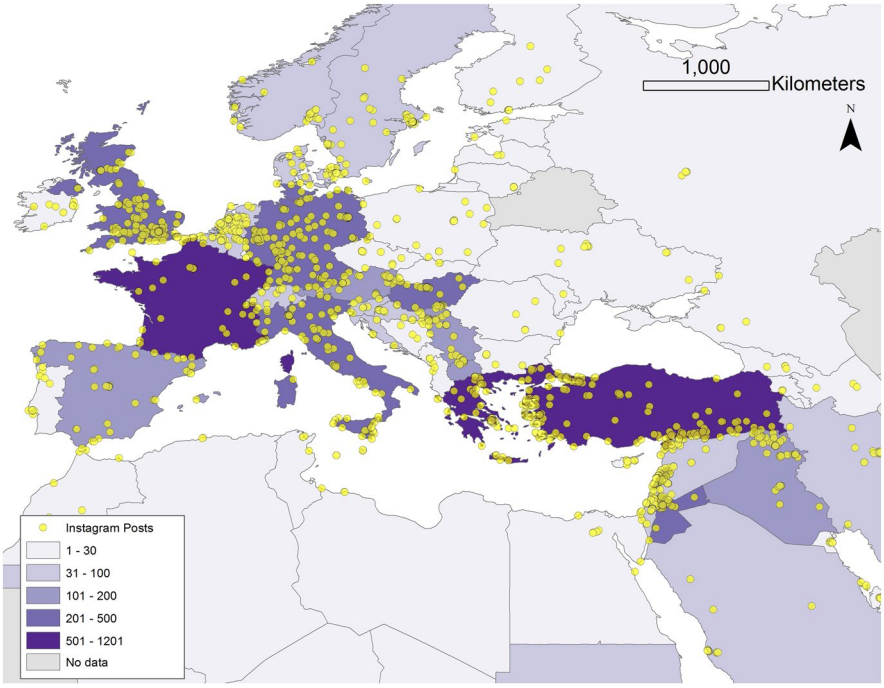

Fig. 7 10,029 geolocated Instagram posts referencing "refugee" from February 2011 to March 2016

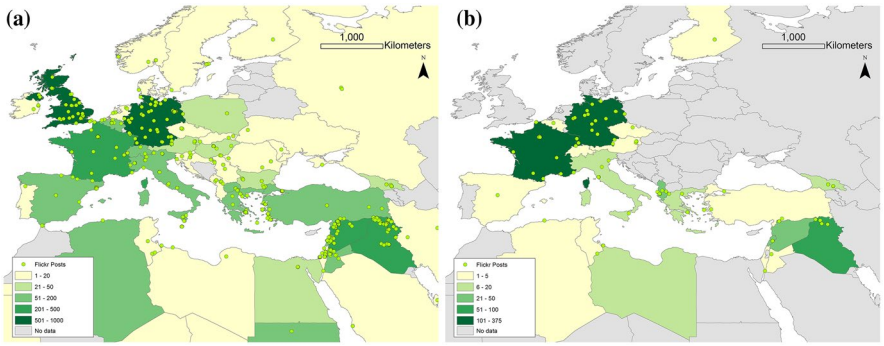

Fig. 8 Flickr posts referencing "refugee": **a** Flickr posts referencing "refugee" in the English language; **b** Flickr posts referencing translation of refugee in other languages ("ىئاجلا," "laji," "mülteci," "πρόσφυγας," "prósfygas," "flüchtling," "réfugié," "profugo")

Many of these countries and regions correspond well with known destinations of Syrian refugees (as shown in Fig. 1).

AGI derived from one social media source about refugee activities can often be corroborated by data from another social media source in order to provide additional support for deriving process-related knowledge from such data. In the case of the Syrian forced migration, multi-source corroboration can be achieved by augmenting the already available Twitter (Fig. 6) and Instagram data (Fig. 7) with data harvested from Flickr. In this case, Flickr data containing the keyword "refugee" in English were collected separately from Flickr data containing the keyword "refugee" in several other languages (e.g., French, German, Greek, etc.). The results of this data collection are shown in Fig. 8, which depicts Flickr photograph data density across Europe, the Middle East, and North Africa in English (Fig. 8a), and the other selected languages (Fig. 8b). Notably, the countries highlighted in these results are in line with the results obtained from Instagram in that they also highlight Greece, Turkey, Spain, Hungary, Italy, Germany, and the UK. However, in the Flickr data, Algeria, Sudan, and Iraq are shown to have more substantial photograph density, thus highlighting other significant migration processes [e.g., the flow of sub-Saharan African refugees from Algeria to Europe (Houttuin and Huson 2016)]. Furthermore, a comparison of the Flickr data in English versus data in the other languages highlights how studying mass forced migration through language use can reveal finer details about where migration-related processes and discourse occur.

