

The Arab Spring on *Twitter*: Language Communities in #egypt and #libya

Axel Bruns and Tim Highfield

With over 600 million signed-up accounts worldwide, *Twitter* has become an important space for the coverage and discussion of unfolding world events, from entertainment and sports to natural disasters and political crises. Especially as breaking news emerges, many of the platform's participants begin to use it as a channel for 'ambient news' (Hermida, 2010): a space in which the collaborative efforts of thousands of contributors who share news links, comment on events, and together 'work the story' (Bruns & Highfield, 2012) as it unfolds, serve to highlight the key issues of the day from the total volume of news coverage. This is an example of gatewatching processes (Bruns, 2005) as conducted on a distributed basis by a large number of users, each contributing only in possibly very minute, *ad hoc* ways.

Such processes are further aided by the operation of topical *Twitter* hashtags that are able to bring together a community of interest around identified themes and events and channel the relevant tweets posted by members of that community into a unified stream of updates. Used in this way – and it should be noted that hashtags can perform other functions than as thematic markers, too – hashtags are able to assist in the formation of *ad hoc* publics on *Twitter* (Bruns & Burgess, 2011); these *ad hoc* publics, in turn, represent a subset of the larger issue publics (Habermas, 2006) that exist in relation to specific issues and events across the overall media ecology and in society itself. The *ad hoc* publics also follow Dahlgren's (2009) depiction of issue publics as temporary assemblages that emerge around a given issue, catalysed in part by media coverage but reinforced and maintained by ongoing engagement with the issue and other participants within the public. Such publics are dynamic, with their lifespan depending on the longevity of the issue at hand and the ongoing contributions and interactions between their members; if an issue is resolved or public interest declines, then the issue public will dissolve.

Issue publics may include multiple perspectives, offering different interpretations or contexts for specific issues and events; in particular, international issues may be discussed in a number of nation- or region-specific issue publics with their own interests, and which may also intersect and overlap with one another. The different perspectives put forward within issue publics might not necessarily interact with one another, though; the single-issue public covering the Arab Spring, for example, may in fact be a number of different publics with their own dynamics and lifespans. These publics may still be connected to one another through the presence of individuals and organisations acting as focal points or bridges between different discussions, yet the patterns of activity within each public may be quite different. Such patterns can be studied on social media platforms such as *Twitter*, as tweeting activity and connections between users can be mapped as networks, highlighting the presence or absence of sub-clusters or isolated groups, and bridging accounts, within a hashtag public. Here, Habermas's (2006) suggestion of a fragmented mass audience in 'isolated issue publics' (423) is not necessarily realised – nor entirely problematic. Although there are reasons behind the fragmentation beyond an individual's political

opinion (including language and region), the development of these mostly-separate discussions can have democratic benefits (Kubitschko, 2011).

Marres (2006), in studying issue networks online, highlights that these discussions can be formed through the promotion of, and interaction between, opposing perspectives around the same issue. On social media, and particularly when studying hashtag publics, these ideas are especially relevant: the hashtag acts as a keyword summarising the central issue or a particular theme, but the surrounding content of tweets using this marker may vary extensively in tone, ideology, and intent. This is especially apparent in politically-oriented hashtags, from ongoing political discussions (the long-standing #auspol hashtag for day-to-day coverage of Australian politics, for instance, is a highly polarised space, characterised by antagonistic comments rather than considered debate) to live responses to media events such as election debates and television panel shows, where the central hashtag may also be promoted by broadcasters for use in their programming. However, these general patterns are not limited to political discussions, and may also be witnessed across contexts from crisis communication and collective action to uses of social media as a backchannel for entertainment broadcasts. The focal function of the hashtag as a collecting mechanism, acting as a central marker for topical information, serves organisational purposes – by encouraging widespread use of a single term to denote related content – and can also contribute to discussions becoming trending topics with increased attention.

The structure of such issue and *ad hoc* publics as they are reflected in social media spaces such as *Twitter* is complicated further where the object of interest shared by such publics spans multiple national or regional public spheres and crosses language, ethnic, religious, or similar identity boundaries. Where this is the case, the same hashtag may play host to highly divergent groups of participants, each engaged in their own processes of gatewatching and ambient news coverage, and variously clashing, interweaving, or cooperating in their efforts to work the unfolding story. Necessarily, such processes are observable most often for major events which draw the attention of a worldwide audience, and – where sharing the same hashtag space becomes unworkable – may lead to a forking of the original hashtag into multiple alternatives designed to accommodate different subsets of the overall *ad hoc* public for the event. Alternatively, the different communities of users populating the same hashtag may seek to develop a mutually beneficial working relationship (with self-selecting participants taking it upon themselves to act as bridges between different language groups, for example), or the attrition of minority users unable or unwilling to engage with the content posted by the majority of contributors may lead to a gradual homogenisation of the hashtag community around the traits, interests and attitudes shared by the majority group.

This chapter explores these processes in the context of the *Twitter* coverage of the Arab Spring uprisings in Egypt and Libya during 2011, with particular attention on the interactions between Arabic- and English-language participants. We build on a computational evaluation of *Twitter* archives of the #egypt and #libya hashtags, which – due to technical limitations to the hashtag system at the time – were adopted by both language groups in spite of their English-language names. Hashtags using the two countries' names in Arabic script could not be used effectively at the time. By distinguishing between Arabic-, English-, and mixed-language participants in these hashtags, we are able to examine the relative prominence of these different language groups within the hashtag communities at any one point, and to chart the changing balance between the groups as the two countries moved through various stages of

conflict (from protests through fighting to regime change and its aftermath). Further, we also specifically identify the language preferences of the lead user groups in each hashtag over time.

Methodology

The data upon which this chapter is based were gathered using a modified version of the open-source tool *yourTwapperkeeper* (Bruns, 2011), which connects to the Application Programming Interface (API) provided by *Twitter* to facilitate programmatic access to public user activity on the platform. *yourTwapperkeeper* (*yTK*) connects to the streaming API to gather, in close to real time, any tweets which match a number of user-defined keywords: in the present case, these keywords were ‘#egypt’ and ‘#libya’, the two most prominent hashtags used to discuss unfolding events in the two countries during 2011. It delivers these tweets and associated metadata – such as the name of the originating user, the exact timestamp of the tweet, and various other information – in a tabular form which allows further processing with a range of computational tools.

In doing so, *yTK* is inherently limited by the fundamental features of the *Twitter* API itself: first (and, in terms of research ethics, most importantly), the API delivers only *public* tweets – that is, tweets posted from accounts which are publicly visible on the platform and which could be accessed even by non-registered visitors to *Twitter.com* itself who happened to search for a relevant term or viewed the profile page of the originating user. The API does not deliver tweets from accounts which have been set to ‘private’ (that is, whose tweets are only visible to followers who have been explicitly approved *as* followers by the account owner), nor direct messages between users or other non-public, internal *Twitter* information.

