Social networking in developing regions

Azarias Reda
University of Michigan
azarias@umich.edu

Sam Shah
LinkedIn
samshah@linkedin.com

Mitul Tiwari
LinkedIn
mtiwari@linkedin.com

Anita Lillie
LinkedIn
alillie@linkedin.com

Brian Noble
University of Michigan
bnoble@umich.edu

ABSTRACT

Online social networks have enjoyed significant growth over the past several years. With improvements in mobile and Internet penetration, developing countries are participating in increasing numbers in online communities. This paper provides the first large scale and detailed analysis of social networking usage in developing country contexts. The analysis is based on data from LinkedIn, a professional social network with over 120 million members worldwide. LinkedIn has members from every country in the world, including millions in Africa, Asia, and South America. The goal of this paper is to provide researchers a detailed look at the growth, adoption, and other characteristics of social networking usage in developing countries compared to the developed world. To this end, we discuss several themes that illustrate different dimensions of social networking use, ranging from interconnectedness of members in geographic regions to the impact of local languages on social network participation.

Categories and Subject Descriptors

H.4 [**Information Systems**]: Information Systems Applications; J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Measurement, Human Factors

Keywords

Developing regions, social networks, emerging markets

1. INTRODUCTION

As mobile and Internet penetration improve worldwide [1], users from developing countries are participating in increasing numbers in online communities. Several studies of web usage patterns in developing country contexts [13, 15] indicate Internet users in these areas are very engaged in online social networking and communication tools, spending a significant portion of their online time on them. These observations have been made across several usage scenarios, ranging from educational institutions in urban India to remote Internet access sites in Africa and Latin America. In addition, surveys and other anecdotal evidence has indicated that users from emerging economies have been driving worldwide membership growth in social networks [6].

This paper provides the first large scale and detailed analysis of social networking usage in developing country contexts. Our analysis is based on profile and activity data from LinkedIn, a professional social networking site with, as of writing, over 120 million members worldwide. Over the past few years, online social networking has been providing a communication platform on a truly global scale unlike anything the world has seen before. Its emphatic adoption by people from every corner of the world builds on many interesting patterns and characteristics that reflect the underlying economic, social and cultural makeup of the participants. Using data from a commercial social networking site with a global membership base, this paper provides an internal look at the social networking phenomena in developing country contexts, and how it compares with the rest of the world.

LinkedIn has members from every country in the world, including several million in Africa. It also has a strong presence in Asia and Latin America, with countries like India and Brazil among the most active in the world. As a professional networking site, LinkedIn also has unique access to information such as career industries and educational level of members. This gives us a rich set of demographic and location data to work with, augmented by detailed activity information for the website. This ranges from how members access the social networking service to how they make connections and interact with other members. We combine profile information with activity data to analyze several aspects of social networking usage in developing countries.

Our analysis is presented in the form of several themes that illustrate different dimensions of social networking use. While we focus on patterns and characteristics from developing countries, we will also present contextual information from the rest of the world, which provides interesting comparisons emerging from the underlying differences and similarities in the member base. Some patterns are unique to the developing world, often shaped by economic, social and cultural factors, or the brief history and attributes of Internet citizenship for many users in these environments. Other patterns transcend geographic and economic barriers, and derive from basic human social behavior in sharing, communication and interaction. The goal of this paper is to provide researchers a revealing look on the growth, adoption and characteristics of social networking in developing countries. While we believe these characteristics are good indicators of social networking use in emerging economies, it is important to note that our analysis is solely based on data from LinkedIn, only one of several commercial social networks.

This paper discusses six characteristics and patterns, ranging from the interconnectedness of members in various geographic regions to the demographic and educational makeup of participants. Social networks enable members to make connections with other members throughout the world, and we will begin by investigating how people choose to connect with each other. In particular, we will look at the geographic locality of social network connections, and trends that emerge from cross-country and cross-continental connections.

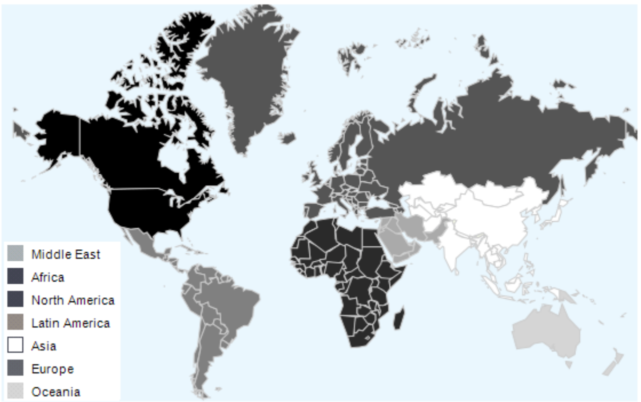


Figure 1: Region classifications

We then consider the overall engagement and activity of members in developing countries, and how it compares with the rest of the world. This can be expressed in several terms, including the growth of personal connections of members in the network, and the frequency and length of visits compared to members from more connected environments. This is augmented by a look at access devices from developing regions. We will investigate how members are accessing social networking sites from various regions, and how this trend maps the increase in mobile and Internet penetration in many developing countries.

Another important component in understanding social networking use is the demographic and educational makeup of members. We will explore generational bias in Internet access and participation, as demonstrated in the age distribution of members across the world. In addition, we look at gender representation, and how cultural and access barriers are reflected in social networking usage. Alongside demographics, we investigate the educational background and industry representation of members from various developing countries and corresponding worldwide trends, discussing biases in membership due to economic status and access to technology.

Finally, we look at the impact of local languages in social networking participation. We will investigate how content access in a local language influences adoption, and the extent of this influence in various regions. As new languages are introduced, we analyze how it affects usage in developing countries. To show this effect, we consider pairs of countries with the same national language, but different economic and cultural backgrounds, and explore the correlation between local languages and adoption.

The rest of this paper is organized as follows. We will first look at the data collection and extraction process, which forms the basis for the rest of the paper (§2.1). We will then look at the logistics of processing large amounts of data, often measured in several terabytes, in an efficient manner (§2.2). This is followed by the analysis of social networking use in developing countries along several dimensions, which makes up the majority of the paper (§3). We will finish by discussing related work (§4), and providing our conclusion (§5).