In addition to the social media content itself, it is also important to recognize that the content contribution rate itself may also provide additional insight into the underlying conditions and activities that drive and shape mass migration. In particular, notable change in the rate of contributions in a social media service is capable of capturing the crowdsourced response to changes in the physical conditions on the ground, policy changes of host countries, or highlight risks along migration pathways. An example of this approach is shown in Fig. 9, which depicts a time series of monthly totals of Instagram posts with the keyword "refugee" from February 2011 to March 2016. In this time series, peaks in monthly contribution rates

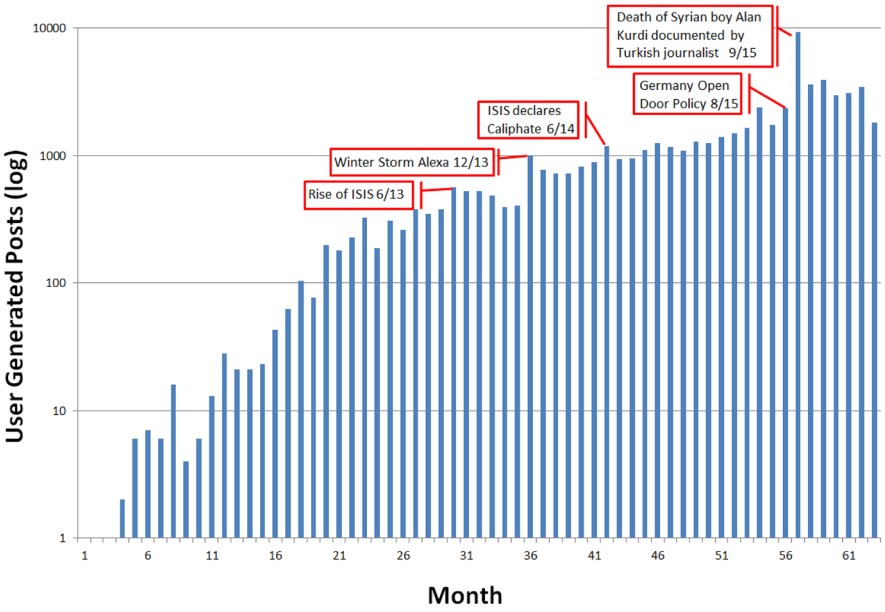

Fig. 9 Time series of monthly totals of Instagram posts referencing "refugee" from February 2011 to March 2016 plotted on a logarithmic scale. Months are designated by numbers along the horizontal axis (e.g., January 2011 = 1, March, 2016 = 63)

can be mapped to key events in the Syrian refugee crisis during the corresponding months, including geopolitical events, social uprisings, weather-related events that can impact refugees, and social events that increase awareness of a migration crisis. Such rate-to-event correspondence can be further refined to a specific date by exploring changes in the daily contribution rates.

4 Digging deeper into mass migration data

AGI and VGI are viable data sources for the study of forced migration due to the ever-increasing foundational information and network connections that are being constantly crowdsourced and provided by social media users. VGI data, such as OSM, have shown that user-generated content can dramatically increase the infrastructure that is mapped both with a quantity that rivals authoritative government and commercial sources and a quantity that would not be sustainable considering financially constrained government resources. The nature of OSM data creation improves over time with the addition of features and attributes that can rapidly provide information for use by a humanitarian application or disaster response scenario (as we showed in Sect. 3).

However, the use of VGI can also offer a more holistic view of unfolding events as there is additional contextual information provided by such data that can yield insight into a deeper understanding of place as it relates to changes in policy or

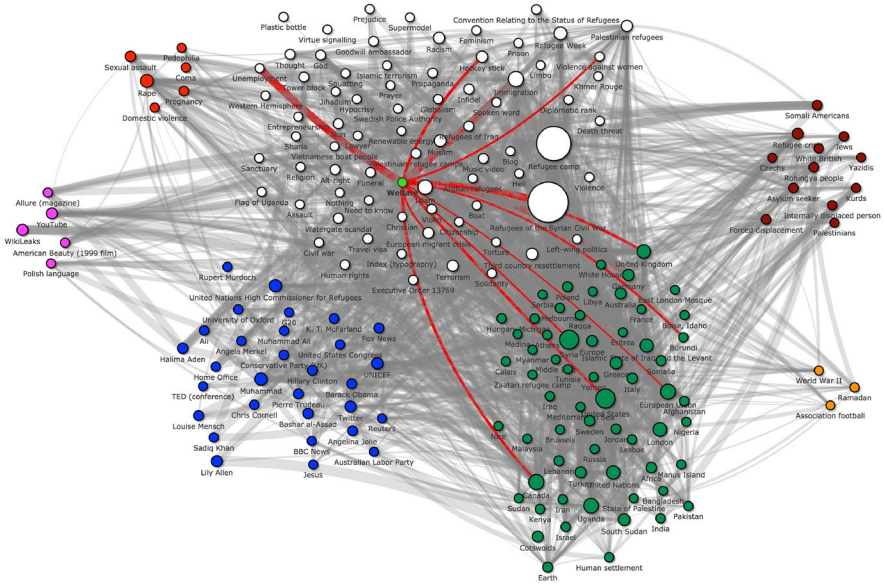

Fig. 10 A concept graph illustrating the associations between a keyword related to root factors of mass migration such as poverty ("welfare") to other keywords, as they appear in our Twitter data corpus. The color of the node refers to specific themes: locations (green), actors (dark red), topics (red), entities and individuals (blue), concepts (white), and events (yellow). Red edges represent active associations between terms; gray edges represent inactive associations between terms (color figure online)

variations in migration patterns. In this section, we specifically explore how such data can be mined to explore the factors that cause refugee production and movement (Sect. 4.1) and how VGI along with other open data sources can be used to model the flow of migration (Sect. 4.2). This moves beyond just analyzing the contribution patterns to that of analyzing the root causes of the crisis and modeling the paths migrants take, informed by data.