Additionally, the public API – as opposed to more comprehensive services offered by *Twitter* itself or its approved third-party data resellers – is inherently throttled at one per cent of the total current volume of tweets on the global *Twitter* platform: for example, if the global *Twitter* userbase were posting one million tweets per minute at present, the API would provide a maximum of 10,000 tweets per minute at the same time, even if global activity around the search terms tracked by *yTK* was higher than that limit (cf. González-Bailón, Wang, & Rivero, 2012; Morstatter, Pfeffer, Liu, & Carley, 2013). For the most part, this limitation remains theoretical as the total global volume of tweets is usually several orders of magnitude higher than the volume of any one current hashtag or keyword; in the case of major world events which attract the attention of a substantial subset of the total *Twitter* userbase, however, it is possible for *yTK* to reach such limits. In the context of the present case, such throttling limits may have been triggered for the #egypt and #libya hashtags on days which saw major developments (such as the resignation of President Mubarak in Egypt, or the killing of Colonel Gaddafi in Libya). It is impossible to determine from the *yTK* data alone whether such limits were reached; we simply note, therefore, that the figures reported in this chapter may undercount total hashtag activity on such major days. Unfortunately, these limits could be removed only by drawing on the services of commercial *Twitter* data resellers, at costs well beyond the resources available to most publicly-funded research projects.

Finally, the long-term approach to gathering hashtag data which we have pursued for this chapter has also meant that a number of gaps exist in the datasets due

to API and network outages or scheduled server maintenance. The #egypt dataset is missing tweets for 31 Jan., 5-7 Feb., 31 Mar., 1 and 2 Apr., 2-4 Aug., 15 Sep., 16 Oct., and 23, 26, 27, and 29 Nov. 2011; #libya is missing data for 31 Mar., 1 Apr., 15 Apr., 2-4 Aug., 15 Sep., 16 and 21 Oct., and 26 and 29 Nov. 2011. (Here and throughout the chapter, all dates are given in Cairo time.) This means that for #egypt, we missed 16 days in over ten months of data collection; for #libya, we missed 11 days in nine and a half months. Again, such data outages are hardly avoidable for projects which gather social media data over lengthy periods of time; while they cause gaps in the day-to-day picture of *Twitter* activity around these events, they do not impact significantly on our overall assessment of user participation patterns in the two hashtags.

We captured tweets in the #egypt hashtag from 23 January to 30 November 2011, and in the #libya hashtag from 16 February to 30 November. In total, this resulted in datasets of 7.48 million tweets for #egypt, posted by some 445,000 unique contributors, and 5.27 million tweets from 476,000 unique contributors for #libya. Notably, therefore, the larger #libya community created a smaller number of tweets, indicating the presence of a greater number of less active users for this hashtag. Such greater breadth of participation, to the detriment of the depth of repeated engagement, is common where a larger number of users engage in the casual retweeting of hashtagged messages, for example. We explore this further in the discussion which follows.

For the purposes of analysis, the raw *Twitter* data were processed over several stages. First, we followed the methodology outlined in Bruns and Stieglitz (2012) to distinguish between different types of tweet for each message: original tweets (which neither @mention nor retweet other users); genuine @replies (which @mention another user but are not retweets); and retweets (which @mention another user in the process of retweeting their message, using formats such as 'RT @user ...', 'MT @user ...', 'via @user ...' or '"@user ..."').

Further, we divided the total base of actively participating users for each hashtag into three groups, according to the volume of their contributions to the hashtag: we distinguished between the one per cent of most active contributors, designating these the *lead users* in the hashtag public; the next nine per cent of still *highly engaged users*; and the remaining 90 per cent of *least active users*. This follows a common approach to drawing distinctions in online communities which are characterised by a power law distribution of activity – described by Anderson (2006) as a 'long tail' curve: a small number of regular, highly active users contributes the vast majority of content, while a large group of far less active users makes only occasional, random contributions. In addition to these three groups, we also identify a fourth group of purely *passive* users: this group encompasses those *Twitter* accounts which are @mentioned (in genuine @replies or retweets) in the tweets in our datasets, but did not themselves make any direct post to #egypt or #libya.

Finally, and most importantly for the purposes of the following discussion, we developed an approach which enabled us to make an approximate distinction between Arabic and English speakers in the dataset. Due to the complexities inherent in accurately identifying different languages, especially in the very short message format of individual tweets, through automated computational means, we based this distinction instead on the character sets used in each tweet: while English-language tweets can usually be expressed in characters that use ASCII codes between zero and 127, Arabic characters utilise higher-code additions and alternatives to the ASCII system. Building in a grace interval of ten high-ASCII characters for nominally

English-language tweets which contained high-ASCII characters such as curled quotation marks, en- and em-dashes, or accented characters, we therefore coded each tweet containing fewer than eleven high-ASCII characters as using the Latin alphabet, and each tweet with more than ten such characters as using a non-Latin alphabet.

Given the predominance of English and Arabic as the key languages of discussion in #egypt and #libya (which we confirmed through qualitative sampling), this Latin/non-Latin distinction serves as a workable approximation of the English/Arabic distinction especially once the coding of each tweet is aggregated over the large datasets upon which this study is based. We acknowledge that the ‘Latin’ tweets will also contain content in various other European languages, while the ‘non-Latin’ tweets will contain tweets in Chinese, Korean, Japanese, Cyrillic, Farsi, and other characters; however, these false positives are well outweighed by the dominant languages, and will eventually emerge as outliers in the network maps which we present in a later section of this chapter. For the purposes of the present analysis, the Latin/non-Latin shortcut is not only sufficient, but also necessary in order to process these very large *Twitter* datasets without having to rely on techniques that sample the total dataset down to a much smaller, manually codeable size.

From the coding of individual tweets as ‘Latin’ or ‘non-Latin’, then, it becomes possible to calculate the language profile of each active participant in the hashtag public. Here, we introduce a threefold distinction: users for whom at least two thirds of their tweets were coded as ‘Latin’ are designated as English-language users; users for whom at least two thirds of their tweets were coded as ‘non-Latin’ are considered to be Arabic-language users; and users who do not fall into either one of these categories are described as mixed-language users. This distinction enables us to examine the relative presence of these three language groups in the overall datasets, as well as within the three activity groups (lead users, highly engaged users, least active users) and over the timeframes covered by our datasets. By charting the @reply and retweet interactions between the participants in each hashtag, we are also able to examine the extent to which these language groups communicate with each other.

