RESEARCH ARTICLE

Geo-Located Tweets. Enhancing Mobility Maps and Capturing Cross-Border Movement

Justine I. Blanford[1]*, **Zhuojie Huang**[1,2], **Alexander Savelyev**[1], **Alan M. MacEachren**[1]

1 Department of Geography, GeoVISTA Center, Penn State University, 320 Walker, University Park, Pennsylvania, 16802, United States of America, **2** Centre for Infectious Disease Dynamics, Penn State University, Millenium Science Complex, University Park, Pennsylvania, 16802, United States of America

* jib18@psu.edu

Abstract

Capturing human movement patterns across political borders is difficult and this difficulty highlights the need to investigate alternative data streams. With the advent of smart phones and the ability to attach accurate coordinates to Twitter messages, users leave a geographic digital footprint of their movement when posting tweets. In this study we analyzed 10 months of geo-located tweets for Kenya and were able to capture movement of people at different temporal (daily to periodic) and spatial (local, national to international) scales. We were also able to capture both long and short distances travelled, highlighting regional connections and cross-border movement between Kenya and the surrounding countries. The findings from this study has broad implications for studying movement patterns and mapping inter/intra-region movement dynamics.

Introduction

Despite the importance of capturing human mobility, reliable data on movement is often scarce, imprecise, and out of date. Therefore efforts needed to aid in the planning and allocation of resources are often hampered by lack of accurate and verifiable mobility data, making it difficult to identify movement trends and patterns [1]. Many data streams have been used to capture aspects of human movement, and each has limits of currency, spatial and temporal precision, accuracy, geographic coverage, and/or representativeness of population of interest. Census data, for example, is widely available and commonly used for analyzing long-term net movement flows at broad spatial scales [2]. But census data are usually aggregated to political districts (thus are imprecise) and collected on a decennial cycle (thus are often out of date) [3]. Spatially referenced data from cell phones [4–6], circulation of the US dollar bill [7–9], air traffic data [10, 11], sequential night-time imagery [12], transportation data [13] and news articles [14] have also been used to estimate movement and serve as proxies for capturing mobility at various temporal scales. Airline traffic data, although good at showing connectivity between geographic locations and potential net flows, are restrictive since they only capture air transportation and long-range spatial patterns between specific origin-destination locations by individuals who can afford to fly. Data about circulation of US dollar bills provides mobility

patterns across different scales but, currency movement is only an indirect proxy for human movement and the available data reflect only a small number of volunteer contributors [8, 9]. Night-time imagery is useful for capturing periodic movement such as seasonal variation [12] as are news articles through the reporting of an event [14]. Although cell phone call record data are useful for capturing population distribution [15] and fine scale spatial movement [16], the analysis of these data are usually aggregated to the nearest cell phone tower, can be difficult to obtain and are restricted in their use and reuse due to both proprietary and privacy issues [17]. Furthermore due to cell phone networks and coverage, these data are generally restricted to country-level, making cross-border human movement difficult to capture. Therefore, more accessible data through which human movement can be analyzed across different spatial and temporal scales and through different transport systems is needed [9].

With the arrival of smartphones, the ability to capture accurate locational information through a variety of applications (commonly referred to as 'apps') is now more common than ever because the smartphones have built-in GPS devices. Many social media apps (Twitter, Foursquare and Facebook are three common examples) have geo-location features that (optionally) includes the ability to attach locational information in the form of coordinates provided by the units GPS or place names provided by the user, thereby enabling individual users to leave a digital 'geographic footprint' of their movement when posting a message. This 'geographical footprint' is similar to that collected when a cell phone call or text message is sent. The major difference is that, unlike cell phone data, these records typically have precise GPS-generated locations rather than location to the nearest cell tower and can be freely collected independently of the software or service provider that generated the location data. Thus, the social activity with GPS coordinates can be aggregated to and summarized by various spatial/political units for different analytical purpose. Of the social media data, Twitter, through their geo-located tweets, has shown great potential for capturing human movement [18–21] across various political/regional boundary and is the focus of the research reported here.

Twitter has over 230 million active users [22] and its use is not confined to high income countries [20]. Twitter posts (tweets) are a maximum of 140 characters in length, but in spite of the brevity they carry a huge volume of publically available data within the average of 500 million tweets/day [22]. This data stream is attracting the attention of the scientific community as a source of information that may provide situational awareness about events (e.g. Sense-place2 [23]; ScatterBlogs [24]) as well as possible insights into society [25, 26]. Since 2010, Twitter provided users with the ability to include their location either by attaching coordinates or a place name [27] while tweeting, therefore making it possible to locate tweets geographically [18] both in space and over time (see [28]). The publicly accessible Streaming API provides around 90.1% coverage of the total geo-tagged tweets [29] and our own analysis of a 209 million tweet sample indicates that 91% of the tweets with geo-tagging are tagged with GPS-derived coordinates and 9% with place names (geolocated to common coordinates for that place). Geo-referenced tweets ordered in time by individual represent semi-continuous movement for that individual and because of the volume of tweets that are often sent, they can provide key insights into human movement patterns. Here we explored the utility of Twitter in capturing human movement in Kenya, where 62% of the population has access to a mobile phone (based on 2009 Census) [30]. The focus of our analysis was to understand movement patterns and connectivity between regions both nationally and internationally.

Table 1. Summary of Twitter data used in this analysis that was collected for Kenya between June 2013 and March 2014 ($N_{unique\ users}$ = 28,335; N_{tweets} = 720,149).

Year	Date	No Unique Users	Total No Tweets per Month (Kenya)	Total No Tweets per Month (Kenya + surrounding countries)
2013	1 Jun– 30 Jun	6,405	88,377	135,845
	1Jul– 17 Jul	3,285	32,651	47,721
	3 Aug– 8 Aug	2,413	17,394	26,457
	1 Sep– 30 Sep	5,530	59,182	91,802
	1 Oct– 31 Oct	7,046	92,794	143,714
	1 Nov– 30 Nov	7,188	90,302	143,452
	1 Dec– 31 Dec	8,197	115,264	187,504
2014	1 Jan– 31 Jan	7,635	113,194	183,942
	1 Feb– 28 Feb	4,982	63,402	103,484
	1 Mar– 23 Mar	3,928	47,589	73,679

doi:10.1371/journal.pone.0129202.t001

Results

Dataset summary

Geo-located tweets were captured over 10 months (June 2013 to March 2014) (Table 1) using the Twitter Streaming API version 1.1 (https://dev.twitter.com/docs/streaming-apis/streams/public) and saved to a text file in JSON format using a node.js application (http://nodejs.org/). To minimize erroneous movement and mobility results, we cleaned the data removed tweets that were related to web advertising (e.g. UNjobs_), traffic updates (e.g. NSC_MombasaRd) and internet bots (e.g. MarsBots). We further removed obvious errors in the data (e.g., users with apparent movement at speeds greater than 1000km/hr) [19]. A total of 720,149 tweets were captured for Kenya resulting in an average number of 4,931 tweets collected per day. Since we were only interested in movement we extracted latitude, longitude, date, time and the unique user id for each tweet. A total of 28,332 unique users,representing an estimated 1.2% of the population who have access to a mobile phone (N = 2.38 million [30]), were identified in Kenya (see Table 1 for summary) and mapped using ArcGIS 10.2 ESRI (Environmental Systems Research Institute). This percentage is very close to that found in a previous study [20]. Tweets were distributed throughout Kenya (Fig 1A) with the highest number of tweets captured in South Central and Western Kenya. Temporally at a daily scale, two peaks in the number of Tweets sent per day were apparent one, at 11am and the second at 9pm (Fig 1B). For the 10 months of data that were analyzed, tweets were highest during December and January and lowest in February (Table 1).