4.1 Concept graphs

As noted in Sect. 2, there are many factors that cause the mass migration of refugees. Social media such as Twitter provides a way to explore this from the bottom up (i.e., user-contributed information) in the form of text narratives as shown in Fig. 10. Through data mining techniques, we can identify *actors*, *locations*, and *concepts* with respect to the refugee crisis. An actor can be broadly defined as a twitter user, locations can be derived from the tweet meta data, and concepts (i.e., keywords) that frame the narrative can be derived through the identification of named entities using a reference ontology, in this case DBpedia (Auer et al. 2007, dbpedia 2017). Readers wishing to know more about this process are referred to Stefanidis et al. (2017). Through the collection of tweets (using the keyword search "refugee") via the Twitter API feed, a retweet analysis and subsequent concept graphs were constructed to

explore the frequency and interconnections between locations (green), actors (dark red), topics (red), entities and individuals (blue), concepts (white), and events (yellow) in Fig. 10. In this figure, the size of the node reflects how frequently a particular term was encountered in the data corpus and the edges between nodes are either red or gray. Red links are directly related to specific terms (e.g., terms frequently appearing in the same tweet) and gray edges represent co-occurrences of words which have not been clustered in the same topic. Such analysis allows for the key concepts that have been captured within the Twitter narrative for the Syrian refugee crisis. These include the root conditions such as welfare (Fig. 10). Examples of other concept graphs are provided in the online resources of this paper.

4.2 Modeling refugee movement

Moving beyond data analysis, there is also the question of how such data can be used for modeling. For example, can the number of migrants and their paths be predicted? While migration models have a long tradition in the geographical sciences (e.g., Stillwell 1978), there are few that have been constructed using structured and unstructured data formats. Through the growth of VGI, we now have at our disposal information not only on the factors that cause migration (as highlighted in Sect. 4.1) but also information about the camp sites and pathways that migrants use and activities and processes they utilize during their migration (as highlighted in Sect. 3). Through such data and analysis, we now have the opportunity to construct highly detailed individual-based models of migration whereby individual micro-level agents with diverse behaviors can provide insight to macro-level patterns of movement (Crooks et al. 2018; Hübl et al. 2017, Edwards 2008). A recent example application of a geographically explicit agent-based model is shown in Fig. 11. This open source model was created to explore individual movement pathways for the current Syrian refugee crisis utilizing open data on the number of refugees, camp sites, and migration pathways across the Middle East, north Africa, and Europe. The sources of these data included UNHCR, IOM, and OSM, and the individual goal selection for each agent was set in accordance with the theory of intervening opportunities (Stouffer 1940). Based on these data, the model was then calibrated using a parameter sweep approach. (Details about this model are available in Hu 2016.) Results from such a model could provide guidance for policy and humanitarian decisions. For example, such models could be used to explore "what if" questions, e.g., if a certain border is closed, how will the refugees change their routes, and where should humanitarian agencies place aid centers?

5 Discussion and concluding remarks

In this paper, we used the recent mass migration that resulted from the ongoing Syrian humanitarian crisis to showcase the type of migration-relevant information that can be harvested from VGI and social media. The information that is gleamed from such sources, complemented by traditional open data resources, can provide rich

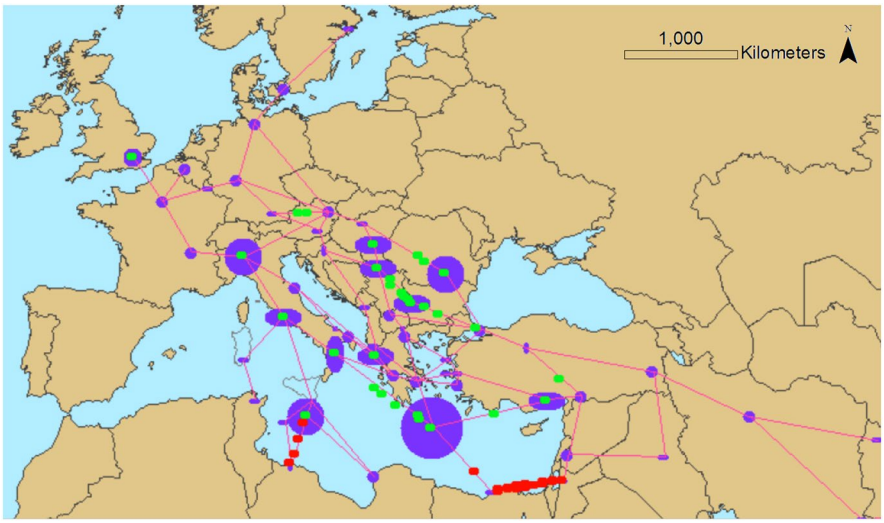

Fig. 11 An agent-based model of migration: top: the spatial environment, where the lines represent migration pathways, and the nodes represent number of migrants. Purple nodes represent final destination sites, red nodes show migrant deaths, and green nodes show migrants *en route* (*source*: Hu 2016)

