Social Multimedia, Diversity, and Global South Cities: A Double Blind Side

Daniel Gatica-Perez
Idiap Research Institute and EPFL, Switzerland
gatica@idiap.ch

Joan Isaac-Biel
Amaris, Switzerland
joanisaac.biel@gmail.com

Darshan Santani
Pepper Cloud, Singapore
dsantani@peppercloud.com

Thanh-Trung Phan
Idiap Research Institute and EPFL, Switzerland
tphan@idiap.ch

ABSTRACT

Social media provides opportunities to examine urban phenomena at scale, and we believe that studying cities in the Global South through citizen-contributed data and AI-driven analytics should be a priority of multimedia research. However, little work has been done in our community, and we argue that this contributes to a double blind side problem. We exemplify this situation by studying Ma3Route, a mobile social media channel to crowdsource and broadcast transit reports in Nairobi, Kenya. Using multimedia data from its Twitter stream, we first conduct a descriptive analysis that shows an active community generating rich traffic-related reports, and then discover latent topics that identify both regular and ephemeral thematic clusters of reports involving accidents, traffic conditions, and attitudes of citizens towards authorities. In the second place, we conduct a deep learning-based analysis of Ma3Route images to understand the kind of visual content shared in the platform, and that shows limitations of using deep neural network models trained with data largely coming from the US and Europe, which do not fully match the reality and diversity of other world regions. We conclude by presenting a multidisciplinary research agenda for future work in this domain.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile computing; Ubiquitous and mobile computing design and evaluation methods.

KEYWORDS

Social Media; Mobile Crowdsourcing; Cities; Urban computing

1 INTRODUCTION

Social media channels are used by citizens across the globe to voice their opinions on social and economical issues faced by the cities they inhabit. According to official statistics, 25 of the 30 largest urban areas in the world are located in countries in the so-called Global South [49]. This would suggest that studying phenomena in large cities through multimedia analytics should include the study of cities in Africa, Asia, and Latin America.

Surprisingly, however, this is not the case for urban multimedia if one examines the body of research in the past years (with the notable exception of China). A review of ACM MM long papers published in the last years shows little or no work on multimedia analytics involving cities in these world regions. In contrast, analytics of multimedia data coming from economically developed cities like New York, London, San Francisco, and others have received substantial attention in the literature [23, 30, 34, 35, 37]. Recent papers specifically published in ACM MM connecting multimedia analysis and such cities include [36] (New York and Boston) and [22] (Rome and Milan).

Our position is that studying cities in the Global South through citizen-contributed data and AI-driven multimedia analytics should be a priority of multimedia research, and that not doing it contributes (by omission) to produce a double blind side: the first one with respect to our understanding, as a research community, of phenomena in Global South cities that have different dynamics and issues to solve than wealthy countries; the second one with respect to what machines can learn (or not) about these cities, when they are represented by data that does not reflect the world’s diversity.

Cities in the Global South have been studied outside the multimedia literature, often in the realm of development studies [52] and social sciences [28], and frequently focused on the study of crises [43] and violence [32]. Needless to say, most Global South cities are not in a permanent state of crisis [47]. We argue that the study of everyday urban life phenomena in African, Asian, and Latin American cities with multimedia analytics is important because the fact that most of the urban world is in the Global South opens a wealth of opportunities (as many of these cities produce both social media and other citizen-contributed data) to produce valuable findings that could support communities, organizations, and government to address some of the issues citizens face. Along with it, however, come a number of challenges, related to the need to understand the social practices and cultural contexts in which citizen-contributed multimedia is created, shared, and appropriated.
We exemplify this situation with Ma3Route, a platform that uses mobile and social media to crowdsource transit reports in Nairobi, Kenya. It has been estimated that this city loses each year an equivalent of 200 million USD in lost productivity as a result of traffic [8]. Ma3Route provides a community-driven channel that asks citizens to share traffic reports, and then broadcasts these reports using Twitter to inform commuters in real-time. Users also use the platform to voice their opinions on transportation and safety policies, as well as mistrust and frustration towards the city administration and the police. As of July 2019, the platform has 1.14M followers on Twitter.