Mobility and connectivity

We analyzed local and regional movement to identify how places are connected and demonstrate the utility of social media data for mapping connectivity. Each person creates a unique digital footprint with each message they send, therefore we ascertained movement patterns by connecting individual points by user. We constructed a series of temporal movement patterns. These included (i) daily movement patterns by connecting tweets for each user within a 24 hour time period; (ii) monthly movement patterns by joining tweets for each user for each month and (iii) total movement patterns by linking all tweets for each user. Movement tracks were created by connecting the tweet by date and time for each unique user. ESRI ArcGIS 10.2
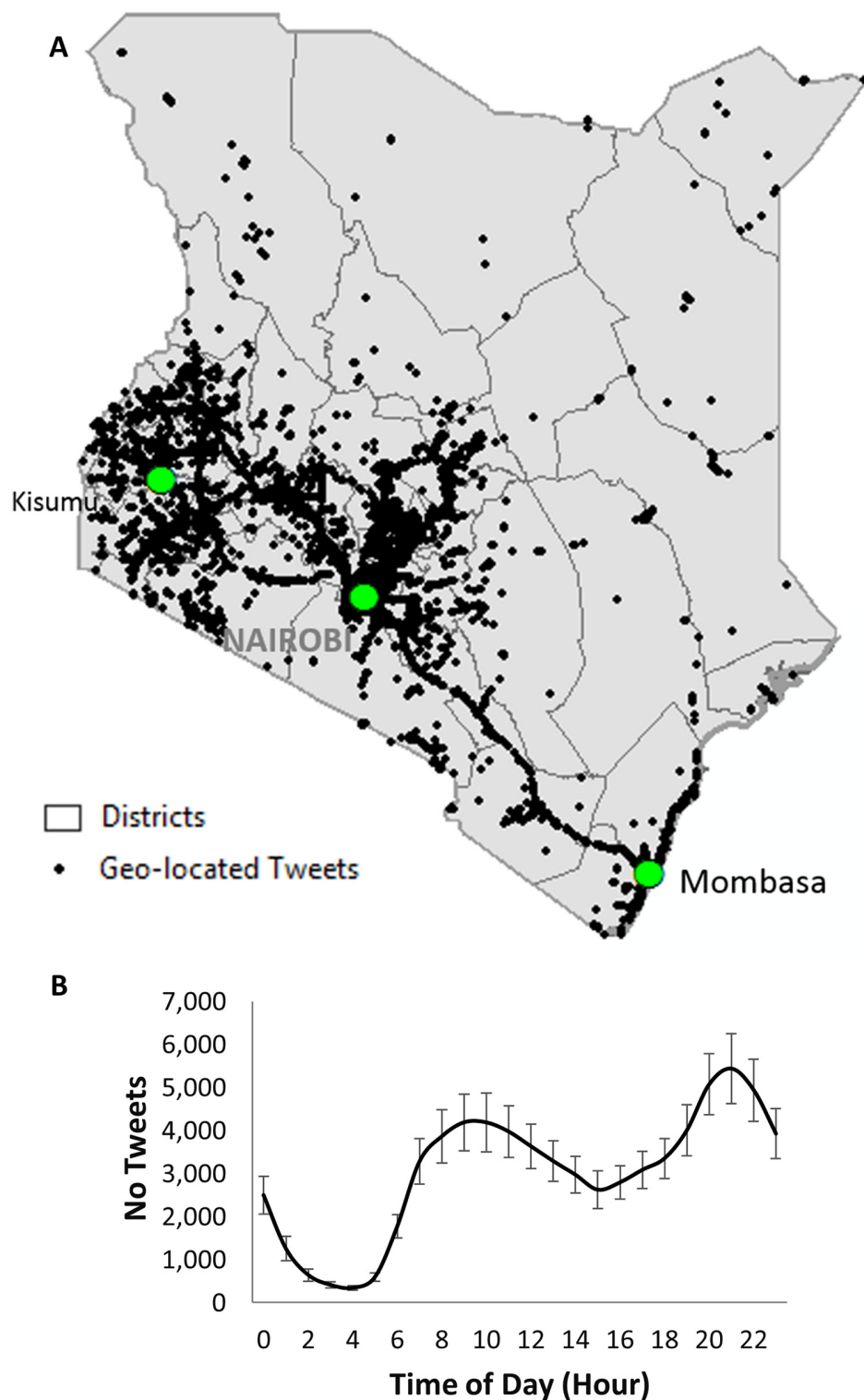
Fig 1. **(A)** Map showing distribution of tweets captured between June 2013 and March 2014. (B) Frequency of tweets per hour (+/- SE) local time. Map created using ArcGIS 10.2.

was used to create movement tracks between each of the locations and calculate the distance between each tweet and the total distance travelled by each user.

Movement patterns captured at different temporal scales highlight changes in connectivity across Kenya (Fig 2). On average 427 movement tracks were recorded each day and much of this movement occurred over short distances (< 5km, Fig 3A). Estimates of the users' spatial spread of movement based on analyzing their radius of gyration also indicated that the majority of users' travelled mostly within a relatively small area (<5km, Fig 3B). As the temporal scale was expanded more long distance movement events were revealed, both within country and across national boundaries (Figs 2 and 3B).

Net regional and international movement flow

Net regional flows of movement were constructed by capturing flow networks at administrative levels, where the administrative unit represents the source or destination nodes in each network. In this case we used centroids for district level boundaries as nodes in the network. Movement was defined as the movement of a person from one district to another. The total numbers of connections between districts were determined and used to investigate centrality of each district. We calculated four measures of centrality that included: degree (the number of ties to the district), betweenness (the number of times a district acts as a bridge between the shortest distance of two other districts), closeness (the inverse of the farness = the total distance of a district to all districts), and eigenvector centralities (the level of influence of a district by measuring the importance of a node based on the node's connections) (Table 2). Fig 4 shows the districts with the highest connectivity. Matching the cell-phone based findings by [4], we also found that the capital, Nairobi, was a central hub in the travel network with movement flowing in and out of the capital to different parts of Kenya (Fig 4). Movement was highest between Nairobi and the neighboring districts Kiambu, Machakos and Kajiado in the Central Region. Additional strong connections were also identified between Nairobi and the coastal area of Mombasa and the districts in the west, Nakuru, Uasin Gishu and Kisumu (Table 2).