information on migration processes as they occur. When collected in fine spatial, temporal, and thematic resolution, such information can lead to a better understanding of migration and can inform the broad and diverse stakeholder communities that are involved with this issue. However, along with the opportunities to do so, there are also several key challenges that are associated with using VGI and AGI sources. Below we address three such challenges which we believe are essential in the context of the use of crowdsourced data from mass migration, along with a discussion on the key benefits of utilizing crowdsourced information in the context of mass migration.

With the increased use of VGI and social media, data quality has emerged as a key concern. A number of researchers have addressed the relative quality of geographical crowdsourced data sets by comparing VGI to corresponding authoritative data sets (Haklay 2010; Dorn et al. 2015; See et al. 2016). In a similar manner, other crowdsourced data sets such as Wikipedia, blogs, and microblogs can also vary in terms of content quality (Kittur and Kraut 2008). Quality issues are related to a variety of factors, including positional accuracy, completeness, quantitative and qualitative attribute values, and generalizations (Fonte et al. 2017). At the same time, it should be recognized that traditional data sources are also affected by data quality issues. UN agencies, governments, and some NGOs have long been providers of migration- and refugee-related data. However, such data can often be affected by count inconsistencies and classification subjectivity (Nature 2017; Dijstelbloem 2017).

However, the same principle of the "wisdom of the crowd" that has served as the foundation of crowdsourced efforts could also be applied to the availability of diverse crowdsourced platforms. By integrating multiple such sources, we are able

to alleviate the accuracy shortcomings of each, as we showed in Sect. 3.4. One could argue that this is an extension of the concept of the wisdom of the crowd to the wisdom of multiple crowds. Furthermore, the gap between authoritative and crowdsourced data is becoming narrower and blurred. Authoritative data providers are increasingly becoming active participants in the creation and promotion of crowdsourced data as well, as we observed in Sect. 3.3 regarding the humanitarian blogs and Twitter feeds from organizations such as the UNHCR and Red Cross. Therefore, the complementary use of these data sources can support the validation and verification of their content (see Antoniou et al. (2016) regarding land cover mapping).

Another challenge relates to data bias. As VGI and AGI rely on crowdsourced efforts, such data can be affected by bias, which can arise from the digital divide (Norris 2001; Van Deursen and Van Dijk 2014). For instance, spatial disparity has been documented regarding user contribution of VGI data (Yang et al. 2016), while social media has been shown to have a spatial disparity, gender, and geographically dependent race/ethnicity bias (Mislove et al. 2011). There are also limitations in demographic information based on census that include static aggregated snapshots in time, coarser scale, low periodicity, fixed locations, limitation on nighttime or daytime collected data, and lack of information on moods (Crooks et al. 2015). Furthermore, bias regarding the perception of mass migration can emerge due to external factors. For example, public awareness of migration flows in Europe has been found to be inexact and variable due to media influence (Brovelli et al. 2017). As a result, such bias factors can propagate into the crowdsourced data, e.g., by focusing on certain geographical locations or on specific populations more than others. While this may, at least in part, be mitigated through the integration of diverse data sources, further research into the impact and mitigation of such bias factors is required.

A third challenge relates to the agility of both traditional and crowdsourced data generation. The use of crowdsourced data offers an obvious advantage in comparison to their authoritative counterparts in terms of their ability to provide near real-time information about mass migration over space and time. However, the dynamic nature of the data, along with its production rate and heterogeneity, poses big data-like challenges regarding the timely processing and analysis of high volumes of heterogeneous data streams, and the dissemination of the derived information from it. Similarly, in the realm of traditional data sources, governments and large organizations around the world have recognized the need to develop big data applications to support key government operations and functionalities (Kim et al. 2014).

Notwithstanding these challenges, we believe that the potential benefits outweigh these challenges. As stated earlier, migration-related information is a commodity, especially so during complex and challenging processes such as mass migration (Wall et al. 2017). Humanitarian agencies and organizations are increasingly relying on up-to-date information for their actions (Palen et al. 2007), and in that sense, timeliness is emerging as an additional parameter of the overall data quality. Arguably, this is an area where crowdsourced data show a clear advantage compared to more traditional sources. Furthermore, the participatory nature of the crowdsourced paradigm is uniquely suited for refugees, who find themselves physically stranded but struggle to avoid information deprivation (Watson et al. 2015). Toward that goal, mobile devices and the apps they

support are becoming essential during refugee migration, ensuring connectivity to family and friends, and finding critical information to aid their migration and settlement (Rao 2015). Advancing this paradigm, humanitarian organizations are increasingly providing GIS training and resources to refugees. In Camp Za'atari in Jordan, for instance, formal GIS training has been provided to refugees to assist the community with safety, planning, and decision making in the allocation of resources and infrastructure (Tomaszewski 2017). In turn, such data can also inform refugees arriving to a camp about the availability and location of key facilities and services (e.g., as shown in Fig. 4).