2. DATASET

This section describes the process of data collection and data analysis used throughout the paper. We will first discuss the pipeline of data collection on the LinkedIn platform, and how data is aggregated from several sources. We will then introduce the data analysis infrastructure used in this paper, which also supports many LinkedIn services.

2.1 Data collection

We collect two main types of data at LinkedIn. The first form is replicated from production databases, which consists of data mostly provided by LinkedIn members. This data includes member profile information, their education, and their connections with other members.

The second form is activity-based tracking data, which corresponds to logins, pageviews, and user agents. This data is aggregated from production services using Kafka [12], a publish-subscribe system for event collection and dissemination developed at LinkedIn. As of writing, Kafka is aggregating hundreds of gigabytes of data and more than a billion messages per day from LinkedIn's production systems.

The data analysis presented in this paper is done over a large amount of data, and is intended to represent interesting characteristics of social networking use in developing countries. In order to classify countries into groups, we use the UN geoscheme for macro geographical regions from the United Nations Statistical Division [4]. Member country information is provided during registration.

This scheme is based on the M49 classification, and is often used for statistical analysis purposes. To better represent socioeconomic differences, we make two common adjustments. First, we group Central America, the Caribbean and South America into "Latin America and the Caribbean" (or "Latin America" for short). In addition, we extract Western Asia (the Middle East) into a group of its own. While we have represented Africa, Asia and Latin America separately in our analysis, we often refer to their combination as the developing world, and compare their statistics with North America and Europe. Figure 1 shows a map representation of our regional classification.

When we represent countries on figures, we have sometimes used the ISO 3166 two letter country code for graphics readability. Further, any chosen countries have at least 50,000 LinkedIn members so as to avoid skew due to sparsity.

2.2 Analysis Infrastructure

One of the core pieces of data analysis infrastructure at LinkedIn is Hadoop, an open source implementation of MapReduce [8]. MapReduce provides a framework for processing big datasets on a large number of commodity computers through a series of steps that partition and assemble data in a highly parallel fashion, simplifying the process of writing parallel programs by providing the underlying infrastructure, failure handling, and simple interfaces for programmers.

At LinkedIn, and for the analyses in this paper, we use MapReduce as well as two scripting languages on top of Hadoop: Pig [14], a high level data flow language, and Hive [17], a SQL-like language. After aggregations are computed on Hadoop, the resulting data is small enough to be processed by common tools locally on a single machine.

3. DATA ANALYSIS

This section discusses six themes in understanding social networking usage in developing country contexts. As this work is a comparison of the developing world against the developed world, we normalize all data to the United States or North America respectively. In a couple of cases, we have had to estimate data, which has been clearly documented.

3.1 Connections

Our first topic focuses on the composition of connections in the social network for members from various regions. A connection is established when a member requests an invitation with another member in the network and is later approved by the invitee. Online social networks enable participants to establish connections with members from around the world, and this section investigates the makeup of these connections. Note that connections are bidirectional, with each connection linking two members in both directions.

An interesting pattern in analyzing social network connections is the interconnectedness of members within a region, or the local-

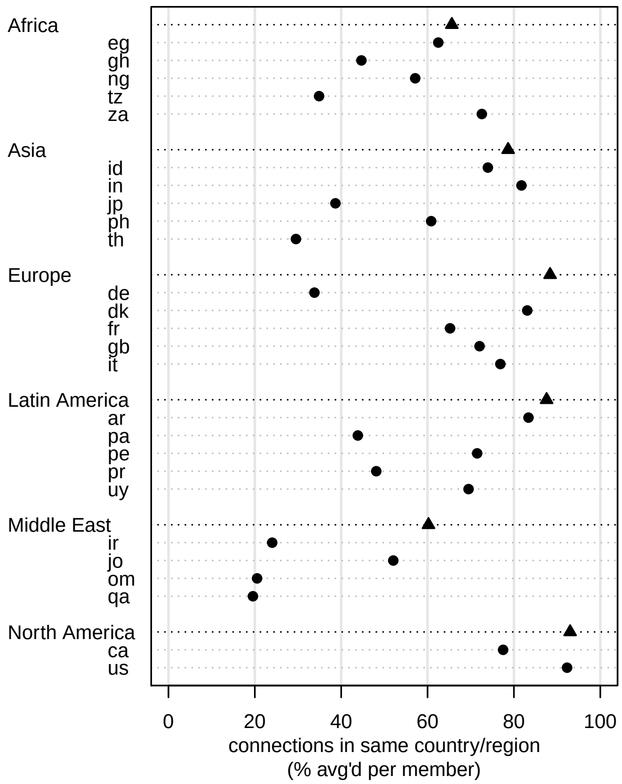


Figure 2: Geographic interconnectedness—fraction of connections originating from the same geographic region as members in that region

ity of relationships. We express the geographic locality of relationships by computing the ratio of connections that are established to members within the same geographic region. When considering macro-geographic classes, we measure the fraction of connections established within the same macro-geographic region. We then break the numbers down by country, and consider connections within the same country.

Figure 2 shows the interconnectedness of various regions. For each macro-geographic region represented, we also provide a few selected countries from the same region, and consider connection locality at the country level. Africa and the Middle East have two of the lowest rates of geographic locality: nearly 40% of connections in each respective region is established with members outside the region. Figure 3 shows geographic locality of connections for all countries in Europe and Africa on a map. As the dots on each country get larger and darker, connection locality increases.

One of the important factors in understanding connection locality is the membership population from each region. Intuitively, as the number of members in a region increases, the chances of establishing a relationship with similarly located members increases. However, this can be balanced out by the increase in membership of other regions, which also provides more opportunities for cross-country and cross-region relationships. We find some correlation between the size of the membership base in a country and the rate of connection locality ($\rho \approx 0.6$, for countries with more than 100,000 members).

Another interesting way of looking at connection distributions is to consider how far connected members are from each other. Using location information, we estimate the distance between members using the Haversine function [16]. Figures 4(a) and 4(b) show connection distance for macro-geographic regions and a selection of countries. For each region, the distance distribution is computed by considering all connections that originate from the region. On average, Africa and Asia have the two longest distances for connections, and this is also reflected in the individual countries represented.