Analysis

The first key aim of this study is to examine how the use of *Twitter* as a medium for the dissemination of information and exchange of opinions in relation to the events of the Arab Spring uprisings in the two countries differs when comparing the period of immediate tension at the point of regime change and the longer-term aftermath of that cataclysmic moment. A working hypothesis in this context is that the replacement of the existing regime (through peaceful or violent means) is a transitional moment which attracts considerable domestic, regional, and international attention in social media spaces, while the lengthy and complicated process of building legitimate new governmental and societal structures to replace the old regime is of less interest to an international audience and is therefore likely to predominantly attract participants who – as citizens or regional neighbours of the affected countries – have a direct stake in the process. In other words, while in each phase of the transition *Twitter* continues to function as an ambient news network which provides a steady stream of updates on current developments, conventional news values apply: the sudden and dramatic moment of regime change will attract a greater and more diverse number of participants who together engage in ‘working the story’ on the social media platform than the slow and difficult project of national renewal.

If correct, we expect this hypothesis to manifest in the various patterns of *Twitter* activity which our methods are able to document. For both cases to be examined here, therefore, we will distinguish between a phase of immediate regime change, and a phase of longer-term transition in the aftermath. For Egypt, we focus on the month of February 2011 for the former, and on the period of 15 June to 15 September 2011 for the latter; for Libya, we compare 16 February to 15 March 2011 and 1 August to 30 September 2011. Given the long aftermath of the overthrow of the established regimes in either country, the latter periods are necessarily chosen somewhat arbitrarily, and periods from 2012 or 2013 could provide equally valuable points of comparison; however, for such more recent periods it would become progressively more difficult to distinguish genuine change in the usage patterns related to the two country hashtags from the effects of general growth in the local and international *Twitter* userbase. Additionally, *Twitter* launched support for Arabic-script hashtags (as well as for hashtags in other right-to-left languages) in March 2012 (Phelan, 2012), meaning that the use of the English-language hashtags #egypt and #libya for tweets which were otherwise in Arabic has generally declined since then.

However, we begin our analysis with a general overview of the *Twitter* activity within the two hashtags during 2011. #egypt (Figure 1) shows a sharp spike in activity during the early days of the period covered by our dataset, on the day of President Mubarak's resignation; this also corresponds to a substantial influx of unique users and a very low percentage of non-Latin tweets on the same day (and in the days immediately following). The obvious explanation for this pattern is the significant world attention paid to the end of the Mubarak regime, generating increased *Twitter* activity especially also from more casual followers of events in Egypt (who may be active during that time predominantly by occasionally retweeting news updates and other tweets which happen to contain the #egypt hashtag, but who are not otherwise committed to following #egypt on an ongoing basis).

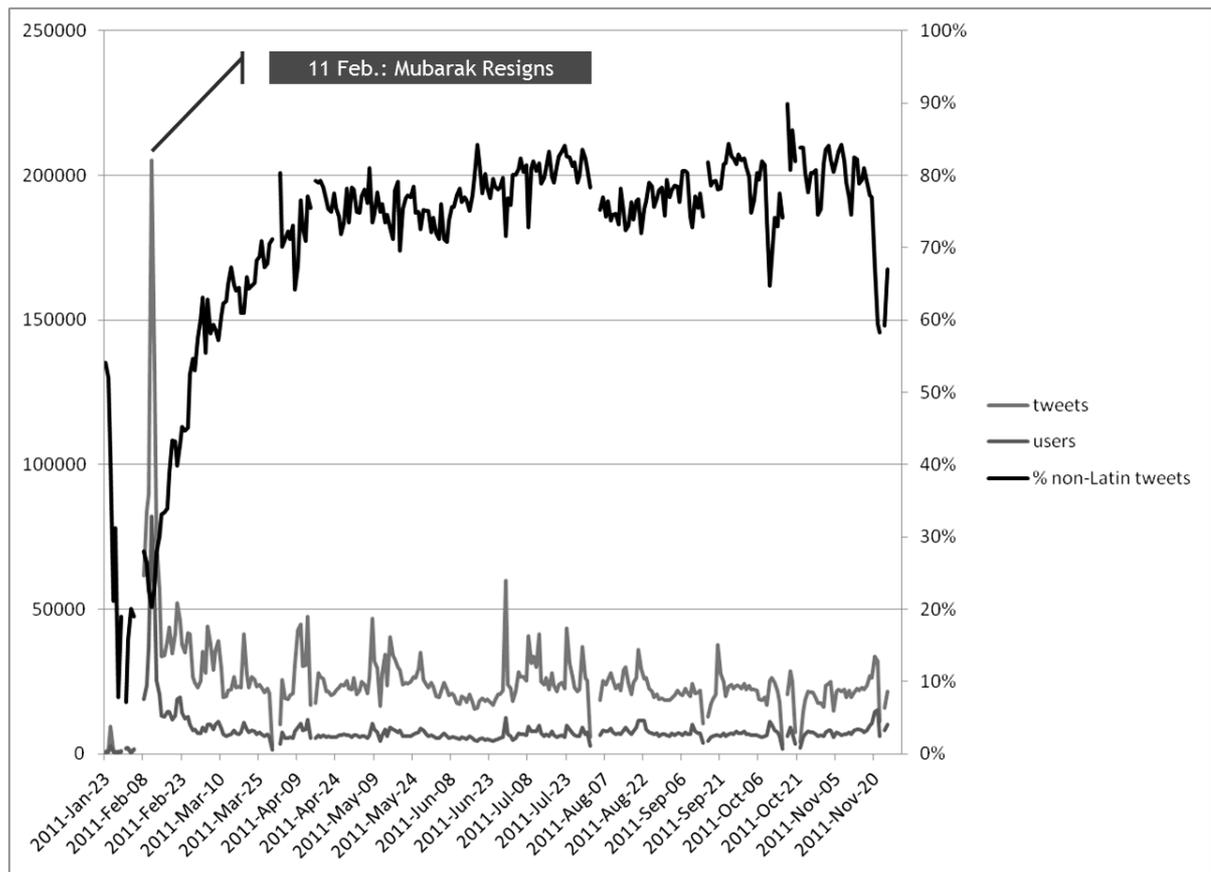


Fig. 1: #egypt, 23 Jan. to 30 Nov. 2011

The comparatively low percentage of non-Latin (i.e. Arabic) tweets during this early phase of the transition in Egypt may appear surprising at first, but is also explained at least in part by such world attention and the presence of more casual, predominantly retweeting non-Arabic participants. Additionally, during late January and early February 2011 a number of alternative hashtags were popular on *Twitter* especially with local Egyptian users, referring for example to the major protests in Tahrir Square on 25 January 2011 (#Jan25 or #25Jan) or to other significant memes of the fledgling revolution. By contrast, international, non-Arabic users with a more limited understanding of local issues were more likely already to use #egypt as their hashtag for discussing the protests.