Cross-border movement and potential flow of parasite

A limitation of utilizing mobile phone data is that they do not capture cross-border movement since these data are restricted to the phone providers and coverage may not extend beyond country boundaries unless subscribers have the necessary roaming capabilities enabled. With Twitter, in contrast, unique user identifiers remain consistent, therefore users of social media sites can continue to log in to their accounts worldwide, provided they have access to the internet, and provide a way of capturing cross-border movement. We explored the ability to map cross-border movement by analyzing an additional 417,451 tweets (Table 1, Fig 5A). Approximately, three percent of the users who tweeted from Kenya also tweeted from the surrounding countries such as Uganda, Tanzania, Somalia and South Sudan (Fig 5A). Analysis of the movements recorded via Twitter found that eighty-five percent of these users only travelled between two countries, 11% between three countries, 2% between four countries and only two people visited five countries. Net flow of movement between countries is calculated using the centroids of each country as nodes in the network. Movement was defined as the movement of a person from one country to another and the total number of connections between each country was calculated. The greatest number of connections from and to Kenya were with Uganda and Tanzania (Fig 5B, darker and bolder the line the higher the number of connections). Since user's movements are highly localized (Fig 3A), we further highlighted variations in long distance travel connections (green) versus more local connections that occur during shorter (brown) and intermediate (yellow) travel journeys (Fig 5C).
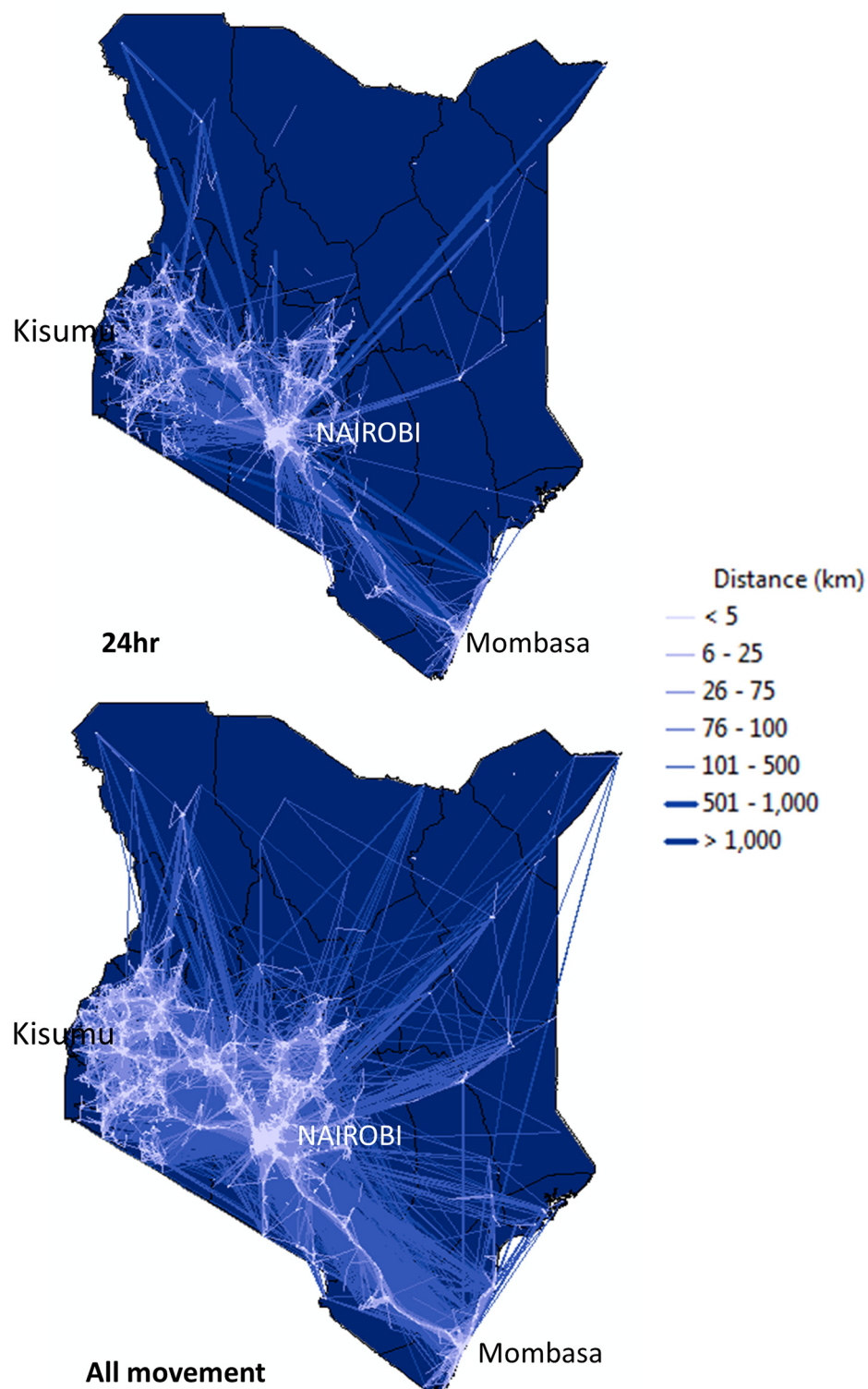
Fig 2. Movement patterns captured at different temporal scales illustrate connectivity between districts in Kenya within a 24-hour time period (N = 90,645 tracks) and during a ten month time period (N = 17,900). Each line represents a movement segment. The long distance tracks indicates population movements by plane or by train within the country. Maps created using ArcGIS 10.2.
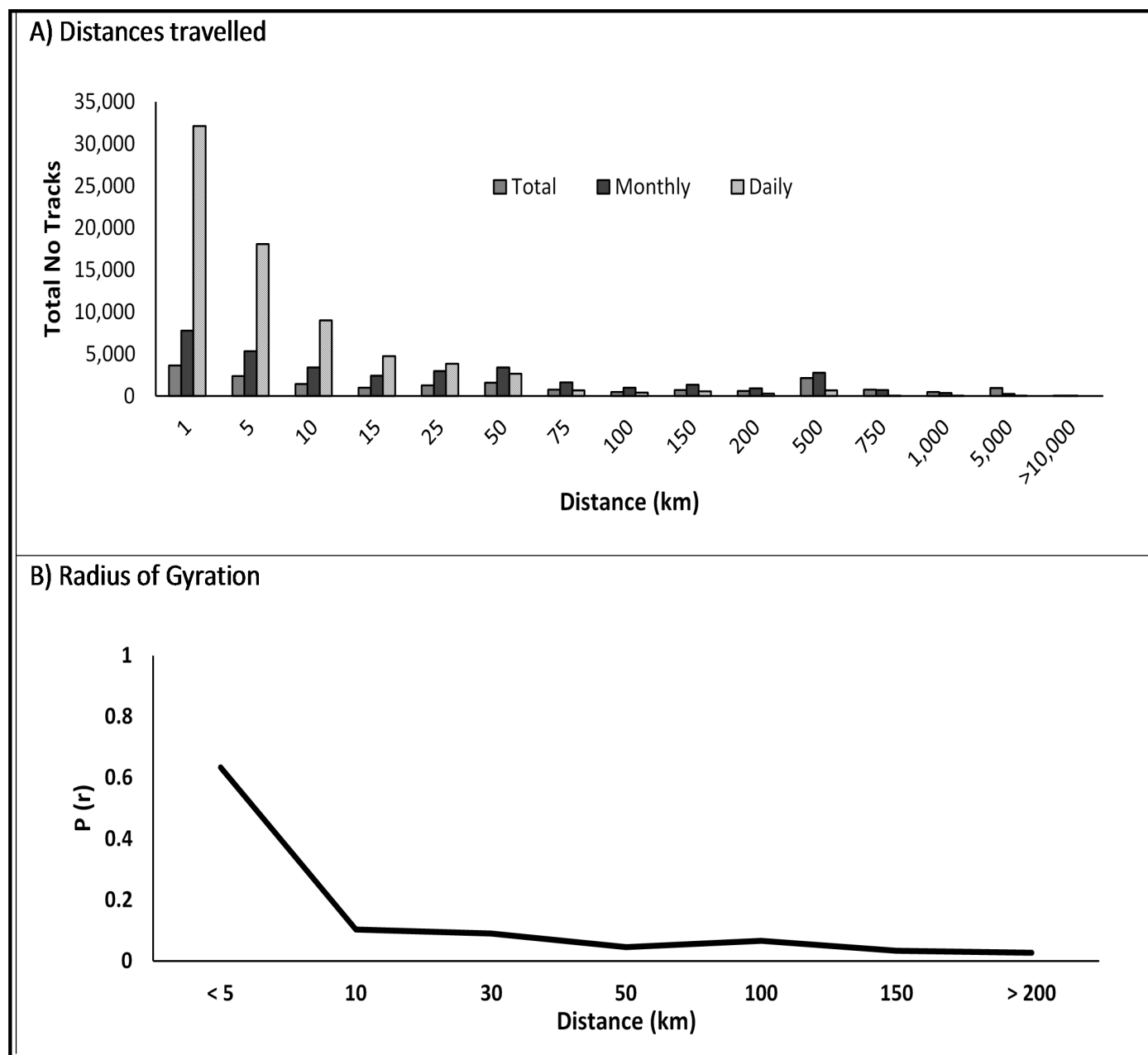
doi:10.1371/journal.pone.0129202.g002