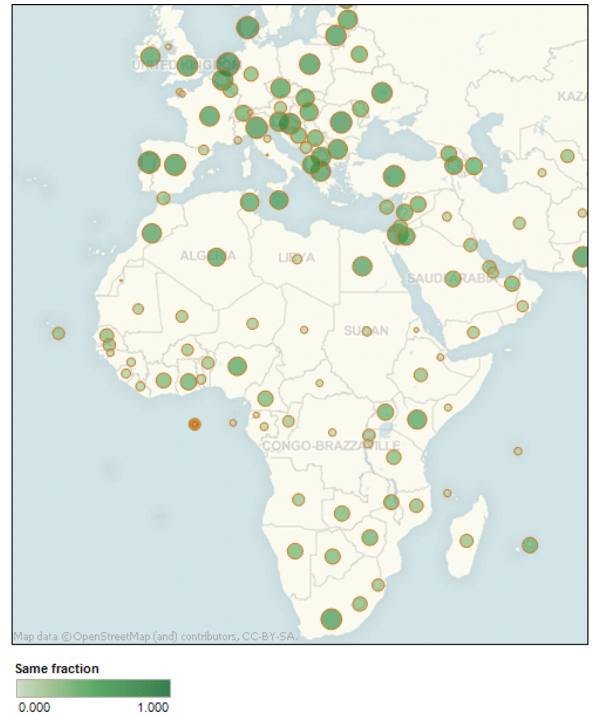


Figure 3: Geographic interconnectedness—locality of connections for members in Africa and Europe, computed per country. Bigger dots represent more connections established to members within the same country.

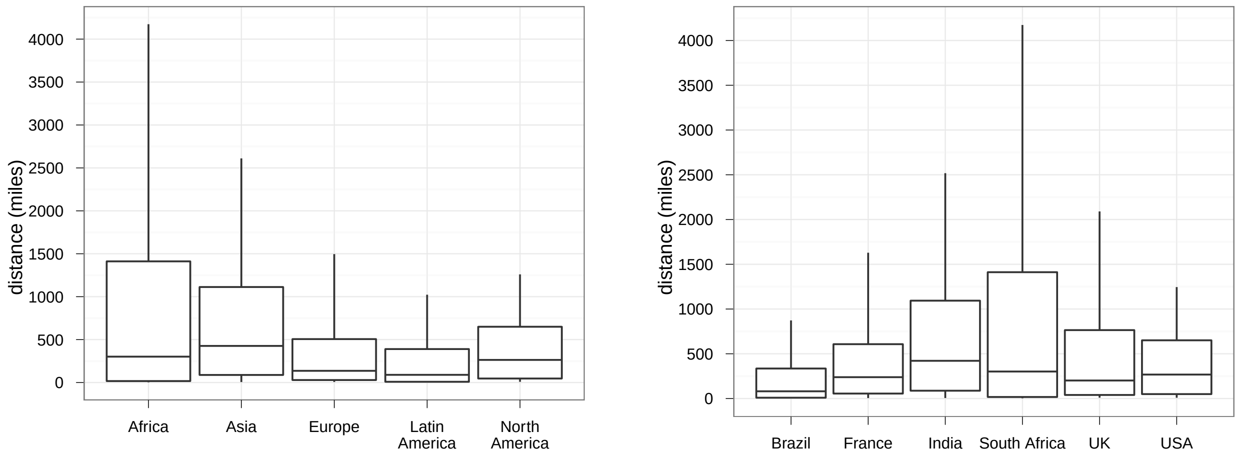
Several reasons, including geographic attributes of the region and the rate of connection locality, affect this distribution.

For connections that do not terminate in the same geographic region, we consider patterns in cross-country and cross-continental relationships. Figure 5 is a branching map that depicts where outbound connections terminate for a few selected countries. For each country, we show a few countries where members in the originating country have connections to. The thickness of the line for each arrow corresponds with the fractions of connections that terminate in the destination country. To avoid cluttering, we have removed self loops, instead providing the fraction of local connections as a percentage.

3.2 Activity

Measuring the activity and engagement of social networking participants is an important component in understanding usage from various regions. This manifests itself in several ways, ranging from how actively members are making connections on social networking sites, to the duration of visits to social networking sites. Several web usage studies in developing country contexts have indicated that users spend a significant fraction of their online time on social networking and communication websites [13, 15]. This section presents a few metrics to further delineate usage across developing countries.

An important aspect of social networking activity is the rate of establishing connections. Much of the utility in social networks is driven from communicating with fellow members, and connection growth is a key indicator of member engagement. Figure 6(a) presents the normalized rate of connection growth for January 2011 across several regions, which is calculated as the average month over month growth in the number of connections for members from each region. This rate is normalized such that the rate of connection growth for North America is one. Figure 6(b) presents the connection growth information for a selection of countries for January 2011, normalized to give the US a value of one.



(a) Regions

(b) Selected Countries

Figure 4: Connection distances—distribution of estimated physical distances (in miles) of network connections for members from various regions