This case study is relevant for several reasons. First, by investigating the Ma3Route Twitter channel, we contribute to the examination of the urban dynamics and mobility in Global South cities as expressed in social media, which are often different in characteristics when compared to cities in the US and Western Europe. Second, by studying Nairobi, we analyze a number of common trends that apply to many large urban areas in the world. Nairobi is one of the top 100 populated urban areas worldwide, 80% of which are in the Global South [49]. Finally, the data reveals current limitations of deep learning systems with respect to the diversity of urban scenes represented—an issue that the fairness, accountability, and transparency emerging literature has demonstrated for individual citizens in visual tasks like face recognition [20], but that is also applicable to urban areas as we show here.

The contributions in this paper are three-fold. First, we collected tweets from Ma3Route’s Twitter multimedia stream over a period of three months; conduct a descriptive data analysis enriched by qualitative observations made on both tweets and users, which reveals the main characteristics of the platform and the kind of contributed content; and apply topic modeling to discover the main themes of conversation, identifying both regular and ephemeral thematic clusters of citizen contributions, which go beyond accident and traffic reports and reflect the concerns of citizens towards road conditions, transportation, and city authorities. Second, we present a deep-learning based analysis of Ma3Route images to understand the visual content shared in the platform, revealing the diversity limitations of using pretrained deep networks trained under the diversity limitations of using pretrained deep networks trained in the Global South [49]. Finally, the data reveals current limitations of using mobile crowdsourcing to gather and document road quality information in Nairobi [40] or to map the informal bus transportation network [50]. Other campaigns to report urban concerns in Latin America include e.g. [39].

The idea of crowdsourcing traffic updates and broadcasting them to commuters is not new. Waze, which was acquired by Google in 2013, is a popular mobile application with millions of users worldwide [15]. Waze works by collecting GPS sensor data from its users in order to provide real-time traffic updates. As Ma3Route, other similar systems have been created in other countries. In India, Trafline [14] is a mobile and web application, that provides traffic information across several major metropolitan areas. TrafficDito in Manila (Philippines) [13], and ma2too3a in Lebanon [10] were created to provide similar services to their residents. Unfortunately, all these channels, which serve large cities in the Global South, have remained invisible to multimedia research until now.

3 BLIND SIDE #1: MA3ROUTE AS CASE STUDY

We first introduce Ma3Route as a sociotechnical system that illustrates the diversity of issues of Global South cities. We then describe the collected dataset, followed by a data-driven descriptive analysis, and a more in-depth analysis based on topic models. This section reveals that cities in the Global South can indeed be analyzed using citizen-contributed data, but that this promising domain has been largely unaddressed by multimedia research.

3.1 Platform Overview

Ma3Route crowdsources transit reports in Nairobi to provide users with information about traffic, matatu (informal buses in Kenya) directions, and driving reports in almost real-time. The platform aggregates, curates, and broadcasts citizen reports using Twitter, and its own web and mobile applications. Reports consist mainly of text and images, and do not rely on GPS or maps to geolocalize information. Instead, users use road names and other well known places and landmarks as text to locate their reports. This allows
We used the public Twitter REST API to collect tweets posted on the Ma3Route Twitter channel for a period of 3 months (Jun-Sep 2015). This resulted in a collection of 60,086 tweets. A tweet contains the text of the tweet, timestamp, user details, hashtags and media items. In addition, the Ma3Route website shows tweets from its Twitter channel plus associated report categories [11]. The produced Twitter data is relatively small in terms of volume compared to those of large US or European cities. However, we argue that missing the importance of studying urban processes of certain regions because of the relatively modest data volumes they produce is actually part of the blind side problem we discuss here.

### 3.2 Ma3Route Dataset

We used the public Twitter REST API to collect tweets posted on the Ma3Route Twitter channel for a period of 3 months (Jun-Sep 2015). This resulted in a collection of 60,086 tweets. A tweet contains the text of the tweet, timestamp, user details, hashtags and media items. In addition, the Ma3Route website shows tweets from its Twitter channel plus associated report categories [11]. The produced Twitter data is relatively small in terms of volume compared to those of large US or European cities. However, we argue that missing the importance of studying urban processes of certain regions because of the relatively modest data volumes they produce is actually part of the blind side problem we discuss here.