As 25 January passed and Mubarak resigned, however, #Jan25 and other time- or issue-specific hashtags became increasingly inappropriate even for Egyptian-based users themselves, and a greater adoption of the generic hashtag #egypt evidently began. By early March 2011, more than half of the tweets posted to #egypt per day used Arabic script, and from April onwards an average of 75 per cent of #egypt tweets were in Arabic. This is a significant reversal of the language balance of the early days, and is due both to this greater adoption of #egypt as a common hashtag for Arabic-language discussion of the transition and to the declining interest of a non-Arabic speaking world audience in continuing to discuss the regime change process other than on a handful of key days during the remainder of the year.

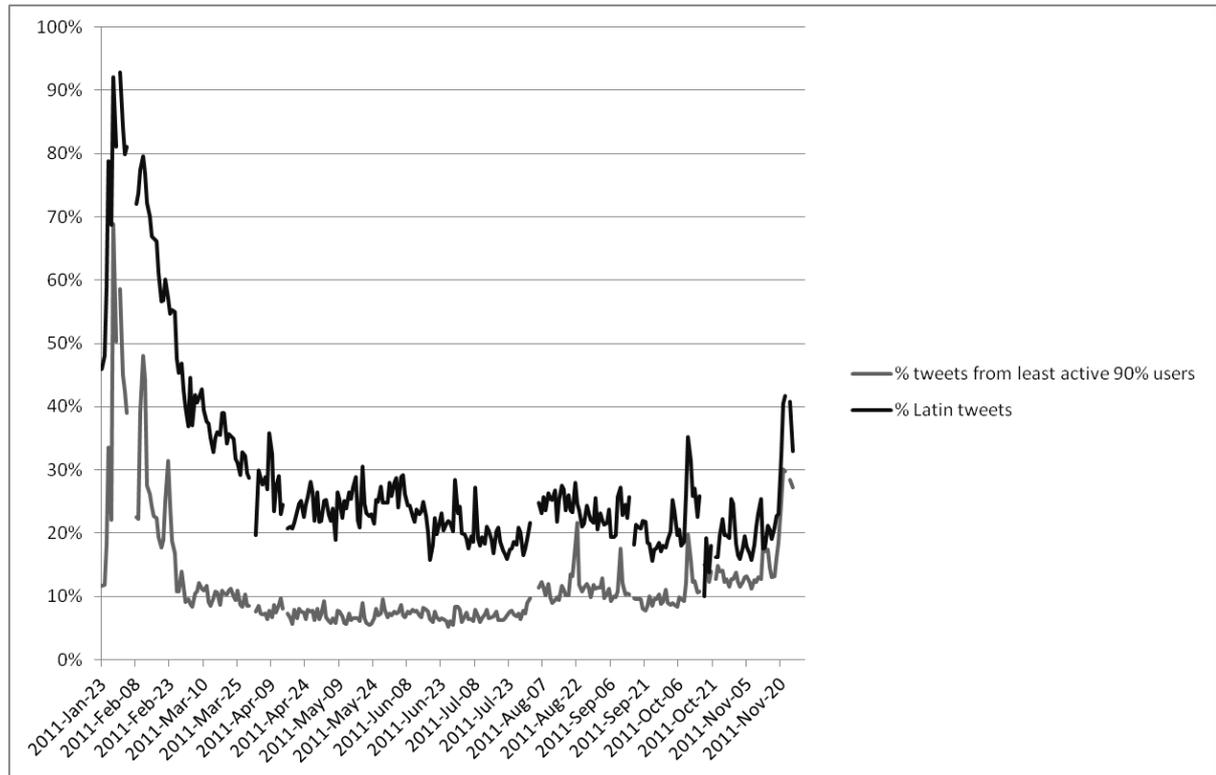


Fig. 2: #egypt, showing Latin tweets and tweets from the least active users

This decline in international interest, and the corresponding takeover of the #egypt hashtag by a smaller number of highly committed Arabic-speaking users, is demonstrated even more clearly in Figure 2. It compares the percentage of tweets made each day which we identified as being in English or other languages using the Latin alphabet, and the percentage of tweets originating from the group of the least active 90 per cent of users (as determined over the entire period covered here, from January to November). The two curves follow each other closely, showing a gradual exodus both of non-Arabic speakers and of less committed participants in the #egypt hashtag after the early, revolutionary months – and thereby suggesting a predominance of English speakers amongst these more casual contributors to #egypt.

The overall patterns exhibited by the #libya dataset are remarkably different (see Figure 3). There is a similar initial spike of interest during the early stages of the revolution, from the first armed conflicts in mid-February through to the UN's declaration of a no-fly zone on 17 March, followed by a much less active period that is punctuated only occasionally by spikes in activity around key dates in the subsequent civil war (such as the storming of Gaddafi's Bab Al-Azizia compound in Tripoli or Gaddafi's death on 20 October). Indeed, the first spike in #libya activity well surpasses that observed for #egypt – pointing perhaps to the comparative absence of rival hashtags for the Libyan case. However, the balance between Latin and non-Latin tweets remains relatively stable throughout the period examined here; while – as in #egypt – the percentage of Arabic tweets is even lower during the first month of the conflict, there is no rapid and decisive rise in contributions from Arabic speakers in the subsequent period, but only a small readjustment of the balance to a steady level of around 20 per cent non-Latin tweets. After Gaddafi's death, this steps up again, to around 30 per cent.