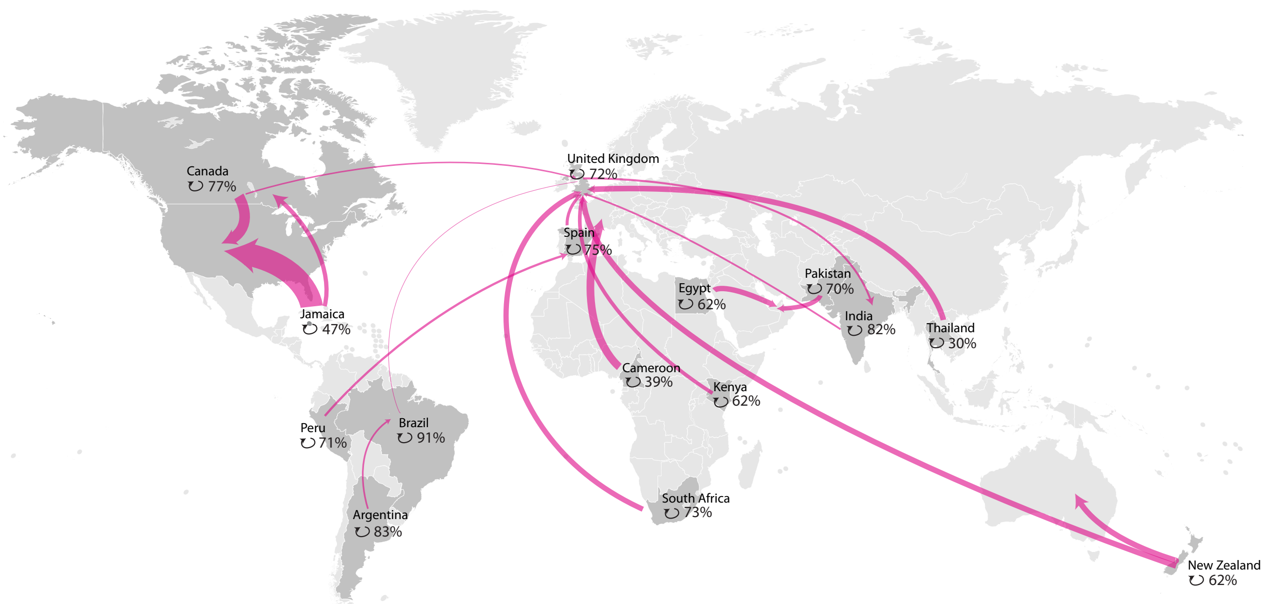


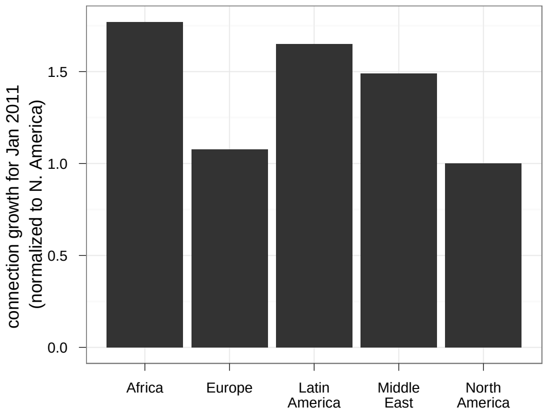
Figure 5: Outbound connections—a representation of cross-country social network connections for various countries

Connection growth is the fastest in developing regions, which is also reflected in the individual countries represented. One of the main reasons for this is the increasing addition of new members from these regions who are actively making connections on the network. In the early stages of social networking use, members actively add connections to their network. Even then, however, some regions are more active in adding connections. For example, during January 2011, members from Africa are adding new connections quicker than their counterparts in Latin America, although the membership base is growing around 30% faster in the latter.

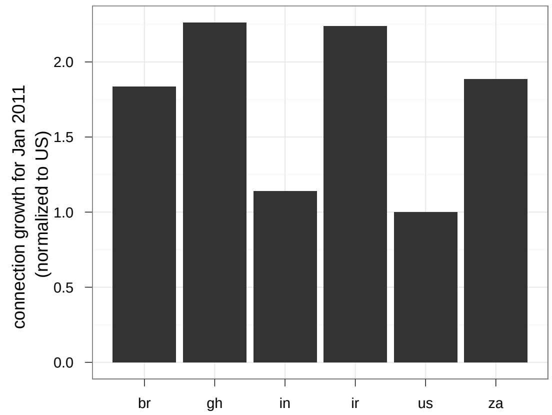
Another metric we consider in analyzing activity on the social networking platform is the duration of visits. A session is defined as a continuous user activity with an idle period of at least 30 minutes indicating a new session. Figure 7(a) shows the average normalized length of sessions for January 2011. In general, sessions established from developing regions generally last longer. A key factor for these differences is network connectivity, which varies significantly for different regions. For example, low bandwidth connectivity in African and Asian countries requires members from those regions to spend more time interacting to obtain the same information as North American members who have shorter lived sessions.

To better represent this relationship, Figure 8 plots the average and peak inbound bandwidth from accesses in developing regions on July 1, 2011 normalized to North America. These measurements are obtained by a system that monitors LinkedIn's inbound network traffic from several endpoints around the world, in part for detecting and preventing network attacks. These numbers correspond to Akamai's state of the Internet report [5] which estimates average connectivity from various regions. As mentioned earlier, the low bandwidth connectivity of members from developing regions is one of the reasons for elongated sessions.

In addition to session duration, visit frequency is an important metric for comparing social networking engagement across developing regions. Figure 7(b) plots the normalized, average number of sessions per member for various regions for January 2011. The data is normalized such that members in North America have a visit frequency of 1. Developing countries generally have a high number of visits per member, although this could be skewed by the influence of

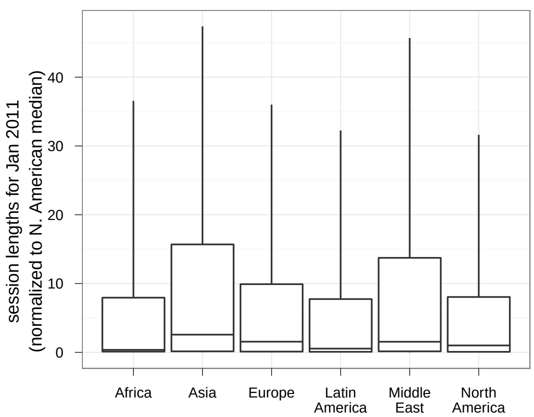


(a) Regions

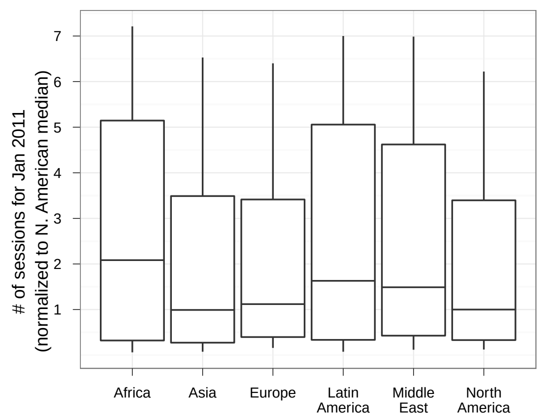


(b) Selected Countries

Figure 6: Connection growth—the rate of establishing new connections for members in various regions



(a) Normalized session length



(b) Normalized visit frequency

Figure 7: Member activity—the duration and frequency of member visits from several regions

newer members. This also correlates with the growth in the number of connections described earlier.