Fig 3. Connectivity between locations and human movement within Kenya. (A) Distances travelled daily, monthly and in total by each user and (B) the proportion of user's radius of gyration (solid line).

doi:10.1371/journal.pone.0129202.g003

Discussion

In this study we used geo-located Twitter data to determine movement flows in Kenya and the surrounding countries. Although Twitter data, like cell phone data, is limited due to the number of users tweeting and biases inherent in who has access to communication technology, these data can serve as a proxy for mobility [18–21] particularly in regions where data is scarce as shown here. Having perfect data is rare [31], yet in many cases the "good enough" principle in making decisions is sufficient [32]. Thus in absence of better data, Twitter may be

Table 2. Centrality values between districts in Kenya.

| District Name	Degree	Betweenness	Closeness	Eigenvector
BARINGO	33	6.210	0.013	0.465
BOMET	33	2.908	0.013	0.469
BUNGOMA	40	11.894	0.014	0.539
BUSIA	33	3.581	0.013	0.467
E. MARAKWET	21	1.482	0.012	0.288
EMBU	34	8.217	0.013	0.450
GARISSA	29	13.653	0.013	0.347
HOMA_BAY	37	4.499	0.014	0.519
ISIOLO	28	7.912	0.013	0.327
KAJIADO	79	125.993	0.018	0.905
KAKAMEGA	49	14.293	0.015	0.653
KERICHO	43	18.557	0.013	0.569
KIAMBU	76	105.624	0.018	0.890
KILIFI	44	25.793	0.014	0.562
KIRINYAGA	36	6.346	0.013	0.497
KISII	44	10.389	0.014	0.606
KISUMU	59	39.176	0.016	0.735
KITUI	36	10.264	0.014	0.475
KWALE	36	7.276	0.013	0.491
LAIKIPIA	44	15.822	0.014	0.580
LAMU	13	0.148	0.011	0.174
MACHAKOS	72	74.364	0.018	0.866
MAKUENI	36	7.405	0.013	0.495
MANDERA	10	0.620	0.011	0.120
MARSABIT	10	0.250	0.011	0.133
MERU	47	23.142	0.014	0.590
MIGORI	36	7.067	0.013	0.490
MOMBASA	77	127.112	0.018	0.881
MURANGA	42	14.809	0.014	0.542
NAIROBI	94	266.909	0.022	1.000
NAKURU	77	98.524	0.018	0.908
NANDI	37	13.789	0.013	0.467
NAROK	59	44.211	0.016	0.742
NITHI	23	0.761	0.012	0.327
NYAMIRA	32	2.058	0.013	0.469
NYANDARUA	24	3.025	0.012	0.335
NYERI	49	19.072	0.015	0.628
SAMBURU	20	1.401	0.012	0.278
SIAYA	30	1.800	0.013	0.428
TAITA TAVETA	35	7.934	0.013	0.478
TANA RIVER	15	0.418	0.012	0.198
TRANS-NZOIA	38	8.474	0.014	0.522
TURKANA	21	2.106	0.012	0.291
UASIN GISHU	75	90.420	0.018	0.881
VIHIGA	42	7.944	0.014	0.585
WAJIR	15	1.760	0.012	0.197
WEST POKOT	21	1.588	0.012	0.299
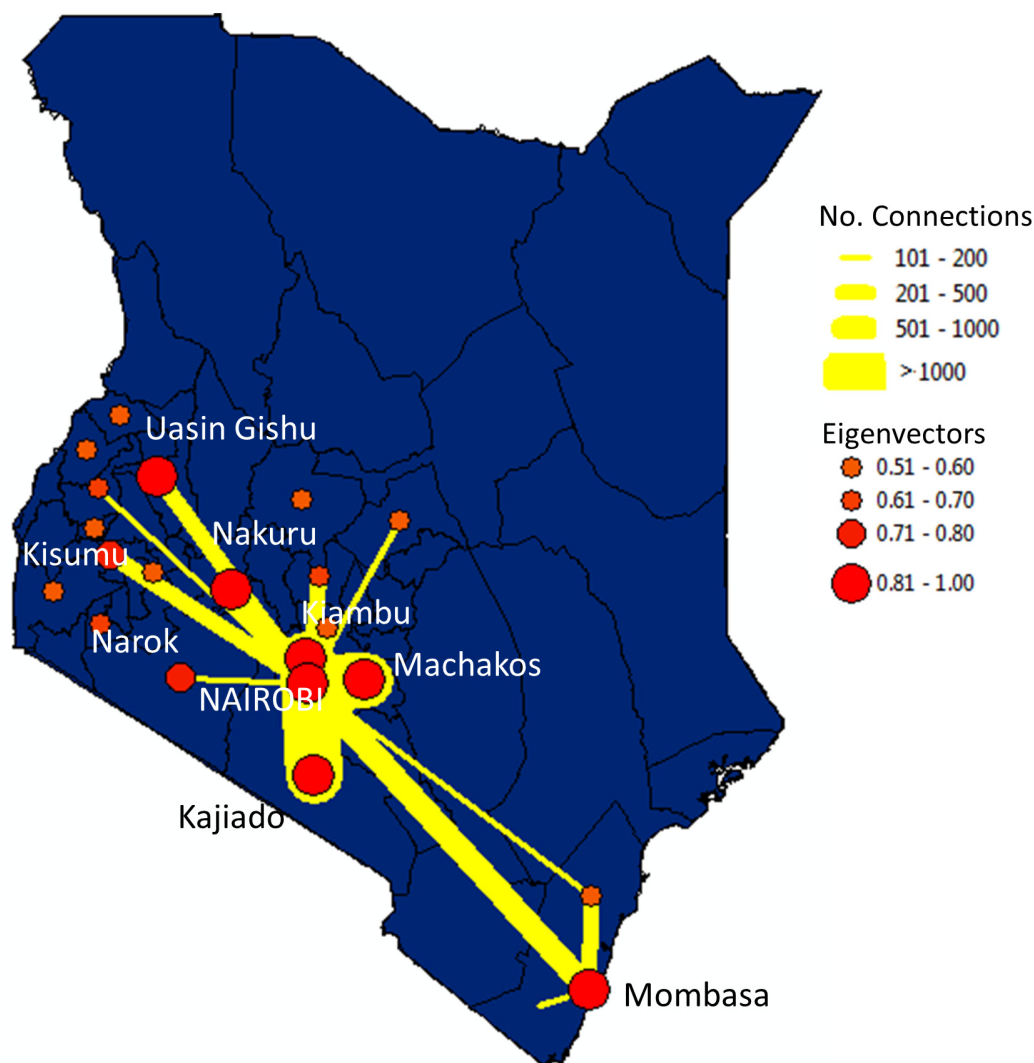
Fig 4. Map illustrating the connectivity between districts and Eigenvector value illustrating the level of influence of each district. For clarity we have only illustrated linkages with more than 100 connections and Eigenvectors greater than 0.50. Map created using ArcGIS 10.2.