In the broader context of mass migration, there is little doubt that migration-related data will become even more essential than they already are today. Such data are required in a myriad of application areas, from informing decision-making processes, to monitoring and early detection of increased risk to migrants' well-being and mass migration modeling and forecasting. However, these data requirements often remain largely (and in some cases entirely) unmet, resulting in a critical data gap. Furthermore, given the increased complexity of the factors that trigger and intensify mass migration processes, in terms of both their scope and their granularity, the mass migration data gap is continually widening. The need for resolving this gap has been well recognized: a recent UN resolution on the global agenda for sustainable development for the year 2030 (United Nations 2015) has identified data as a systemic issue for global sustainable development. Moreover, it has called on nations to significantly increase the availability of high-quality, timely, accessible, and reliable disaggregated data on migratory status and geographical location in developing countries by 2020. Similarly, a recent report by The World Bank (2017b) has specifically identified data availability, quality, and reliability, as critical for managing large forced displacement events. The report has called for a substantial coordinated effort to enhance the availability, reliability, and comparability of data on forced displacement. The report also identified a need for nations to move toward an open data paradigm through coordination of collection efforts and harmonization of multiple data sources.

Forecasting the number of future refugees in the world is difficult. The processes and factors that lead to mass migration are often volatile and can be influenced by unexpected trigger events. However, given the increase in the number of refugees and asylum seekers around the world over the last several years (UNHCR 2016), it is likely that the challenge of collecting data related to mass migration and refugees will become even more significant. Given the potential of currently existing open and crowdsourced data—and in particular VGI and AGI—as showcased in this paper, and given their rapidly evolving nature, we believe that such data will increasingly play a vital role in filling the mass migration data gap in the next 10 years and beyond. Within a narrow scope of studying mass migration, VGI and AGI can provide essential data regarding camp sites and pathways between sites, as well as information about the conditions and activities on the ground. In the broader scope of studying mass migration, such data can inform the nexus of a larger scientific inquiry to develop a better understanding of the roots, evolution, and consequences of mass migration in the environmental, economic, social, cultural, and behavioral dimensions. Ultimately such data can pave the way to a new cross-disciplinary

migration information science, which will adopt a much-needed holistic approach to the study of mass migration.

Acknowledgements We would like to thank the four anonymous reviewers for providing invaluable comments and suggestions.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Alencar A (2017) Refugee integration and social media: a local and experiential perspective. Inf Commun Soc. https://doi.org/10.1080/1369118X.2017.1340500

Andrade AD, Doolin B (2016) Information and communication technology and the social inclusion of refugees. Manag Inf Syst Q 40(2):405–416

Andrienko N, Andrienko G, Pelekis N, Spaccapietra S (2008) Basic concepts of movement data. In: Giannotti F, Pedreschi D (eds) Mobility, data mining and privacy—geographic knowledge discovery. Springer, Berlin, Heidelberg, pp 15–38. https://doi.org/10.1007/978-3-540-75177-9

Antoniou V, Fonte CC, See L, Estima J, Arsanjani JJ, Lupia F, Minghini M, Foody G, Fritz S (2016) Investigating the feasibility of geo-tagged photographs as sources of land cover input data. ISPRS Int J Geo Inf 5(5):64. https://doi.org/10.3390/ijgi5050064

Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R (2007) DBpedia: a nucleus for a web of open data. In: Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (eds) The semantic web. Springer, New York, pp 722–735

Baker I, Card B, Raymond N (2015) Satellite imagery interpretation guide: displaced population camps. Harvard Humanitarian Initiative. Retrieved on 24 July 24 2017, from https://tinyurl.com/yayxfufo

Boyd M (1989) Family and personal networks in international migration: recent developments and new agendas. Int Migrat Rev 23(3):638–670

Brekke JP, Brochmann G (2015) Stuck in transit: secondary migration of asylum seekers in Europe, national differences, and the dublin regulation. J Refug Stud 28(2):145–162

Brovelli MA, Minghini M, Kilsedar CE, Zurbarán M, Aiello M, Gianinetto M (2017) Migrate: a foss web mapping application for educating and raising awareness about migration flows in Europe. Int Arch Photogramm Remote Sens Spat Inf Sci 42(4):51–55

Charmarkeh H (2013) Social media usage, Tahriib (Migration), and settlement among Somali refugees in France. Refug Can J Refug 29 (1). https://tinyurl.com/zq3qu2z

Chatfield AT, Scholl HJJ, Brajawidagda U (2013) Tsunami early warnings via twitter in government: net-savvy citizens' co-production of time-critical public information services. Gov Inf Q 30(4):377–386

Clark L (1989) Early warning of refugee flows. Refugee Policy Group. Washington DC. Retrieved on 16 Mar 2017, from https://tinyurl.com/zr2ouax