3.3 Access devices

Mobile penetration has been one of the singularly most important factors in connecting developing countries locally and across the globe. By the end of 2010, there were more than 5.3 billion telephone subscribers around the world, with nearly a billion of them having access to 3G data services [1]. This growth has been largely driven by Asia and Africa, which have the two highest growth rates in the world. Mobile access is available to nearly 90% of the world population. A number of studies in the developing world indicate that for many people, the phone is the first, and sometimes only, gateway to the Internet [9].

In this section, we focus on how users access social networking services from various regions. We broadly divide access verticals to mobile and desktop. Mobile accesses include visits that were directly made from mobile web browsers, or through applications that access the social network over a set of API's. In addition to smartphone applications for platforms like the iPhone, Android, and BlackBerry, LinkedIn also has a native Symbian application that runs on Nokia phones, which are more common in developing countries. The fraction of mobile accesses is the metric of interest in this case.

For each region, we compute the ratio of accesses made from mobile devices. To compute access ratios, we look at the fraction of sessions that were made from mobile browsers and applications, aggregated by geographic regions. Each ratio is computed as a fraction of accesses from mobile devices in that region to the total accesses from the same region on a monthly basis. Each month has been normalized to the mobile access fraction in North America or the US, respectively.

Figure 9 shows the ratio of mobile sessions by region and select countries for a period of 10 months, each month normalized to North America or the US, respectively. At a regional level (c.f. Figure 9(a)), Latin America and the Middle East have the highest fraction of mobile accesses, and mobile accesses have been increasing significantly. Figure 9(b) presents a selection of countries in developing regions with high mobile access ratios. For example, in Africa, Nigeria has some of the highest mobile access rates with nearly 3 times as much as mobile accesses from North America.

3.4 Demographics

This section focuses on the age and gender composition of social networking participants from developing regions. The age of members is estimated from their profile education information: we assume members were 21 years old when they start their career. While this technique might be a good approximation for members in western countries, a caveat is that its effectiveness might vary in different areas where career start ages might be generally different.

Generational bias is an important factor in online participation. Perhaps more interesting is that this bias tends to operate on a global scale. When we look at the average and median ages of social net-

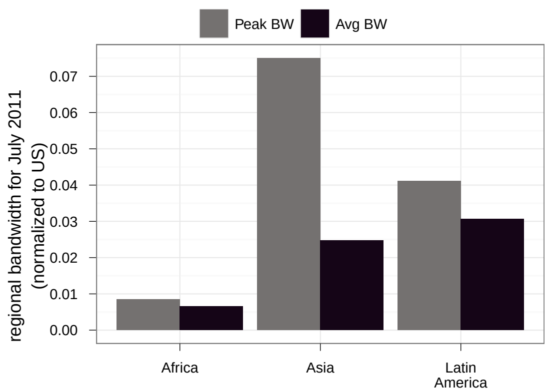


Figure 8: Regional bandwidth—peak and average inbound bandwidth from members in developing regions

working members from across the world, participation is naturally skewed towards people of younger age. The age distribution is also interesting when considering the underlying makeup of the population in different regions. Countries in developing regions generally have a younger population, which is reflected in the social network representation of age groups.

As shown in Figure 10(a), the median ages for members from North America are a few years higher than those in Africa or the Middle East. The age distribution for Asia and Latin America is also quite similar, roughly within a year compared to members in Africa. Broadly speaking, younger median ages for members from developing regions correlate with the differences in median ages of the underlying population. Figure 10(b) shows a selection of some countries from each region with median ages in the highest or lowest quantiles.

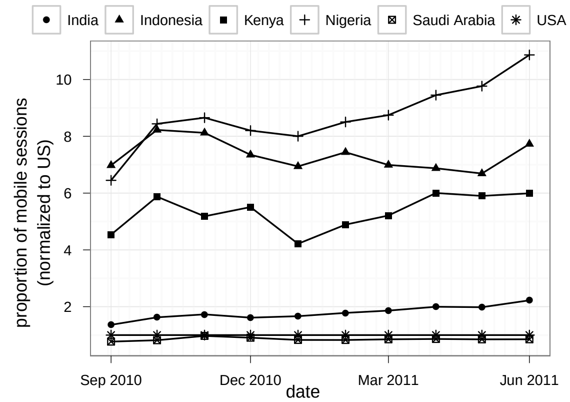
Gender representation is another interesting characteristic for many regions. We approximate gender information by classifying member first names using a large annotated catalog retrieved from several baby name books. Names that could not be mapped to a gender—the catalog of baby names is biased to Western names—or are ambiguous are labelled "unknown." In all of the gender information figures, we have represented the fraction of users we were not able to map to genders. The unknowns are rather high, so any conclusions should be viewed with some suspicion.

Figure 11(a) shows the female membership ratio for a few regions. Globally, males are generally overrepresented by membership, but the differences are more pronounced in many developing countries. Many of these differences can be attributed to social gender roles and economic differences. For example, the Middle East has the lowest female membership ratio in the world, with females making up less than 25% of the total membership base. The ratio is slightly higher for Africa, but with significant differences from country to country.