considered a "good enough" data set for constructing human mobility networks and these data are likely to be more comprehensive than can be derived from the decennial census data, one-off surveys, or night lights methods.

Twitter has several advantages over mobile phone data records for tracking population movements: geo-located tweets provide greater locational accuracy than cell phone data since coordinate information is collected using built-in GPS receivers whereas cell phone data is linked to the nearest tower [18]. While this may not affect our ability to analyze large scale movement of people, the accuracy of small scale movement may be misrepresented when solely relying on cell phone data. Furthermore, cell phone data typically stops at country borders but Twitter data does not. A limitation of the study by [4] is that mobility was only quantified for Kenya. We have extended this study by capturing cross-border movement (Fig 5) and highlighted key connections between Kenya and surrounding countries (Fig 5B). For example, based on the number of connections shown in (Fig 5B), movement between Kenya and
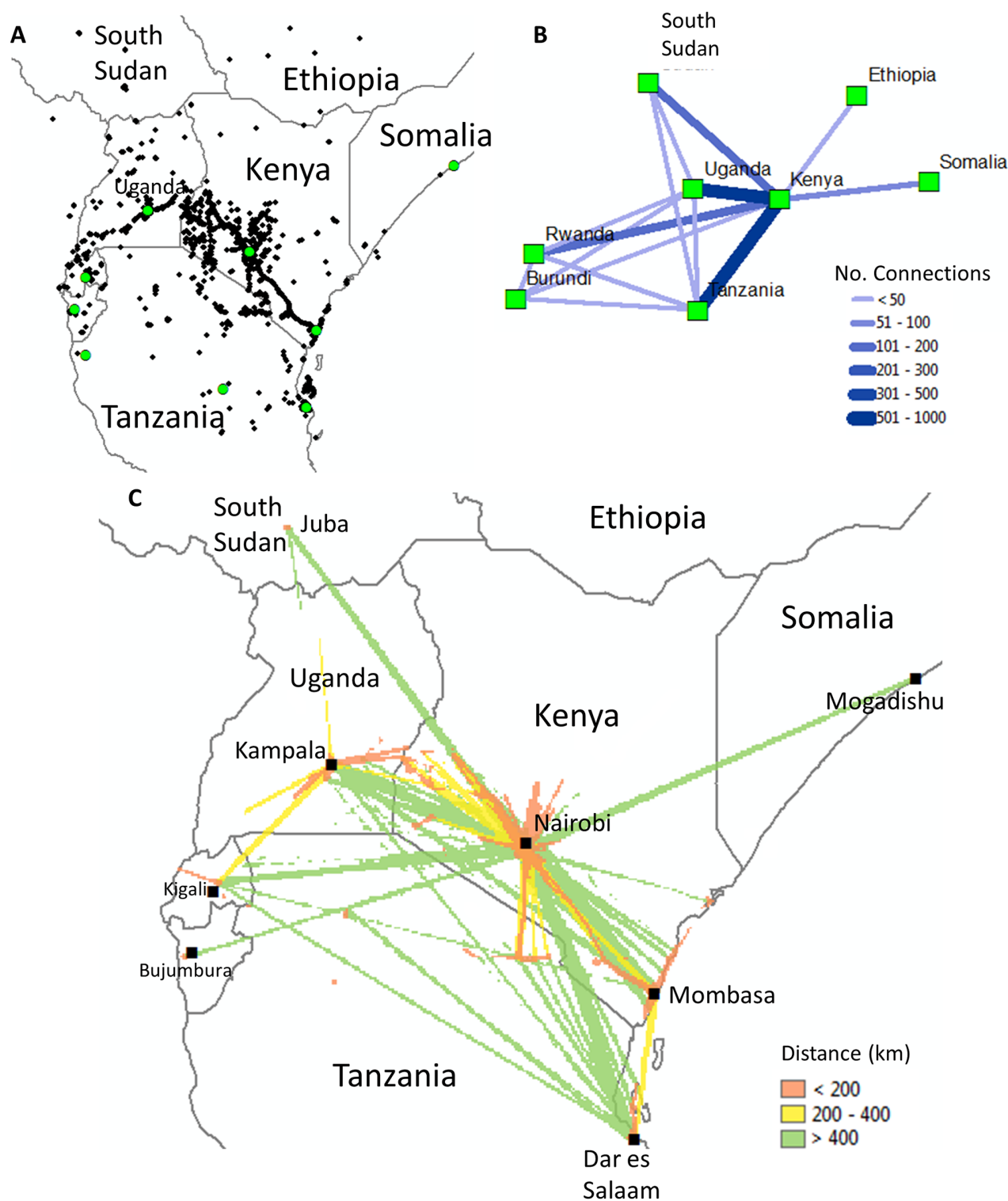
Fig 5. Maps illustrating (A) the distribution of geo-located tweets in the study area for users who crossed-borders (N = 770) **(B)** connections between Kenya and the surrounding countries and **(C)** a flow map showing the connectivity between different geographic locations by travel distance. Map created using ArcGIS 10.2.

doi:10.1371/journal.pone.0129202.g005

Tanzania is well established. By breaking this down further, we were able to distinguish travel patterns between specific geographic locations such as direct connections between Nairobi and Dar es Salaam and trips between Nairobi and Arusha (Figs 3 and 5C). Not only is this important for examining cross-border movement and importation of disease [33] but also for identifying disease epicenters by determining regional travel hubs that people pass through. For example, in light of the current Ebola outbreak, identifying major transport hubs, such as Nairobi in Kenya and their specific connections to other small as well as large locations at different times is important for determining places that may be at high risk to the virus [34]. Finally, the context of the tweets can reflect a population's collected social behavior [35], which can be utilized to understand the spatial-temporal dynamics of peer influence and social contagion of behavioral spread [36].