Crawley H, Duvell F, Sigona N, McMahon S, Jones K (2016) Unpacking a rapidly changing scenario: migration flows, routes and trajectories across the Mediterranean. Unravelling the Mediterranean Migration Crisis. Research brief no. 1, March 2016. Retrieved on 12 Mar 2018, from https://tinyurl.com/y8ar7jzu

Cresswell T (2014) Place. In: Lee R, Castree N, Kitchin R, Lawson V, Paasi A, Philo C, Radcliff S, Roberts SM, Withers C (eds) The sage handbook of human geography. Sage, Thousand Oaks, pp 3–44

Crooks A, Pfoser D, Jenkins A, Croitoru A, Stefanidis A, Smith D, Karagiorgou S, Efentakis A, Lamprianidis G (2015) Crowdsourcing urban form and function. Int J Geogr Inf Sci 29(5):720–741

Crooks AT, Malleson N, Wise S, Heppenstall A (2018) Big data, agents and the city. In: Schintler LA, Chen Z (eds) Big data for urban and regional science. Routledge, New York, pp 204–213

dbpedia (2017) mappings.dbpedia. Ontology Classes. Retrieved on 2 Oct 2017, from https://tinyurl.com/obzs5d7

Dekker R, Engbersen G (2014) How social media transform migrant networks and facilitate migration. Glob Netw 14(4):401–418

Dijstelbloem H (2017) Migration tracking is a mess. Nature 543:32–34. https://doi.org/10.1038/543032a
Dodge S, Weibel R, Lautenschütz AK (2008) Towards a taxonomy of movement patterns. Inf Vis 7(3–4):240–252
Dorn H, Törnros T, Zipf Z (2015) Quality evaluation of VGI using authoritative data—a comparison with land use data in Southern Germany. ISPRS Int J Geo Inf 4(3):1657–1671
Edwards S (2008) Computational tools in predicting and assessing forced migration. J Refug Stud 21(3):347–359
Eurostat (2017) Eurostat: your key to European statistics. Retrieved on 16 Mar 2017, from https://tinyurl.com/jt4q84d
Flickr (2015a) Refugees charging phones at a makeshift charging station courtesy of Joshua Zakary. Retrieved on 2 Oct 2017, from https://tinyurl.com/y6woev3y
Flickr (2015b) Refugees arriving at the greek island of kos courtesy of Christopher Jahn, International Federation of the Red Cross and Red Crescent. Retrieved on 2 Oct 2017, from https://tinyurl.com/y7x5wze3
Fonte CC, Antoniou V, Bastin L, Estima J, Arsanjani JJ, Bayas J-CL, See L, Vatseva R (2017) Assessing VGI data quality. In: Foody G, See L, Fritz S, Mooney P, Olteanu-Raimond A-M, Fonte CC, Antoniou V (eds) Mapping and the citizen sensor. Ubiquity Press, London, pp 137–163
Galton A (2000) Qualitative spatial change. Oxford University Press on Demand, Oxford
Goodchild M (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221
Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ Plan B 37(4):682–703
Hanson K (2016) Refugee crisis? There's an app for that. Londonist. Retrieved on 3 Sept 2017, from https://tinyurl.com/y7vtp4c3
Herfort B, de Albuquerque JP, Schelhorn SJ, Zipf A (2014) Exploring the geographical relations between social media and flood phenomena to improve situational awareness. In: Joaquín Huerta J, Schade S, Granell C (eds) Connecting a digital Europe through location and place. Springer, New York, pp 55–71
Houttuin S, Huson E (2016) Algeria: the new migrant staging post for Europe. IRIN: The Inside Story on Emergencies. October 25. Retrieved on 2 Oct 2017, from https://tinyurl.com/ybgwo77e
Hu E (2016) An agent-based model of the European refugee crisis. GitHub repository, https://github.com/elizabethhu/refugee-abm. Retrieved on 22 Sept 2017, from https://tinyurl.com/yd5s6sjc
Hübl F, Cvetojevic S, Hochmair H, Paulus G (2017) Analyzing refugee migration patterns using geotagged Tweets. Int J Geo Inf 6(302):1–23
Hutchinson A (2017) Top social network demographics 2017. Social Media Today, March. Retrieved on 3 Sept 2017, from https://tinyurl.com/y75qvt7m
International Organization for Migration (IOM) (2016a) Mediterranean migrant arrivals reach 358,403; Official Deaths at Sea: 4,913. December 23. Retrieved on 16 Mar 2017, from https://tinyurl.com/j6larhb
International Organization for Migration (IOM) (2016b) Missing migrants project: tracking deaths along migratory routes worldwide. Retrieved on 16 Mar 2017, from https://tinyurl.com/hgus5jz
Kim GH, Trimi S, Chung JH (2014) Big-data applications in the government sector. Commun ACM 57(3):78–85
Kittur A, Kraut RE (2008) Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In: Proceedings of the 2008 ACM conference on computer supported cooperative work, pp 37–46
Komito L (2011) Social media and migration: virtual community 2.0. J Am Soc Inf Sci Technol 62(6):1075–1086
Latonero M, Kift P (2018) On digital passages and borders: refugees and the new infrastructure for movement and control. Soc Media Soc. https://doi.org/10.1177/2056305118764432
Mabogunje AL (1970) Systems approach to a theory of rural–urban migration. Geogr Anal 2(1):1–18
Maitland C, Xu Y (2015) A social informatics analysis of refugee mobile phone use: a case study of Za'atari Syrian refugee camp. In: Proceedings of the 43rd research conference on communication, information and internet policy. Retrieved on 2 Oct 2017, from https://tinyurl.com/ydhzsl68
MAPS.ME (2017) Free offline world maps. Retrieved on 10 Mar 2018, from https://tinyurl.com/y8tvkq23
Miranda CO (1990) Toward a broader definition of refugee: 20th century development trends. Calif West Int Law J 20(2):9
Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN (2011) Understanding the demographics of Twitter users. In: Proceedings of the 5th international AAAI conference on weblogs and social media (ICWSM), Barcelona, Spain, pp 554–557