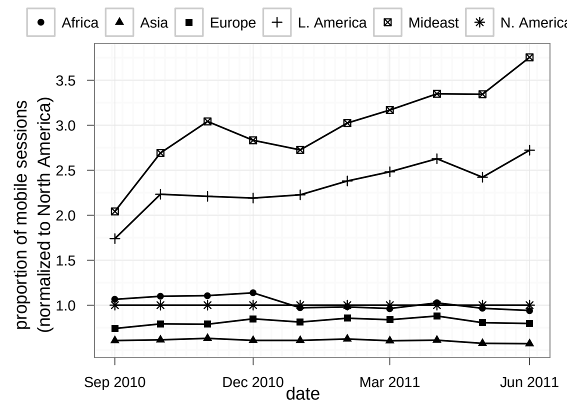
Figure 11(b) shows a selection of countries with various female representations from several regions. Latin America has one of the highest female ratios in the world, and several Asian countries have gender representation in line with the general population. However, countries like India and Bangladesh have a highly skewed male representation. In Africa, South Africa has one of the higher ratios with nearly 45% female makeup (compared to 49% in the general population [2]). North African countries share similar traits with the Middle East with lower female representation compared to the rest of the continent.

3.5 Education and Careers

The fifth theme in our analysis focuses on educational levels and career industries of members. As a professional social network, LinkedIn encourages its members to enter their educational and



(a) Regions



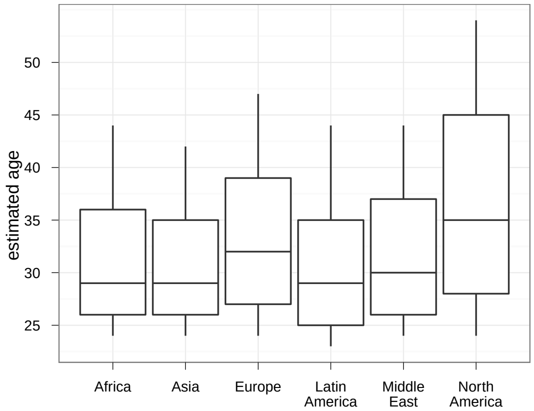
(b) Selected Countries

Figure 9: Mobile Access Growth—the ratio of sessions established from mobile devices, each month normalized to the fraction of mobile accesses in the US

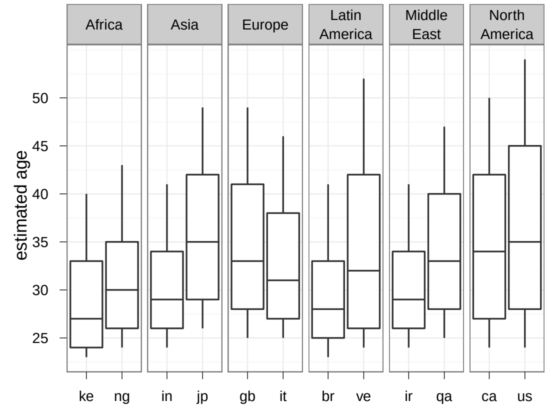
work history to their profiles. Education levels can include one or more user provided description of the member's educational history. Education levels are described differently across the world. For example a "diploma" in Ethiopia corresponds to a 2 year degree that is equivalent to an associate degree obtained from a community college in the US. To mitigate this problem, we broadly divide education levels to four: high school, college, masters, and doctorate. When a member has listed more than one education level on their profile, we pick the highest one. Members must also provide an industry when they describe their career. Industry captures a high level classification of career paths, and there are over 120 industries represented on LinkedIn.

We consider educational levels in several regions in comparison to educational makeup in North America. Figure 12 shows the distribution of educational levels in relation to North America or the US respectively for several regions and countries. With the exception of high school graduates, it is interesting to note the near uniform distribution of members with higher educational levels in several regions. Africa and the Middle East have a high fraction of high school graduates in the membership base. When considering Education Indices from the UN Human Development Report [3], we note that professional social networking membership is not representative of the underlying literacy rate and education index in many developing countries. Rather, it is skewed towards to relatively more educated members, which can translates to relative economic affluence, and improved access to connectivity.

Table 1 shows the top-5 industries represented from each region. Some differences appear as we look down the list of industries from

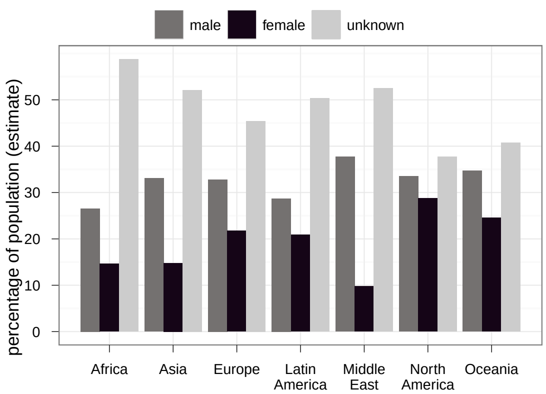


(a) Regions

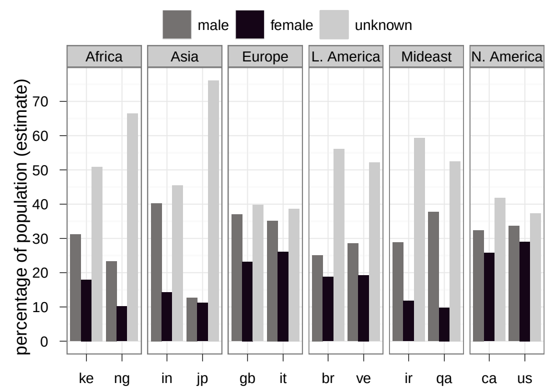


(b) Selected Countries

Figure 10: Demographics—estimated age of members in various regions



(a) Regions



(b) Selected Countries

Figure 11: Demographics—approximated gender of members in various regions

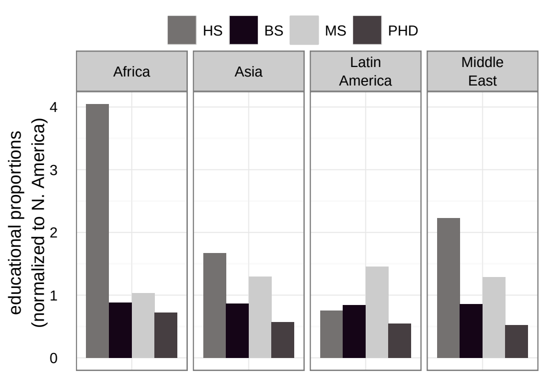


Figure 12: Educational levels—the proportion of educational levels making up the membership base in various regions

each region. These differences are more apparent when looking at a selection of countries as shown in Table 2. As expected, industry representation in a country tends to reflect regionally established industries, such as oil/energy in Nigeria or computer and software in India. In addition to the volume of professional regionally established industries hire, they also tend to have more international contacts, which has an impact on technology adoption.