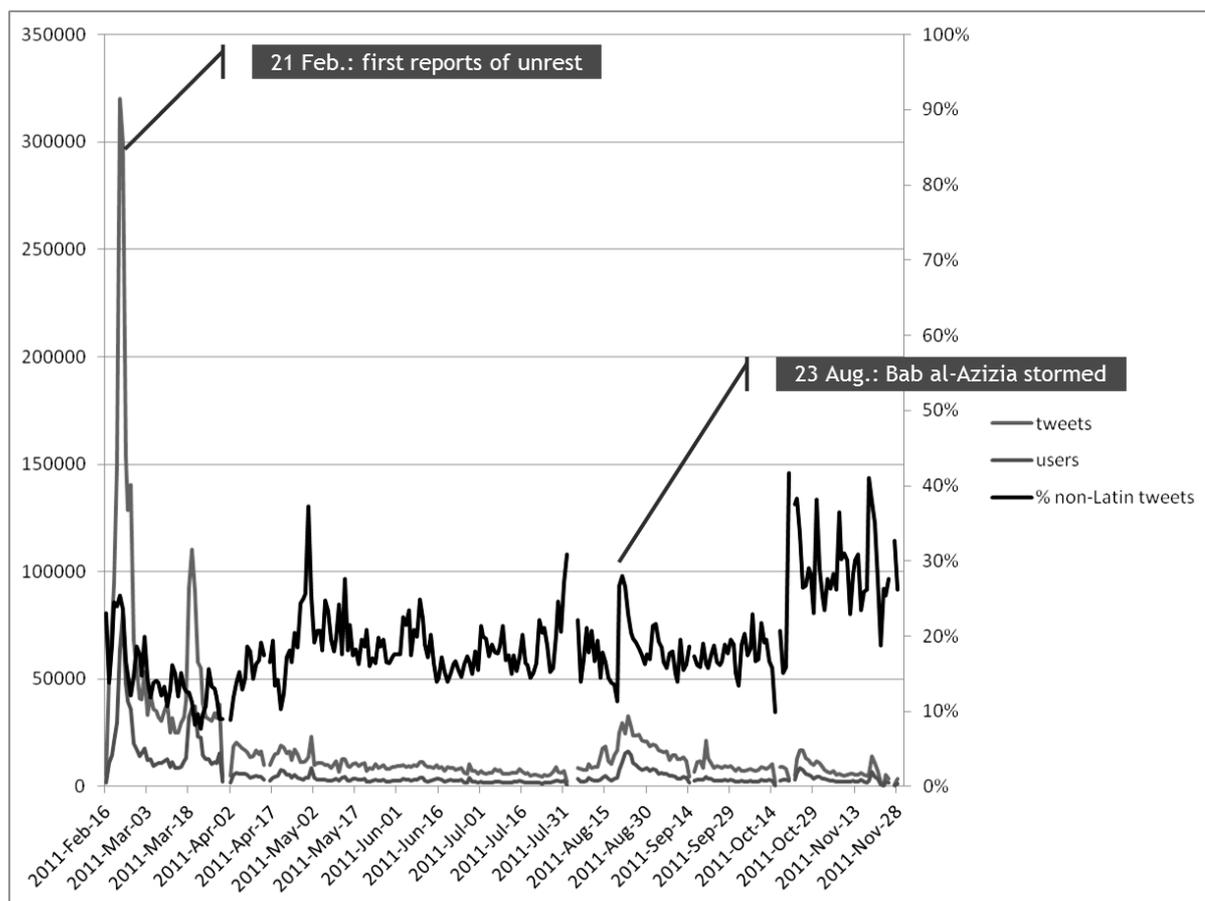


Fig. 3: #libya, 16 Feb. to 30 Nov. 2011

This remarkably low rate of participation by Arabic speakers in the #libya hashtag may be explained by a number of factors. First, compared to Egypt there was considerably less domestic Internet infrastructure available in Libya which would have enabled local users to participate in the *Twitter* discussion, and during its final months the Gaddafi regime took further steps to limit Internet access especially for the residents of restive regions in the country. The available domestic userbase of social media platforms in Libya was therefore considerably smaller than its Egyptian counterpart, and the subset of that userbase who had reliable Internet access was smaller still.

Additionally, Gaddafi's portrayal as the 'Mad Dog of the Middle East' by US President Ronald Reagan had positioned him – alongside dictators such as Saddam Hussein in Iraq or the Kim dynasty in North Korea – as one of the United States' 'favourite' enemies, representing the stereotypical third-world despot. Even in spite of a comparative *rapprochement* between his regime and the West during the first decade of the 21st century, therefore, a revolution directed at overthrowing Gaddafi (and one involving NATO forces imposing a no-fly zone and Western military advisers on the ground) was more likely to attract English-speaking *Twitter* commenters who would participate over the long term than one aiming to overthrow staunch US ally Hosni Mubarak in Egypt.

Examining the identified 'crisis' and 'aftermath' phases for each hashtag, then, enables us to examine how the balance between English and Arabic speakers

evolves over time. Here, we also begin to further utilise the distinction between lead users, highly engaged users, and other contributors which we introduced in the discussion of our methodology, above: not only is the overall shift in the language balance itself of interest, but also the distribution of languages across the three contributor groups in each time period.

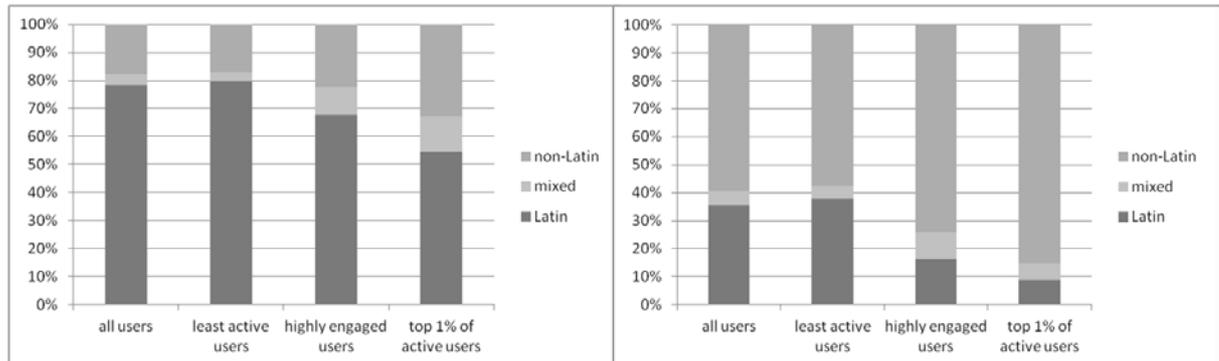


Fig. 4: #egypt language balance, 1-28 Feb. 2011 and 15 June to 15 Sep. 2011

Figure 4 compares the language balance between the two key periods chosen for the #egypt hashtag, and across the contributor groups. First, it demonstrates again the substantial shift in language balance, from an early dominance of English-language contributors (with over 75 per cent of users in the first period tweeting mainly in English) to a mainly Arabic-speaking hashtag community (with some 60 per cent of users tweeting mainly in Arabic). During both periods, only a small percentage of the total userbase are genuinely multilingual participants (that is, post more than 33 per cent but fewer than 66 per cent of their tweets in Arabic or English).

If the total participant base in each period is broken down into the three groups of lead users (top 1 per cent), highly active users (next 9 per cent), and least active users (lowest 90 per cent), according to the volume of tweets they posted during that timeframe, it becomes possible to determine the specific language patterns pertaining to each of these groups. During the immediate crisis phase in Egypt, it is notable that while English-language tweets dominate the hashtag, the lead users are already considerably more likely to tweet in Arabic or use a mix of languages – compared to the less than 20 per cent of least active users who tweet in Arabic, some one third of lead users do so, and nearly half of all lead users tweet in Arabic at least some of the time. It is also notable during this period that there is a small but important presence of mixed-language users both in the lead user group and amongst the highly active users; given the combination of predominantly English- and predominantly Arabic-language participants in these top contributor groups, such bilingual participants are likely to play a crucial role in facilitating the flow of information across the language boundary.