Mooney P, Minghini M (2017) A review of OpenStreetMap data. In: Foody G, See L, Fritz S, Mooney P, Olteanu-Raimond A-M, Fonte CC, Antoniou V (eds) Mapping and the citizen sensor. Ubiquity Press, London, pp 37–59

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci 105(49):19052–19059

Nature (2017) Data on movements of refugees and migrants are flawed (editorial). Nature 543:5–6. https://doi.org/10.1038/543005b

Neis P, Zipf Z (2012) Analyzing the contributor activity of a volunteered geographic information project—the case of OenStreetMap. ISPRS Int J Geo Inf 1(3):146–165

Norris P (2001) Digital divide: civic engagement, information poverty, and the internet worldwide. Cambridge University Press, Cambridge

O'Reilly T (2005) What is web 2.0: design patterns and business models for the next generation of software. Retrieved on 22 Sept 2017, from https://tinyurl.com/p8z24dw

OpenStreetMap (2017) Za'atari refugee camp OSM object. Retrieved on 10 Mar 2018, from https://tinyurl.com/y8l9lflo

OpenStreetMap Wiki (2017) Refugee camp mapping. Retrieved on 3 Sept 2017, from https://tinyurl.com/ydavj4r7

Palen L, Vieweg S, Sutton J, Liu SB, Hughes AL (2007) Crisis informatics: studying crisis in a networked world. In: Proceedings of the third international conference on e-social science, Ann Arbor, MI. https://tinyurl.com/ydgl2zoh

Ram A (2015) Smartphones being solace and aid to desperate refugees. Wired. Retrieved on 16 Mar 2017, from https://tinyurl.com/hpakt82

Rao M (2015) A digital survival guide for the modern refugee: a compendium of helpful apps for migrants on the move. Huffington Post: The World Post. Retrieved on 24 July 2017, from https://tinyurl.com/ybuja5d7

Rehrl K, Gröchenig S (2016) A framework for data-centric analysis of mapping activity in the context of volunteered geographic information. ISPRS Int J Geo Inf 5(3):37. https://doi.org/10.3390/ijgi5030037

ReliefWeb (2017) Syrian Arab republic dashboard. Retrieved on 24 July 2017, from https://tinyurl.com/y8buo589

Sampson R, Gifford SM (2010) Place-making, settlement and well-being: the therapeutic landscapes of recently arrived youth with refugee backgrounds. Health Place 16(1):116–131

Schmeidl S (1997) Exploring the causes of forced migration: a pooled time-series analysis, 1971–1990. Soc Sci Q 78(2):284–308

Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman ME, Ungar LH (2013) Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS ONE 8(9):e73791

Sebti B (2016) 4 smartphone tools Syrian refugees use to arrive in Europe safely. Non-Government Organization. Retrieved on 16 Mar 2017, from https://tinyurl.com/jp3ks62

See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, Fritz S, Kerle N, Jiang B, Laakso M, Liu HY (2016) Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int J Geo Inf 5(5):55. https://doi.org/10.3390/ijgi5050055

Seltzer EK, Jean NS, Kramer-Golinkoff E, Asch DA, Merchant RM (2015) The content of social media's shared images about Ebola: a retrospective study. Public Health 129(9):1273–1277

Shacknove AE (1985) Who is a refugee? Ethics 95(2):274–284

Shelton T, Poorthuis A, Graham M, Zook M (2014) Mapping the data shadows of hurricane sandy: uncovering the sociospatial dimensions of 'big data'. Geoforum 52:167–179

Sloan L, Morgan J, Burnap P, Williams M (2015) Who Tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. PLoS ONE 10(3):e0115545

Stefanidis A, Cotnoir A, Croitoru A, Crooks A, Rice M, Radzikowski J (2013a) Demarcating new boundaries: mapping virtual polycentric communities through social media content. Cartogr Geogr Inf Sci 40(2):116–129

Stefanidis A, Crooks A, Radzikowski J (2013b) Harvesting ambient geospatial information from social media feeds. GeoJournal 78(2):319–338

Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, Pfoser D, Croitoru A, Crooks AT (2017) Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 3(2):e22