Another interesting aspect of represented industries in the network is how members connect across various industries. In Figure 13, we look at the industry similarity of connections, which is defined in a similar manner as geographic interconnectedness in Section 3.1. For each member, we compute the fraction of connected members that also work in the same industry as the member. Interestingly, the rate of industry similarity remains very close for all the regions we considered, with members having only 20–25% of their connections from a similar industry. Intuitively, one might have expected most connections to remain within the same industry, where professional relationships are natural to establish.

3.6 Local languages

The last topic we consider in understanding social networking in developing regions is the impact of local languages on adoption. LinkedIn is available in a multitude of languages, a few of which are local languages for many regions in Africa and Latin America. We focus on a few chosen languages that have been available to members for at least one year.

We first look at the the membership growth rate for various languages. Figure 14 shows a month-over-month time-line from January 2008 through May 2011 of five languages normalized to English. There is a substantial increase in membership in the language when it is first introduced, which often remains high for about six months before it regresses to the mean.

It is often interesting to see the impact of local languages on particular countries. In order to make some comparisons, we choose three pairs of countries from different regions such that the national

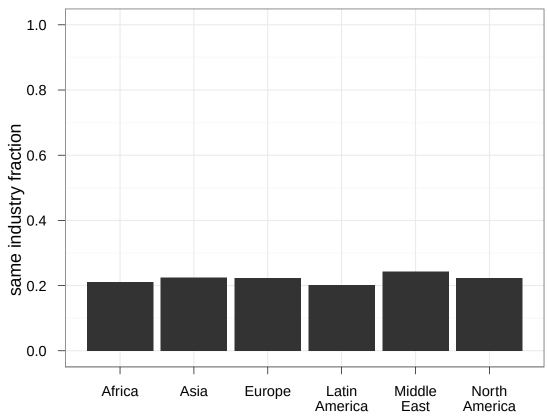


Figure 13: Industry interconnectedness—fraction of connections made to members in the same career industry for members in various regions

Region	Top 5 industries
Africa	accounting, banking, education management, information technology and services, telecommunications
Asia	computer software, education management, financial services, information technology and services, telecommunications,
Europe	computer software, financial services, information technology and services, marketing and advertising, telecommunications
Latin America	construction, higher education, information technology and services, marketing and advertising, telecommunications
Middle East	banking, construction, information technology and services, oil and energy, telecommunications
North America	education management, financial services, hospital and health care, information technology and services, real estate

Table 1: Top 5 industries by region (ordered alphabetically)

language for each pair is the same. These include Cameroon and France (French), Argentina and Spain (Spanish) and Brazil and Portugal (Portuguese). The top half of Figure 15 shows the average normalized month over month growth of languages for 2008–2010. The bottom half shows the average normalized rate of membership growth for each country, and how the rate changes as languages as introduced. In all of the cases, we observe membership growth responding more positively for countries from developing regions as the national languages are added. Some of this difference can attributed to the overall difference in membership growth across several regions, but the adjustment in the rate of growth a year after the language has been introduced indicates that local languages play an important role in early adoption, and even more so in the developing world.

4. RELATED WORK

One class of related projects come from web usage studies that provide a macro classification of how users spend time online. There are several web usage studies that have been conducted in developing country settings [7, 10, 13, 15]. Du et al. [10] evaluated HTTP traffic captured from shared access sites in Ghana and Cambodia. Their results demonstrate several features of web usage in developing coun-

Country	Top 5 industries
India	computer software, education management, financial services, information technology and services, telecommunications
Malawi	accounting, banking, education management, information technology and services, non-profit organization management
Nigeria	accounting, banking, information technology and services, oil and energy, telecommunications
Saudi Arabia	construction, hospital and health care, information technology and services, oil and energy, telecommunications
United States	education management, financial services, hospital and health care, information technology and services, real estate

Table 2: Top 5 industries by country (ordered alphabetically)

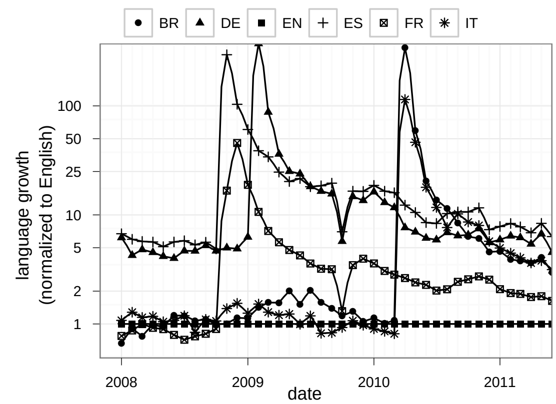
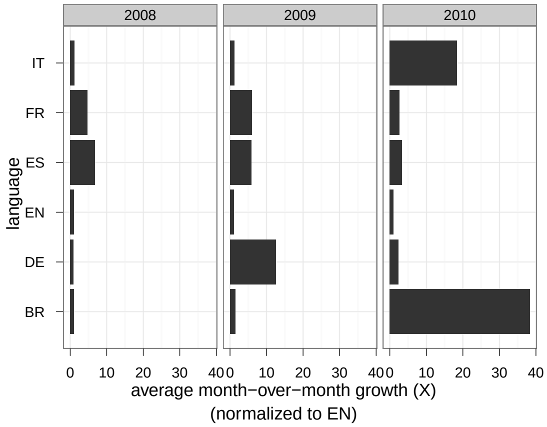