The situation changes almost completely during the aftermath phase. Across the entire community of participating users during this time, some one third of users tweet mainly in English; however, these users are mainly found amongst the least active user group. English-language participants constitute only some 15 per cent of the highly engaged group, and less than 10 per cent of the lead users; in these groups, Arabic-language users are dominant. Indeed, as a result of that domination, the role of

mixed-language users is also diminished: fewer members of that group are now found in the lead user group.

This decrease in the bridging role of the mixed-language users also indicates that the level of interaction between the different language groups has declined. If during the crisis phase English- and Arabic-language users were more frequently collaborating (especially with the help of mixed-language participants) in ‘working the story’ of the emerging anti-regime protests, in the aftermath phase we find a greater disconnect between the English- and Arabic-language subsets of the overall hashtag userbase: two language communities which mainly simply happen to share the same hashtag but no longer interact closely.

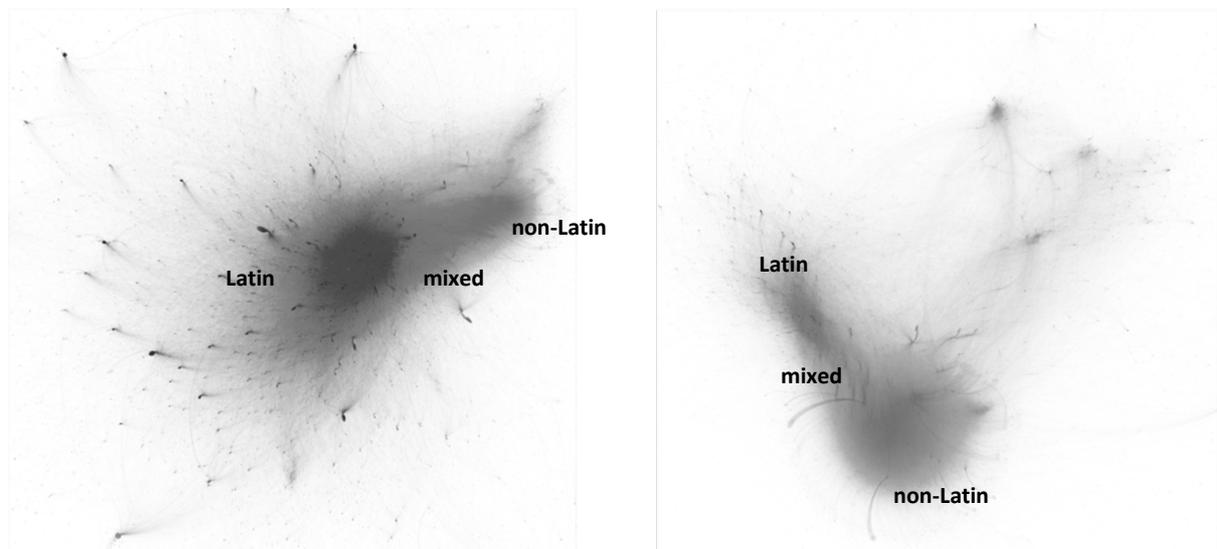


Fig. 5: #egypt @mention networks, 1-28 Feb. 2011 and 15 June to 15 Sep. 2011

A visualisation of @mention networks based on the genuine @replies and retweets across the #egypt hashtag during both periods also demonstrates this (fig. 5)¹. During the early crisis phase, English-language users dominate, as we have seen, but they are also well-connected with their Arabic-language counterparts, not least through the presence of a smaller but important bridging community of mixed-language participants which join the two language groups into one large network. By contrast, during the aftermath phase the English-language group is both diminished in overall numbers and less closely interconnected with the majority Arabic-language group; the latter has contracted to form a more self-centred network with only peripheral links to the English speakers (as well as to a number of smaller communities of non-Latin character users, potentially including users of languages other than Arabic).

¹ Full colour versions of these network graphs can be found online at <http://mappingonlinepublics.net/2012/06/23/some-brief-updates-twitter-and-television-arab-spring-symposium/>

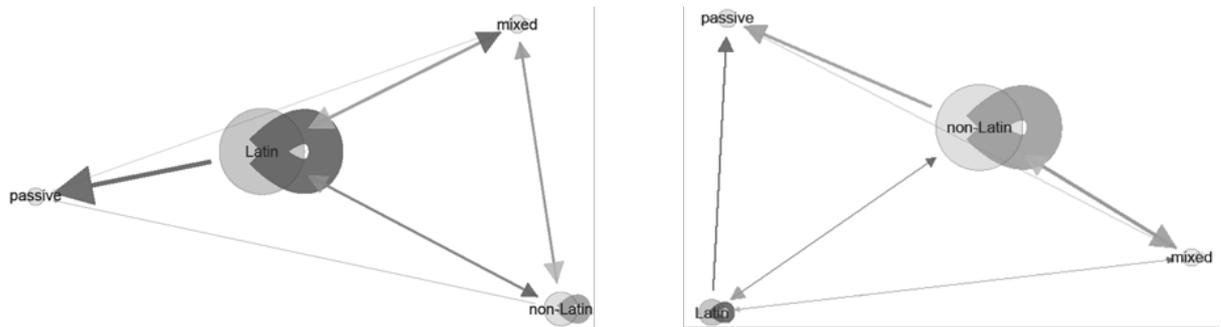


Fig. 6: #egypt @mention networks (aggregate), 1-28 Feb. 2011 and 15 June to 15 Sep. 2011

A simplified depiction of the @mention networks between the three language groups (and non-tweeting, passive accounts which are cited by them), aggregated for each group, also shows this (Figure 6). During each phase, predictably, the majority of @mention interaction for the respective dominant group is with itself (indicated by a horseshoe-shaped self-referential arrow); however, during the immediate crisis phase, the then-dominant Latin users also engage to considerable extent with the smaller mixed group, and these mixed-language users in turn also engage strongly with the non-Latin user group. There is even some significant direct interaction between Latin and non-Latin users, indicating that while most of the Arabic-speakers, for example, may prefer to tweet predominantly in Arabic, many of them are also able to understand and engage with users who only use English in their tweets.

The reverse is less true, however. During the second, aftermath phase, Arabic speakers dominate the #egypt hashtag, continuing to interact with mixed-language participants fairly frequently, but their interactions with English speakers (and especially the @mentions of Arabic-speaking participants by English speakers) have declined considerably. Similarly, the mixed-language group continues to @mention English speakers frequently, but the favour is rarely returned. The English-language group, it appears, has become isolated within a hashtag populated by *Twitter* users whose language it does not understand.