Stillwell J (1978) Interzonal migration: some historical tests of spatial-interaction models. Environ Plan A 10(10):1187–1200
Stouffer SS (1940) Intervening opportunities: a theory relating mobility and distance. Am Sociol Rev 5(6):845–867
The Migrants' Files (2014) Counting the dead. Retrieved on 16 Mar 2017, from https://tinyurl.com/z72f6q5
The World Bank (2015) The Kurdistan region of Iraq: assessing the economic and social impact of the Syrian conflict and ISIS. World Bank, Washington. Retrieved on 18 May 2018, from https://tinyurl.com/yccokqu7
The World Bank (2016) The welfare of Syrian refugees: evidence from Jordan and Lebanon. World Bank, Washington. Retrieved on 18 May 2018 from https://tinyurl.com/y9v5loub
The World Bank (2017a) DataBank home. Retrieved on 24 July 2017, from https://tinyurl.com/cq9guf3
The World Bank (2017b) Forcibly displaced: toward a development approach supporting refugees, the internally displaced, and their hosts. World Bank, Washington. https://doi.org/10.1596/978-1-4648-0938-5. Retrieved on 25 Sept 2017 from https://tinyurl.com/y9tnoojf
Tomaszewski B (2017) GIS for refugees, by refugees. ArcNews 39(3):4–5
United Nations Conference of Plenipotentiaries on the Status of Refugees and Stateless Persons (1951) Final act and convention relating to the status of refugees. United Nations, New York
United Nations General Assembly (2015) Transforming our world: the 2030 agenda for sustainable development. United Nations (UN) resolution A/RES/70/1. Retrieved on 25 Sept 2017, from https://tinyurl.com/od9mens
United Nations High Commissioner for Refugees (UNHCR) (2013) UNHCR statistical online population database: sources, methods and data considerations. Retrieved on 26 Mar 2018, from https://tinyurl.com/ycyzmabd
United Nations High Commissioner for Refugees (UNHCR) (2016) Global trends: forced displacement in 2015. United Nations High Commissioner for Refugees (UNHCR). Retrieved on 16 Mar 2017, from https://tinyurl.com/h2uw8h5
United Nations High Commissioner for Refugees (UNHCR) (2017a) UNHCR statistics: the world in numbers. Retrieved on 16 Mar 2017, from https://tinyurl.com/hks95am
United Nations High Commissioner for Refugees (UNHCR) (2017b) Operational portal refugee situations: Syria regional refugee response. Retrieved on 2 Oct 2017, from https://tinyurl.com/ycd3aq96
Ushahidi (2017) Read the crowd. Retrieved on 22 Sept 2017, from https://tinyurl.com/y8ydcuc3
Valtonen K (2004) From the margin to the mainstream: conceptualizing refugee settlement processes. J Refug Stud 17(1):70–96
Van Deursen AJ, Van Dijk JA (2014) The digital divide shifts to differences in usage. New Med Soc 16(3):507–526
Wall M, Campbell O, Janbek D (2017) Syrian refugees and information precarity. New Med Soc 19(2):240–254
Watson I, Nagel C, Bilginsoy Z (2015) 'Facebook refugees' chart escape from Syria on cell phones. CNN Today. Retrieved on 2 Oct 2017, from https://tinyurl.com/y9u5nk2m
Weiss-Meyer A (2017) Apps for refugees: how technology helps in a humanitarian crisis. The Atlantic, May. Retrieved on 8 Aug 2017, from https://tinyurl.com/y9nzzwsv
Widener M, Horner M, Metcalf S (2013) Simulating the effects of social networks on a population's hurricane evacuation participation. J Geogr Syst 15(2):193–209
Yang A, Fan H, Jing N (2016) Amateur or professional: assessing the expertise of major contributors in OpenStreetMap based on contributing behaviors. ISPRS Int J Geo Inf 5(2):21
Zhong C, Arisona SM, Huang X, Batty M, Schmitt G (2014) Detecting the dynamics of urban structure through spatial network analysis. Int J Geogr Inf Sci 28(11):2178–2199
Zook M, Graham M, Shelton T, Gorman S (2010) Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Med Health Policy 2(2):6–32
Zottarelli LK (1998) Determinants of refugee production: an exploratory analysis. Thesis, University of North Texas, Denton, Texas. Retrieved on 16 Mar 2017, from https://tinyurl.com/ht7jb7l

Affiliations

Troy Curry[1] · Arie Croitoru[1,2] · Andrew Crooks[1,2,3] · Anthony Stefanidis[1,2,4]

Troy Curry
tcurry4@gmu.edu

Andrew Crooks
acrooks2@gmu.edu

Anthony Stefanidis
astefani@gmu.edu

[1] Department of Geography and GeoInformation Science, George Mason University, Fairfax, VA, USA

[2] Center for Geospatial Intelligence, George Mason University, Fairfax, VA, USA

[3] Department of Computational and Data Sciences, George Mason University, Fairfax, VA, USA

[4] Criminal Investigations and Network Analysis (CINA) Center, George Mason University, Fairfax, VA, USA