### 3.3 Descriptive Analysis

#### 3.3.1 User contributions

We used a standard regular expression mechanism to process the “via @username” part of the tweet to recover the user IDs who authored the original tweet. In the dataset, 97% of tweets were generated by 15,534 unique Twitter users other than Ma3Route itself. The rest of the posts were authored by Ma3Route using their admin handle, or seem to be created by community managers that directly respond or retweet to content using standard Twitter mechanisms, i.e. without the use of the “via @username" convention. The number of contributions per user follows a tailed distribution (max=660; median=3.8; mean=1) during the period of our analysis. We inspected the profiles of the top twenty contributors and found local public organizations promoting initiatives on road safety, in addition to Ma3Route itself. The rest of the top contributors are local Nairobians who contributed between 120 and 550 tweets each.

#### 3.3.2 Multimedia content

We quantified the amount of multimedia content available in terms of user mentions, hashtags, and photos. Table 2 shows the frequencies of these entities in terms of tweets, raw frequency, and unique values.

**User mentions:** References to users are used to direct content or simply mention users. Twitter notifies users when someone mentions them. Mentions reflect an explicit interest from users towards others, thus looking at the top mentioned users is informative of the content. First, we processed the metadata to avoid counting mentions generated using Ma3Route’s “@via username” convention, discarding 86% of the tweets and leaving a total of 10,225 user mentions. We inspected the profiles of the top twenty mentioned users and found that most of them were public figures (e.g., Nairobi’s Mayor, Governor Office, etc.) and public institutions (e.g., Kenya’s National Transit and Safety Authority, Kenya Police, etc.) responsible for traffic control and security, road infrastructure, and public transport. This shows how the Ma3Route community actively points to the main actors of the city administration.

**Hashtags:** Hashtags reveal an explicit interest of users to link to specific conversations. We inspected the top 20 hashtags and found that half of them are road names such as #waiyakiway or #mombasaroad. However, we did not find any evidence of users using road names in hashtags as a general convention to create traffic reports around specific landmarks. We also found other hashtags that refer to road safety campaigns such as #zushaleo or #trafficswatch or to major national events such as the US president Obama visit to Nairobi in July 2015 (#obamareturns, #kiderogras), or a matatu incident reported in the media (#arrestrongaimatatus).

**Photos:** Photos complement text reports, and are valuable to document road hazards [40], as testimonial data to investigate accidents, to support road safety campaigns, and to validate the authenticity or reliability of text reports. As shown in Table 2, 18% of tweets contained photos. This is a considerable proportion of photos compared to systems like FixMyStreet that have an estimate of 11% of posts including photos [44].

#### 3.3.3 Reports over time

The time of day of a tweet provides evidence of the situational context that motivates people to report. In Figure 1, we plot the daily evolution of traffic reports as a result of aggregating tweets by time of the day (normalized by the total number of days). The trend peaks in the morning around 7AM and in the evening around 5PM, which corresponds to peak commuting hours in Nairobi. At peak hours, Ma3Route receives between 1–2 reports reports per minute, whereas between these two peaks, the number of reports decreases by half.

---

**Table 1: Ma3Route categories of traffic and driving reports.**

<table>
<thead>
<tr>
<th>Tweet Entity</th>
<th>% Tweets</th>
<th>Total Freq</th>
<th>Num Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>User mentions</td>
<td>14%</td>
<td>10225</td>
<td>2072</td>
</tr>
<tr>
<td>Hashtags</td>
<td>14%</td>
<td>10299</td>
<td>4141</td>
</tr>
<tr>
<td>Photos</td>
<td>18%</td>
<td>14138</td>
<td>14023</td>
</tr>
</tbody>
</table>

**Table 2: Multimedia entities contained in the dataset.**
we filtered and preprocessed the data. We first selected tweets using Twitter automatic language detection. We then performed standard text preprocessing which included lowercasing, URL, stopword and punctuation removal, and stemming. We also removed user mentions, but kept hashtags as they might potentially contain meaningful keywords for content analysis. We also removed words that appeared in less than 10 tweets. After processing, the dataset consists of 52,626 tweets with a vocabulary size of 3,422 words.