Finally, the role of the passive group of @mentioned accounts also shifts. What unites English- and Arabic-language users during the aftermath phase, and what likely explains the continued presence of both in a hashtag that is divided on language lines, is that both groups refer to passive accounts as outside sources which do not themselves tweet into the #egypt hashtag, but whose tweets may be marked as pertaining to the situation in Egypt as active #egypt participants retweet them. During the earlier crisis phase, on the other hand, only the English-language users refer to a significant extent to such outside sources – and thereby serve as conduits through which such outside information may reach Arabic-language users themselves. This points to a shift in the role of outside reporting on the transition: where earlier, locals may have been engaged in first-hand information exchange through alternative channels while outsiders had to rely on media reporting, now such reporting is readily available to all parties.

Additionally, across both countries and timeframes, some non-participating passive accounts may also be shared across all language groups if they represent politicians, activists, parties and organisations which are relevant to the unfolding conflicts and are mentioned as such by active users (from @UN through @BarackObama to @ArabLeagueSec), but which are not themselves participating in the hashtag exchange. It is now common practice for *Twitter* users to refer to well-

known people and institutions by their *Twitter* handles rather than their full names, without an expectation of receiving responses from them; such usage transcends language barriers and will lead to a considerable overlap in the passive accounts addressed by different language communities.

The #libya hashtag does not see such dramatic changes across the two periods we examine here. As noted above, the English-language group remains dominant in #libya throughout 2011, even if the level of its dominance declines slightly: as Figure 7 demonstrates, more than 80 per cent of all users during the crisis phase tweet predominantly in English, while by the aftermath phase that proportion has declined to just over two thirds.

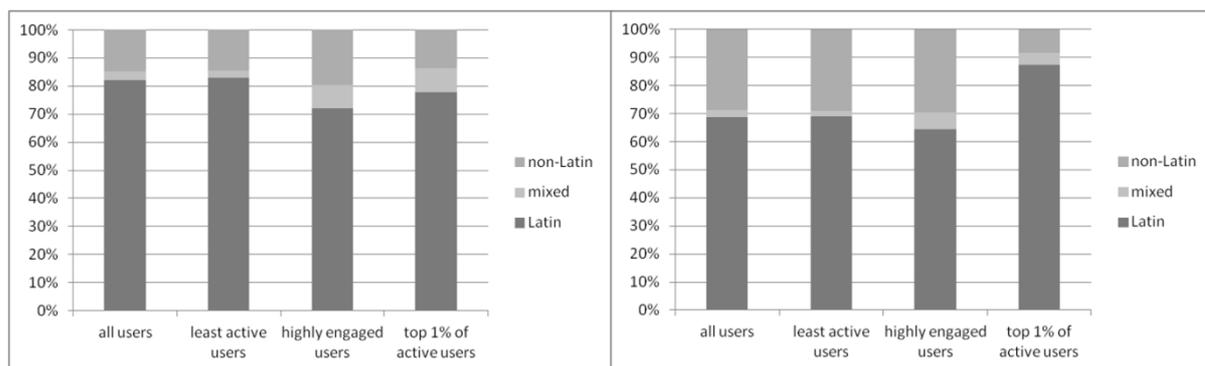


Fig. 7: #libya language balance, 16 Feb. to 15 Mar. 2011 and 1 Aug. to 30 Sep. 2011

At least during the early crisis phase, this dominance is also stable across the different user groups once the total userbase is split according to user activity. There is no significant difference between the percentage of English-only users amongst the least and most active users; the more active user groups exhibit a somewhat greater level of participation from mixed-language users, but (contrary to #egypt) Arabic-only users are even less represented amongst the lead users than they are amongst the least active user group. This paints a picture of a hashtag community which is very strongly dominated by English-language users, then.

During the second phase, Arabic users gain a stronger presence in all aspects, except for the top per cent of users active during this timeframe – indeed, this group of lead users is even more predominantly English-speaking than it had been during the previous phase. The presence of mixed-language users – both in general, and amongst the top user groups – has also declined further. This points to a strongly bifurcated userbase, where the majority of the discussion – from the lead users on down – is conducted by English speakers, while a smaller group of Arabic speakers is generally less active than its non-Arabic counterparts, and is poorly interconnected with them.

A visualisation of @mention networks between #libya participants, in analogy to that for #egypt, bears this out (see Figure 8). English speakers dominate throughout, but during the earlier, crisis phase, the small community of predominantly Arabic-speaking users remains reasonably well-connected to the English-language mainstream both directly and with the help of mixed-language users serving to bridge the language divide. During the second phase, however, the two communities have separated. The community of users tweeting predominantly in Latin characters is strongly self-centred, with a smaller cluster of users at some distance (possibly indicating the presence of a second language community using Latin characters, such

as French, German or Spanish), while the smaller group of non-Latin users remains at a distance and is only loosely connected through a handful of direct links and mixed-language users. It is also internally divided, indicating in this case less a further subdivision caused by different languages using non-Latin scripts, but a lack of cohesion even amongst Arabic-speaking participants in the #libya hashtag.

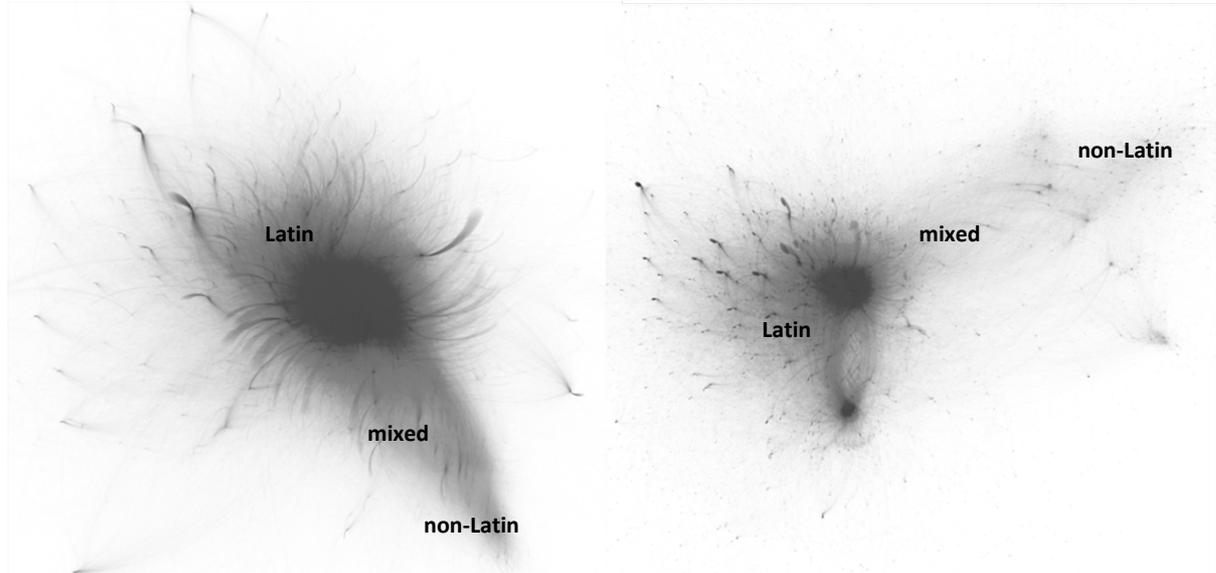


Fig. 8: #libya @mention networks, 16 Feb. to 15 Mar. 2011 and 1 Aug. to 30 Sep. 2011

Figure 9 further explores this by again showing aggregate networks between the three language groups and the group of passive accounts. While here, too, the domination of English-language accounts is inescapable, during the first phase some notable interaction between the English- and mixed-language users, and even directly between English- and Arabic-language users does remain. During the second phase, such interactions are increasingly scant: the mixed-language group is much diminished in size, and English- and Arabic-language participants hardly engage with each other directly any more. During both phases, the attention of English-language participants in #libya – where it does not remain within the English-speaking group itself – is predominantly directed to the outside, passive sources rather than to either of the other language groups, substantially more so than we have observed in the case of #egypt.