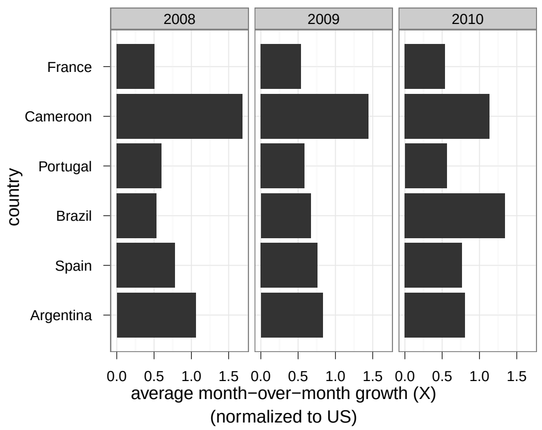
Figure 14: Locale growth—month over month growth of languages, with a focus on the impact of newly added languages

tries prior to the widespread adoption of social networks. Another study of Internet usage and performance in Zambia [13] points at the increased adoption of social networking and communication tools even in rural villages in Africa. Analysis of web usage in Macha, Zambia, some 350 kilometers from the capital city Lusaka, reveals several interesting findings, including social networking sites as the top visited destinations. More specific web access analysis from a school setting in India [7] also indicate wide usage of email communication and social networking. Our work complements this body of work by providing analysis on the adoption and usage patterns of social networking in developing regions. As social networking continues to be a dominant web usage scenario around the world, this paper provides researchers with some insights on the adoption and characteristics of social networking, particularly in developing regions.

A large scale study of web traffic using data collected from a world wide content distribution network (CDN) by Ihm et al. resembles our work in the scale of data analysis [11]. Their work analyzes web content that represents one week's worth of browsing data from nearly 350K users across 190 countries. They observe a number of interesting characteristics of web usage in developing regions, including the desire for rich media and differences in download type distributions. In contrast, our work focuses social networking usage at a global scale by using data from over a 120 million members that come from every country in the world, with tens of millions of those members from developing regions. We combine individual profile information



(a) Locale growth



(b) Membership growth

Figure 15: Membership with languages—the impact of language introductions on countries with the same national language, but different socioeconomic backgrounds

with member activity logs for providing researchers a revealing look on some patterns and characteristics of social networking usage in developing regions.

5. CONCLUSION

This paper provides the first large scale and detailed analysis of social networking usage in developing countries. As Internet access improves for users in those regions, online participation has been naturally increasing. Using profile and activity data from LinkedIn, a social networking site with over a 120 million members worldwide, this paper has presented several themes in social networking usage in developing regions. We looked at the characteristics of the nature of interconnectedness and geographic locality for members, the activity and engagement level of members in developing regions, as well as access verticals for content from various regions. We also discussed the demographic and educational makeup of members in developing regions, and the impact of local languages in social network adoption and growth.

The goal of this paper was to provide researchers with a detailed look on the characteristics of social networking usage in the developing world compared to the rest of the world. As several studies in developing regions have indicated [13, 15], users spend a sizable portion of their online time on social networking and communication websites. This study further explored social networking usage, further delineating its characteristics in developing regions. Using data from one of the largest commercial social networking entities with global reach, the paper focuses on several interesting dimensions of social networking use in developing county contexts.

REFERENCES

[1] Measuring The Information Society, 2010. International Telecommunication Union.

[2] The CIA world factbook. `http://www.cia.gov/cia/publications/factbook`, August 2011.

[3] United Nations Human Development Report, 2011. United Nations Development Programmee.

[4] Standard country and area codes classifications (M49), 2011. United Nations Statistics Division.

[5] Akamai Reports. The State of the Internet, 2nd quarter, 2009.

[6] Pew Research Center. Global Publics Embrace Social Networking. *Pew Global Attitudes Project*, 2010.

[7] Jay Chen, David Hutchful, William Thies, and Lakshmi Subramanian. Analyzing and Accelerating Web Access in a Shool in Peri-Urban India. In *Proceedings of the 20th International World Wide Web (WWW) Conference*, Hyderabad, India, 2011.

[8] Jeffrey Dean and Sanjay Ghemawat. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51:107–113, January 2008.

[9] Jonathan Donner, Shikoh Gitau, and Gary Marsden. Exploring mobile-only Internet use: results of a training study in urban South Africa. *International Journal of Communication*, 5:574–597, 2011.

[10] Bowei Du, Michael Demmer, and Eric Brewer. Analysis of WWW traffic in Cambodia and Ghana. In *Proceedings of the 15th International World Wide Web (WWW) Conference*, pages 771–780, Edinburgh, Scotland, May 2006.

[11] Sunghwan Ihm, KyoungSoo Park, and Vivek S. Pai. Towards understanding developing world traffic. In *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions (NSDR)*, pages 8:1–8:6, San Francisco, California, June 2010.

[12] Jay Kreps, Neha Narkhede, and Jun Rao. Kafka: A distributed messaging system for log processing. In *Proceedings of 6th International Workshop on Networking Meets Databases (NetDB)*, Athens, Greece, June 2011.

[13] Karel Matthee, Gregory Mweemba, Adrian Pais, Gertjan van Stam, and Marijn Rijken. Bringing Internet connectivity to rural Zambia using a collaborative approach. In *Proceedings of the 2nd ACM International Conference on Information and Communication Technologies and Development (ICTD)*, pages 47–58, Bangalore, India, December 2007.

[14] Christopher Olston, Benjamin Reed, Utkarsh Srivastava, Ravi Kumar, and Andrew Tomkins. Pig Latin: a not-so-foreign language for data processing. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 1099–1110, Vancouver, BC, Canada, June 2008.

[15] Azarias Reda, Edward Cutrell, and Brian Noble. Towards improved web acceleration: leveraging the personal web. In *Proceedings of the 5th ACM Workshop on Networked Systems for Developing Regions (NSDR)*, pages 57–62, Bethesda, Maryland, USA, June 2011.

[16] Roger W. Sinnott. Virtues of the Haversine. *Sky and Telescope*, 68(2):159, 1984.

[17] Ashish Thusoo, Joydeep Sen Sarma, Namit Jain, Zheng Shao, Prasad Chakka, Suresh Anthony, Hao Liu, Pete Wyckoff, and Raghotham Murthy. Hive: a warehousing solution over a Map-Reduce framework. *Proceedings of the VLDB Endowment*, 2: 1626–1629, August 2009.