Movement can take place for many reasons ranging from work and economic well-being [1] through conflict [37] to displacement caused by loss of livelihoods due to natural hazards (e.g. climate- and weather-related events [38] such as flooding [39, 40], drought [41, 42],[14] and heat stress [43]) at different temporal (e.g. daily to migration) and spatial scales (neighborhood, city-to-city, to international). Twitter has the potential to capture human movement across these different spatial and temporal scales as highlighted by the red box in (Fig 6). Also, with continuous growth and wide-spread user distribution, geotagged tweets show great potential to serve as a complement to cell phone based data [15] for population distribution estimation [44].

Twitter is free currently, easily accessible and can provide on-the-ground information at local, national and international levels. The key contributions of the study presented here are demonstrating the usefulness of Twitter for mapping human movement and, for Kenya, highlighting key movement corridors and convergence zones that may serve as social hotspots. Having the ability to pinpoint these areas will be useful for policy development and planning, particularly with limited public resources [45]. Although this study examined the utility of Twitter for delineating human movement patterns, the usefulness of this type of data could easily be extended to other applications including, but not limited to, improve disease modelling [9], understand spatial/temporal epidemic dynamics and disease movements [16], enhance disease surveillance [46], understanding of community structure [47, 48], delineating people's activity space [49], assessing access to healthcare or vaccination centers [50] and understanding people's movement response during an emergency (e.g. Hurricane Sandy [51, 52]). Therefore the work outlined here provides a lens for examining the complexities and heterogeneities of movement in both space and time that enables us to gain further insights into the motivations associated with mobility patterns in different settings such as data-rich and data-poor countries and urban vs rural settings.

Methods

Mobility measures

The mobility of an individual was measured using the radius of gyration trajectory, $r(g)$ where the location of an individual, i, at a location in space and time is given by the geolocation of that individual's tweet. Since we do not know an individual's home location, the center of mass, or centroid of all known user locations, was used to represent the user's usual location and was calculated as:

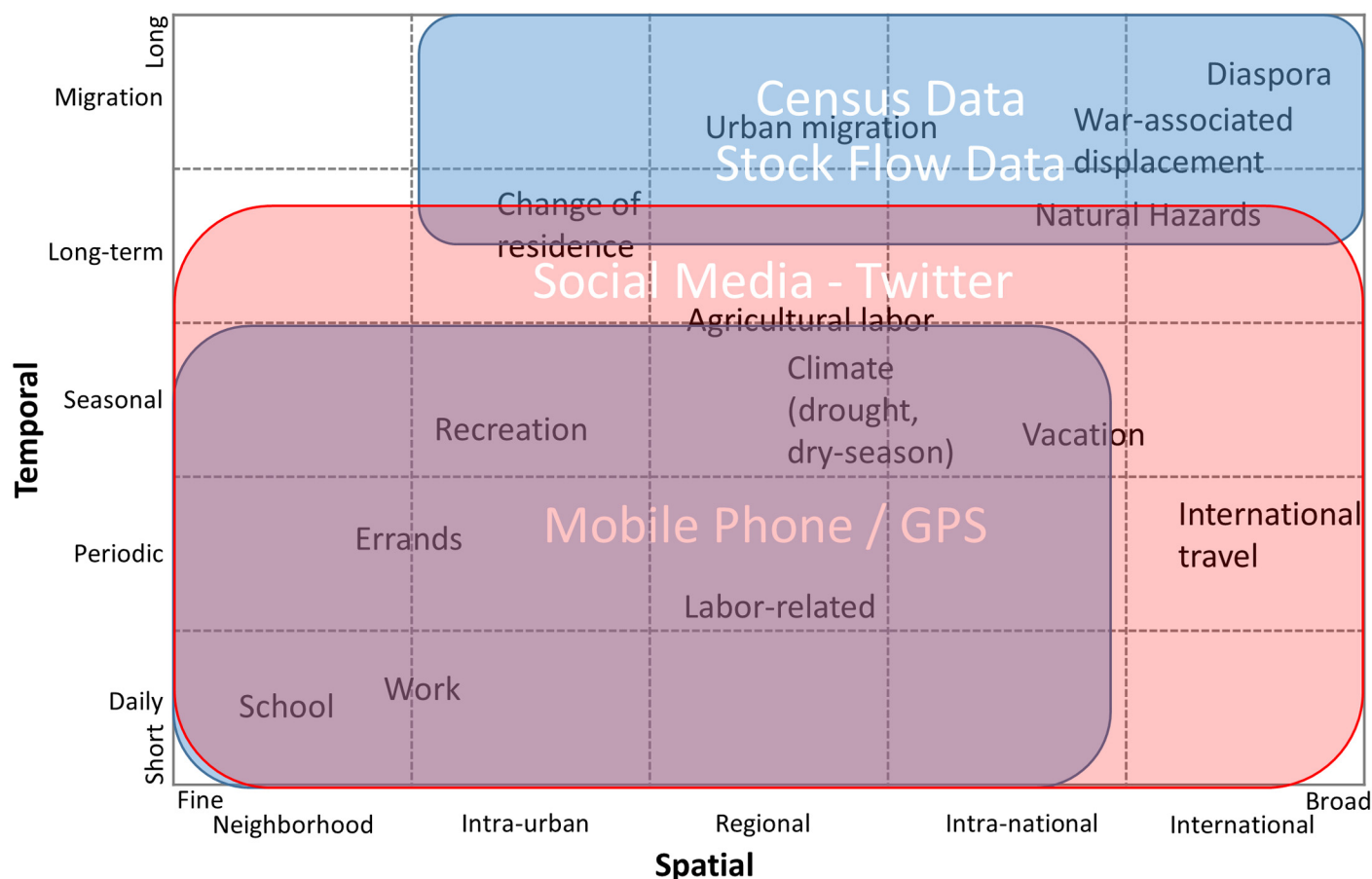$$m_{cm} = \frac{1}{n} \sum_{i=1}^{n} m_i \qquad (1)$$

Fig 6. A framework and associated data sources useful for capturing human mobility in time and space. Movements are characterized in terms of their spatial and temporal scale, which are defined in terms of physical displacement (*spatial*) and time spent (*temporal*, frequency and duration) (Source: adapted from [45], [57]).

doi:10.1371/journal.pone.0129202.g006

where m_i is the location of a geo-located tweet during the time period for which data was collected.