We then used Latent Dirichlet Allocation (LDA) [19]. To train LDA, we treated each tweet as a single document. Previous research has shown that despite the short length of tweets, LDA can obtain meaningful topics from Twitter data [38, 53]. We used a Gibbs sampling implementation [27] requiring two parameters to control the Dirichlet priors on the distribution of topics ($\alpha$), and the distribution of terms per topic ($\beta$). It is common practice to use $\alpha = 50/T$ and $\beta = 0.1$ [26]. We found it appropriate to use a smaller value of alpha $\alpha$ to increase the likelihood of assigning one single topic to each document. Finally, based on an empirical procedure, we found $k = 10$ topics to be suitable for our corpus.

### Topic Inspection

In Table 3, we list the ten discovered topics along with their most relevant words. The relevances of the top words is computed using a measure that combines the probability of terms per topic $\beta$ with the probability of each term in the full corpus $\gamma$. The discovered topics expand our understanding of how citizens use the platform. Topics 6 to 10 are traffic-related topics, while Topics 1 to 5 highlight citizens’ concerns towards local issues related to transportation and safety policies. Below, we describe each topic in more depth. Following guidelines for social media research [46], we paraphrase (rather than quote) some of the top tweets for each topic to reflect the opinions of users.

**Topic 1 – Matatus ($P(z) = 0.12$).** Matatus are privately owned minibuses used for inter and intra-city commute in Nairobi. Due to the lack of public transportation, they provide the necessary transit means to millions of Nairobians every day. Matatu drivers are unfortunately famous for reckless driving, breaking traffic laws, cutting off other vehicles, using sidewalks to bypass traffic jams, and over-speeding [1, 5]. This perception of matatu drivers is reflected in Topic 1 from the top relevant keywords comprising this topic: matatus, drivers, passengers, suspend, loud, reckless, wrong, and shame. One example complaint is that certain crossing points are rather dangerous, and that pedestrian crossing signals are needed.

As both Matatus and Embassava are cooperatives of matatu industry. Matatu drivers’ income and job security depends on the daily passenger load [25, 33]. In addition, matatus are known for being the target of violence against drivers, bystanders, and commuters. This appears in some of the top tweets, which tell users to beware of specific routes where teams of pickpockets operate.

**Topic 2 – Police ($P(z) = 0.12$).** This topic reflects the attitude and mistrust of citizens towards police authorities. The relevant words in this topic are police, speed, traffic, arrest and bribe. In the top tweets, users argue that the city’s video camera network might be used by the police for bribing purposes, or complain that the NTSA (the National Transport and Safety Authority, a government body tasked to manage road transport) does not do a proper job.

**Topic 3 – Obama ($P(z) = 0.11$).** This topic corresponds to the state visit of the U.S. President to Kenya in July 2015. Road closures and disruptions by the city administration along the five major road segments [6] prompted citizens to voice their discomfort to their daily commute and to share traffic updates on road closures as a result of the visit. This is captured in the top terms – obama, road, close, mombusa or friday. Other terms such as kidero and grass make reference to a related controversy created by the governor’s decision to plant grass along the roads only a few days before President Obama’s visit in an effort to give the city a facelift. The #kiderograss hashtag [4] was used to criticize the situation, as shown in top tweets, that comment with irony about how grass grew overnight, and how this seriously affected traffic. Topic 3 is an example of an ephemeral topic, capturing the spike of reports generated during the visit.

**Topic 4 – Accidents ($P(z) = 0.11$).** As indicated by the top word (i.e. accid), this topic concentrates on tweets related to road accidents. Tweets report the location of the accident (bypass), the probable causes for accidents (cause, hit), the parties involved (lorri, truck, car, bus), and the resultant injuries or deaths (dead, injury).

**Topic 5 – Safety Advise ($P(z) = 0.11$).** This topic includes comments on safety and tips to fellow commuters, including conversations partly initiated by Ma3route together with partners to inform citizens about various initiatives such as the ZushaLeo initiative on road safety [16] which recommends drivers to stay vigilant, avoid drinking alcohol while driving, and respect pedestrians (and children in particular) as they cross the street.

Topics are not mutually exclusive, e.g. Topic 1, Topic 2, and Topic 5 capture related issues. For example, there is urban development...
literature that has documented a poor relationship between police and matatu drivers, leading to harassment of matatu operators, and matatu drivers paying bribes to police to avoid harassment [25]. This interplay between Topic 1 and Topic 2 is illustrated in reports that describe checkpoints in the city where police receive bribes from matatu drivers and criticize the consequences of such behavior. In addition, matatus are often criticized for being insecure, which explains why Topic 2 and 5 may overlap.