For each user the radius of gyration was calculated:

$$r(g) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|m_i - m_{cm}|^2}$$ (2)

where n is the number of locations from where a tweet was sent, m_i is the location of a tweet and m_{cm} is the center of mass of a trajectory [6].

Centrality calculations

Firstly, we calculated the degree of a node as the number of edges it is connected to. This is the first implication of how districts are connected by population movements. Afterward, we calculated the betweenness centrality, closeness centrality and eigenvector centrality. Betweenness centrality calculates the number of shortest paths going through a specific vertex in a network. In transport network analysis, it provides an approximating measure of the traffic handled by

the vertices, as well as an indicator of the importance of the vertices in a network [53]. The betweenness centrality is calculated as follows:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\delta_{st}(v)}{\delta_{st}} \tag{3}$$

In which δ_{st} is the total number of shortest paths from node s to node t and $\delta_{st}(v)$ is the number of those paths that pass through v. Under our population movement network context, the vertices betweenness centrality can be interpreted as the intensity that population movement going through vertices.

Closeness centrality measures the distance from a node to all other nodes in a network [54]. It is calculated based on the length of the average shortest path between a vertex and all vertices in the graph:

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1} \tag{4}$$

In which $d(n_i, n_j)$ is the distance between two nodes. In our population movement network setting, the closeness centrality can be interpreted as how close the districts are connected by population movement.

Eigenvector centrality measures relative importance of a node in a network [55, 56]. If a node has more connections to highly connected nodes, it has a relatively higher score than for one that has more connections to less connected nodes. The eigenvector centrality is calculated as follows:

$$x = \frac{1}{\lambda} Ax, \; x_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} x_j \tag{5}$$

in which A would be the adjacency matrix, λ is the largest eigenvalue and x is the corresponding eigenvector. Spatial clusters of eigenvector centrality can be observed on (Fig 4), which can be interpreted as high volume of population movements vertices are close to each other in geographic distance.

Author Contributions

Conceived and designed the experiments: JIB AMM. Performed the experiments: AS JIB. Analyzed the data: JIB ZH. Contributed reagents/materials/analysis tools: JIB ZH AS. Wrote the paper: JIB ZH AS AMM.

References

1. Koser K, Laczko F, Solomon MK. World Migration Report 2010. The future of migration: building capacities for change2010 Aug 10, 2014:[295 p.]. Available from: http://publications.iom.int/bookstore/free/WMR_2010_ENGLISH.pdf.

2. Dalziel BD, Pourbohloul B, Ellner SP. Human mobility patterns predict divergent epidemic dynamics among cities. Proceedings Biological sciences / The Royal Society. 2013; 280(1766):20130763. Epub 2013/07/19. doi: 10.1098/rspb.2013.0763 PMID: 23864593; PubMed Central PMCID: PMC3730584.

3. Massey DS, Zenteno RM. The dynamics of mass migration. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96(9):5328–35. PMID: 10220465; PubMed Central PMCID: PMC21863.

4. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science. 2012; 338(6104):267–70. Epub 2012/10/16. doi: 10.1126/science.1223467 PMID: 23066082; PubMed Central PMCID: PMC3675794.

5. Huang Z, Kumar U, Bodnar T, Salathe M. Understanding population displacements on location-based call records using road data. MOBIGIS '13; November 05–08; Orlando, Florida2013. p. 5.

6. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. Nature. 2008; 453(7196):779–82. Epub 2008/06/06. doi: 10.1038/nature06958 PMID: 18528393.

7. Thiemann C, Theis F, Grady D, Brune R, Brockmann D. The structure of borders in a small world. PloS one. 2010; 5(11):e15422. Epub 2010/12/03. doi: 10.1371/journal.pone.0015422 PMID: 21124970; PubMed Central PMCID: PMC2987795.

8. Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel. Nature. 2006; 439(7075):462–5. doi: 10.1038/nature04292 PMID: 16437114.

9. Brockmann D. Human Mobility and Spatial Disease Dynamics. 2009. In: Reviews of Nonlinear Dynamics and Complexity [Internet]. Weinheim: WILEY-VCH Verlag GmbH & Co.; [24].

10. Huang Z, Tatem AJ. Global malaria connectivity through air travel. Malaria journal. 2013; 12:269. doi: 10.1186/1475-2875-12-269 PMID: 23914776; PubMed Central PMCID: PMC3766274.

11. Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y. Air travel and vector-borne disease movement. Parasitology. 2012; 139(14):1816–30. doi: 10.1017/S0031182012000352 PMID: 22444826.

12. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. Science. 2011; 334(6061):1424–7. Epub 2011/12/14. doi: 10.1126/science.1210554 PMID: 22158822.

13. Balcan D, Colizza V, Goncalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(51):21484–9. doi: 10.1073/pnas.0906910106 ISI:000272994200006. PMID: 20018697

14. Tomaszewski B, Blanford JI, Ross K, Pezanowski S, MacEachren AM. Supporting Geographically-aware WebDocument Foraging and Sensemaking. Computers, Environment and Urban Systems. 2011;(35):192–207.

15. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, et al. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences. 2014:6.

16. Tatem AJ, Huang Z, Narib C, Kumar U, Kandula D, Pindolia DK, et al. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malaria journal. 2014; 13:52. doi: 10.1186/1475-2875-13-52 PMID: 24512144; PubMed Central PMCID: PMC3927223.

17. Wesolowski A, Buckee CO, Pindolia DK, Eagle N, Smith DL, Garcia AJ, et al. The use of census migration data to approximate human movement patterns across temporal scales. PloS one. 2013; 8(1): e52971. Epub 2013/01/18. doi: 10.1371/journal.pone.0052971 PMID: 23326367; PubMed Central PMCID: PMC3541275.

18. Frank MR, Mitchell L, Dodds PS, Danforth CM. Happiness and the patterns of life: a study of geolocated tweets. Sci Rep. 2013; 3:2625. Epub 2013/09/13. doi: 10.1038/srep02625 PMID: 24026340.

19. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopulos P, Ratti C. Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sc. 2014:12.

20. Leetaru K, Wang S, Cao G, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: The geography of Twitter. First Monday. 2013; 18(5–6):32.

21. Savelyev A, MacEachren AM. Exploratory visualization of movement patterns in Twitter data. North American Cartographic Information Society Annual Meeting (NACIS); Greenville, USA2013.

22. Twitter. About Twitter, Inc.2014 Jan 27 2014. Available from: https://about.twitter.com/company.

23. MacEachren AM, Jaiswal A, Robinson AC, Pezanowski S, Savelyev A, Mitra P, et al. SensePlace2: GeoTwitter analytics support for situational awareness IEEE Conference on Visual Analytics Science and Technology (VAST); 23–28 Oct; Providence, RI 2011. p. 181–90.

24. Bosch H, Thom D, Heimerl F, PUttmann E, Koch S, Kruger R, et al. ScatterBlogs2: Real-Time Monitoring of Microblog Messages through User-Guided Filtering. IEEE Transactions on Visualization and Computer Graphics 2013; 19:2022–31. doi: 10.1109/TVCG.2013.186 PMID: 24051768

25. Andrienko G, Andrienko N, Bosch H, Ertl T, Fuchs G, Jankowski P, et al. Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics. Computing in Science & Engineering. 2013:12.

26. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PloS one. 2013; 8(5):e64417. Epub 2013/06/05. doi: 10.1371/journal.pone.0064417 PMID: 23734200; PubMed Central PMCID: PMC3667195.

27. Twitter. FAQs about the Tweet location feature2014. Available: https://support.twitter.com/articles/78525-faqs-about-the-tweet-location-feature.

28. Blanford JI, Bernhardt J, Savelyev A, Wong-Parodi G, Carleton AM, Titley DW, et al., editors. Tweeting and Tornadoes. 11th International ISCRAM Conference (Short Paper); 2014 May; State College, USA: ISCRAM.

29. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. Seventh International AAAI Conference on Weblogs and Social Media; July 8–11; Cambridge, Massachusetts, USA: Association for the Advancement of Artificial Intelligence; 2013. p. 9.

30. IPUMS. Integrated Public Use Microdata Series, International: Version 6.3 [Machine-readable database]. In: Minnesota_Population_Center, editor. Minneapolis: University of Minnesota; 2014.

31. Palen L, Vieweg S, Anderson KM. Supporting "Everyday Analysts" in Safety- and Time-Critical Situations. The Information Society. 2010; 27(1):52–62.

32. Granger-Happ E. The Good Enough Principle—What we can learn about technology from the pragmatic solutions of nonprofits. Save the Children. 2008:1–25.

33. Malakooti MA, Biomndo K, Shanks GD. Reemergence of epidemic malaria in the highlands of western Kenya. Emerging infectious diseases. 1998; 4(4):671–6. doi: 10.3201/eid0404.980422 PMID: 9866748; PubMed Central PMCID: PMC2640260.

34. BBC. Ebola outbreak: Kenya at high risk, warns WHO. BBC News [Internet]. 2014. Available from: http://www.bbc.com/news/world-africa-28769678.

35. Lu X, Brelsford C. Network structure and community evolution on twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami. Sci Rep. 2014; 4:6773. doi: 10.1038/srep06773 PMID: 25346468; PubMed Central PMCID: PMC4209381.

36. Salathe M, Vu DQ, Khandelwal S, Hunter DR. The dynamics of health behavior sentiments on a large online social network. EPJ Data Science. 2013; 2(4):12.

37. Albuja S, Arnaud E, Beytrison F, Caterina M, Charron G, Fruehauf U, et al. Global Overview 2012: People internally displaced by conflict and violence2013 Aug 16, 2014. Available from: http://www.internal-displacement.org/publications/2013/global-overview-2012-people-internally-displaced-by-conflict-and-violence/.

38. Yonetani M. Global Estimates 2012: People displaced by disasters2013 Aug 10, 2014. Available from: http://www.internal-displacement.org/publications/2013/global-estimates-2012-people-displaced-by-disasters.

39. Bohra-Mishra P, Oppenheimer M, Hsiang SM. Nonlinear permanent migration response to climatic variations but minimal response to disasters. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(27):9780–5. doi: 10.1073/pnas.1317166111 PMID: 24958887.

40. Dun O. Migration and Displacement Triggered by Floods in the Mekong Delta. International Migration. 2011; 49:e200–e23. doi: 10.1111/j.1468-2435.2010.00646.x

41. Doevenspeck M. The Thin Line Between Choice and Flight: Environment and Migration in Rural Benin. International Migration. 2011; 49:e50–e68. doi: 10.1111/j.1468-2435.2010.00632.x PMID: 21306803

42. Gila OA, Zaratiegui AU, De Maturana Diéguez VL. Western Sahara: Migration, Exile and Environment. International Migration. 2011; 49:e146–e63. doi: 10.1111/j.1468-2435.2010.00665.x

43. Mueller V, Gray C, Kosec K. Heat stress increases long-term human migration in rural Pakistan. Nature Climate Change. 2014; 4:182–5. PMID: 25132865

44. Linard C, Patel N, Stevens FR, Huang Z, Gaughan AE, Tatem AJ. Use of active and passive VGI data for population distribution modelling: experience from the WorldPop project. GIScience 2014 workshop: Role of Volunteered Geographic Information in Advancing Science: Effective Utilization2014.

45. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis. 2009; 3(7): e481. Epub 2009/07/22. doi: 10.1371/journal.pntd.0000481 PMID: 19621090; PubMed Central PMCID: PMC2710008.

46. Salathe M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. The New England journal of medicine. 2013; 369(5):401–4. doi: 10.1056/NEJMp1307752 PMID: 23822655.

47. Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, et al. Redrawing the map of Great Britain from a network of human interactions. PloS one. 2010; 5(12):e14248. Epub 2010/12/21. doi: 10.1371/journal.pone.0014248 PMID: 21170390; PubMed Central PMCID: PMC2999538.

48. Austwick MZ, O'Brien O, Strano E, Viana M. The structure of spatial networks and communities in bicycle sharing systems. PloS one. 2013; 8(9):e74685. doi: 10.1371/journal.pone.0074685 PMID: 24040320; PubMed Central PMCID: PMC3765359.

49. Zenk SN, Schulz AJ, Matthews SA, Odoms-Young A, Wilbur J, Wegrzyn L, et al. Activity space environment and dietary and physical activity behaviors: a pilot study. Health Place. 2011; 17(5):1150–61. Epub 2011/06/24. doi: 10.1016/j.healthplace.2011.05.001 PMID: 21696995; PubMed Central PMCID: PMC3224849.

50. Blanford JI, Kumar S, Luo W, MacEachren AM. It's a long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger. Int J Health Geogr. 2012; 11:24. Epub 2012/06/29. doi: 10.1186/1476-072X-11-24 PMID: 22737990; PubMed Central PMCID: PMC3515413.

51. Sullivan KD, Uccellini LW. Service Assessment: Hurricane/Post-Tropical Cyclone Sandy, October 22–29, 20122013 Jan 3, 2014:[66 p.]. Available: http://www.nws.noaa.gov/os/assessments/pdfs/Sandy13.pdf.

52. Wang Q, Taylor JE. Quantifying human mobility perturbation and resilience in hurricane sandy. PloS one. 2014; 9(11):e112608. doi: 10.1371/journal.pone.0112608 PMID: 25409009; PubMed Central PMCID: PMC4237337.

53. Guimera R, Mossa S, Turtschi A, Amaral LA. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(22):7794–9. doi: 10.1073/pnas.0407994102 PMID: 15911778; PubMed Central PMCID: PMC1142352.

54. Newman MEJ. A measure of betweenness centrality based on random walks. Social Networks. 2005; 27:39–54.

55. Bonacich P. Factoring and weighting approaches to clique identification. Journal of Mathematical Sociology. 1972; 2:113–20.

56. Bonacich P. Some unique properties of eigenvector centrality. Social Networks. 2007; 29:555–64.

57. Tatem A. Mapping and modelling infection movements in low income regions using novel digital datasets. Challneges in modelling the spatial and temporal dimesions of the ecoology of infectious diseases workshop; Columbus, Ohio2012.