The remaining topics capture reports about traffic flow. LDA splits reports based on the geolocation of landmarks and type of traffic (slow, moderate, clear, traffic jam, etc.). Figure 2 shows the map of Nairobi with the regions highlighted as in Topics 6-10.

**Topic 6 – West Traffic** ($P(z) = 0.10$): This topic includes reports of traffic on the west of the city, in the area that extends between Ngong road and Langata road and towards the Ongate Rongai, an informal urban settlement that lies south of Nairobi.

**Topic 7 – Southeast Traffic** ($P(z) = 0.09$): This topic centers around Jogoo road, Mombasa road, extend including T-mall (south CBD), Mbagathi way and City stadium.

**Topic 8 – Northeast Traffic** ($P(z) = 0.08$): This topic covers traffic updates and jams in areas surrounding Thika road, Kiambu road, Pangani area and Multhaiga area, and Juja road.

**Topic 9 – CBD Traffic** ($P(z) = 0.08$): This topic covers traffic updates on the central business district encompassing Uhuru highway, Haile Selassie and Moi avenue, covering important landmarks such as the City Hall and the Parliament. It includes other derivative roads such as Bunyala, Lusaka, and Valley road, in addition to Mombasa road (continuation of Uhuru highway in the south.)

**Topic 10 – NorthWest Traffic** ($P(z) = 0.08$): This topic includes areas that expands Wayaki way before it converges into Uhuru highway, towards Kangemi (an informal urban settlement in the outskirts of Nairobi) and Limuru (a satellite town north of Nairobi). It also includes Gichuru road, Westland residential district and Chiromo Campus (from Nairobi university), and the museum.

**3.4.3 Topics over time.** We also found that topics are not only different on their description based on relevant terms but also on their temporal patterns, as observed from their hourly evolution. Figure 3 shows the aggregate hourly distribution for each topic. All topics peak in the morning and evening rush hours. However, we see that Topics 1 to 5 have less abrupt peaks (more regularly spread from 5AM to 7PM), as opposed to traffic-focused topics (6–10), which show stronger decays in the middle of the day. Some topics like topic 6 peak equally in mornings and evenings, while others, like topics 8 and 9 peak more in the morning and the evening hours, respectively. This suggests different temporal traffic patterns.

**3.4.4 Topic validation.** As an alternative to collecting human annotations to validate the quality of the discovered topics [21], we leveraged the Ma3route categories to compare topics and categories in terms of how many documents are assigned to each topic-category pair. We expect that a well-fitted model would directly capture some of the Ma3route categories, while at the same time we expect broad categories such as General Info to be distributed over multiple topics. We focus on the most frequent categories: General Info, Clear, and Accident, and aggregate other ones into one single Traffic category. Figure 4 summarizes the tweet assignments for each topic-category combination. First, most posts categorised as Accident are indeed captured by our topic 4. This is mainly due to the keywords used to describe these type of events, as opposed to other traffic info. Second, the General Info category is distributed over several topics (1, 2, 3, 5 and 6), which indicates that a refinement of such broad category can be obtained. Finally, Traffic and Clear reports are mainly included in our traffic-focused topics (topics 6 to 10).

In this section, we showed that urban phenomena of interest for multimedia researchers can indeed be studied from social media channels in Global South Cities, and that this blind side could be alleviated if other researchers turned their attention to these regions as sources of relevant research questions. We now move to show the second blind side, this time propagated to machines.

4 BLIND SIDE #2: DEEP VISUAL ANALYSIS

We performed a deep learning-based analysis to explore the kind of visual content shared in the platform. To extract deep learning features from these images, we chose a pre-trained convolutional neural network (CNN) model which uses the GoogLeNet CNN architecture [45] trained on ImageNet. To extract the CNN descriptors for each image, we obtained the final layer class probabilities across all 1,000 ImageNet classes. We begin our analysis by examining the distribution of the most likely ImageNet category assigned to each image. Given the probability distribution across categories, we chose the scene category with the highest probability as the dominant class for each image. We performed the analysis on 9,805 images from the 14K Twitter photos (Table 2). Figure 5 shows the histogram of the top-10 recognized ImageNet categories. Overall, the dominant class distribution exhibits long-tail characteristics, with a total of 544 unique ImageNet categories. Around 40% of the images were assigned to these top-10 scene categories.

As illustrated in Figure 5, the dominant category was minivan with 8% of the images being classified in this category. Most of the top ten dominant categories are associated with medium to heavy-sized vehicles (minivan, taxi, trailer trucks, tow trucks, garbage trucks), which is not surprising given the nature of the traffic content. However, many of the images are misclassified, in particular images classified as racing car or limousine. Manual inspection reveals that these are not race cars or fancy vehicles. More specifically, images classified as racing car correspond to images of cars circulating on dirt roads, cars changing tires on busy roads, and car crashes. Furthermore, images classified as limousine correspond to views taken inside cars or matatus that capture the top of the dashboard of these vehicles. These misclassifications are the result of limited representations of the world encoded in the pretrained CNN models using ImageNet. It has been recently shown that 53%
of ImageNet images come from only two countries (US and UK) [41]. It seems evident that such sample bias towards wealthy world regions are propagated in our results. It is also troubling that these results mimic results reported in the fairness, accountability and transparency research community about how sample biases affect the recognition of individuals belonging to minority groups [20], but expanding the concerns to urban areas in less wealthy countries. While our analysis will continue in the future, this result demonstrates that pretrained CNNs can reveal basic visual content, but also have limitations with respect to using them to examine images from Global South cities given the sample biases towards wealthy regions. This represents the second blind side, given the current state of multimedia research.

5 AN AGENDA FOR FUTURE WORK

The issues that give rise to the double blind side situation discussed here are complex and multifaceted: while some are technological, others go beyond this realm. We conclude by proposing an agenda for future work that includes tackling the following issues:

The socio-economic bias issue. Although the adoption of mobile phones and mobile internet in Global South cities is increasing, this is a reality that varies across countries, from low-income to lower-middle-income to upper-middle-income countries, using the classification system by the World Bank. Twitter in particular and social media in general are restricted to people having internet access, and therefore our research cannot document the transit issues that affect all citizens. Methods to account for this bias have to be developed, avoiding overgeneralizations.

The data volume issue. Compared to large developed cities, cities in the Global South range from cases where data is produced at similar rates (Istanbul, Sao Paulo, or Mexico City), to cities where data volumes are small, especially for the multimedia research community, accustomed to ever larger data volumes. Africa in particular still generates modest volumes of social media data, but this should not preclude the study of the practices of the current users, as done in this paper. On one hand, we showed that the multimedia data already produced provides rich input to analyze urban dynamic phenomena. Moreover, traffic and public transportation problems are common to many Global South cities, facing challenges related to unsafe driving practices, dangerous road conditions, and overcrowded public transportation. Research could explore transfer learning techniques e.g. to learn data collected from one city and applying it to others of similar characteristics.

The machine perception issue. We showed how pretrained CNNs reveal basic attributes of the image content. While this is already useful, a fundamental question has to do with the need to collect and curate multimedia datasets that reflect urban life in Global South countries, and that can be used to learn models that are better tuned to their realities and alleviate biases, in the spirit of the increasing body of research highlighting the issues of diversity and fairness [48]. We envision the creation of the Global South equivalent of the ImageNet and Places datasets. This would immensely help to make progress on machine perception tasks that are culturally sensitive and useful for these cities.

The human-centered, diversity-aware design issue. We need to improve our understanding of how crowdsourced information is used by citizens in the Global South. This requires a mix of quantitative and qualitative research, and complementary data sources. Do citizens change their transit behaviour or their attitudes based on the information made available in these systems? How much trust is there in the authenticity of systems, specially from the perspectives of fairness and accountability as algorithms become more complex? Finally, can we design tools that could amplify the positive impact of a citizen-driven platform to the community? Answering these questions would require qualitative work interviewing citizens and other urban stakeholders to understand their expectations, beliefs, and usage patterns, and integrating this with multimedia analytics. Human-centered and diversity-aware multimedia could find a wealth of relevant research problems here.

Acknowledgments. This work was supported by the SNSF through the Dusk2Dawn project, and by the EC through the H2020 WeNet project. We thank the Ma3Route Twitter community for their inspiring civic engagement.
REFERENCES