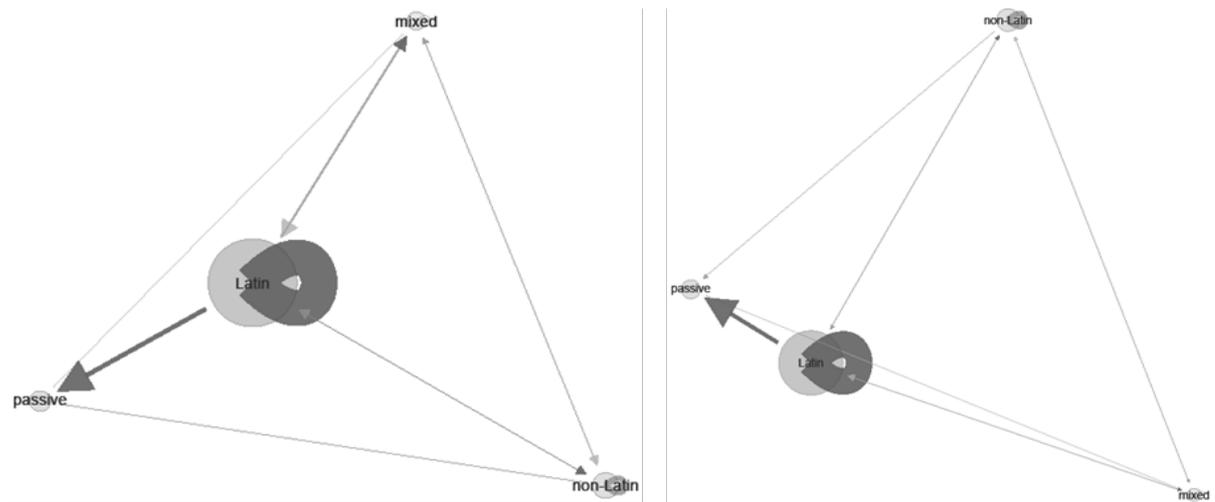


Fig. 9: #libya @mention networks (aggregate), 16 Feb. to 15 Mar. 2011 and 1 Aug. to 30 Sep. 2011

Again, it is likely that these differences between #egypt and #libya are due in part to domestic factors – a lower take-up of social media in general and *Twitter* in particular, poorer overall Internet infrastructure, and more state censorship of Internet traffic in Libya compared to Egypt – and in part to broader circumstances – such as Egypt’s role as a lynchpin in Middle East power structures and international relations, compared to Libya’s status under the Gaddafi regime as a pariah state and long-standing enemy of the United States. In combination, these factors will have resulted in different potentials for the development of a strong domestic social media user base, and for sustaining the interest of onlookers from further afield in participating in the hashtag discussion.

Conclusion

The processes observed here across the #egypt and #libya hashtags show that – in spite the superficial similarities between the two hashtags, as part of *Twitter*’s coverage of Arab Spring uprisings in a number of North African and Middle Eastern nations – the formation and development of international issue publics around major political events can follow some very different paths, and result in vastly different constellations of participants.

In both cases, we observed a rapid early spike in interest by *Twitter* users as rumours and news of protests, unrest, and uprisings spread through the social media platform. Given the remaining predominance of English and other European languages on *Twitter* at the time and now, and the continuing role of English as a global *lingua franca*, we saw the predominance of English as the language of engagement during these early days, and a brief but substantial influx of more casually interested users who were attracted to each hashtag as a locus of significant current *Twitter* activity (and who participated in such activity at least peripherally, especially by retweeting already-hashtagged messages), but we also documented the similarly rapid dispersal of such comparatively inactive participants, leaving the longer-term lead users to contribute the vast majority of tweets to the hashtag. Having

started as *ad hoc* publics (Bruns & Burgess, 2011), the hashtags now transformed into longer-term vehicles for smaller, longer-term publics to engage and interact.

Subsequent to these early days of intense interest in the conflicts in both nations, in both social media and mainstream media, the two issue publics represented by these hashtags diverged markedly in their trajectories. #egypt became a space which supported a domestic and regional issue public tracking continuing developments in the country's transition, at the time, towards free and democratic elections – an issue public whose interactions were predominantly in Arabic, with a smaller number of mainly English-language participants also participating and remaining linked loosely but notably to the Arabic mainstream of the community. #libya, by contrast, hosted a mainly English-speaking community which discussed events in that country's continuing civil war from the outside and which contained few domestic or regional, predominantly Arabic-speaking voices – or where it did, such voices remained marginal and disconnected from the English-language centre. If domestic and regional issue publics which were concerned with the situation in Libya at the time existed, they did so elsewhere – using hashtags other than #libya on *Twitter*, or indeed (and more likely) using channels of communication other than social media.

Even in spite of the limitations of existing tools for the automatic detection and analysis of the language and location distribution in large *Twitter* datasets, and of the limited availability of previous studies which provide comparison data from different contexts, our findings already serve to demonstrate the changeable (over time) and variable (from context to context) nature of hashtag publics and of the wider issue publics of which they form part, especially in the context of major international events and crises. Our work here has been able to establish only the broad patterns of interaction within the Arab Spring Twittersphere; significant further development of tools and methodologies for the study of *Twitter* will be required to enable a finer-grained analysis of the dynamics of these issue publics as they emerged and developed. Substantial further research is also required to document a broader range of such cases and examine the different possible roles which *Twitter*-based hashtag publics, in particular, can play in the broader national and international public sphere.

Such work must also take into account the continuing transformation of *Twitter* and other social media platforms themselves, of course, as their userbases grow and diversify (or possibly shrink and solidify), as well as the shifting technological affordances of these platforms as communication features are added or removed and as the devices and networks used to access such social media platforms themselves continue to evolve. This is especially important in the context of still-unfolding processes such as the Arab Spring and its consequences, which have also resulted in a considerable change in the communicative opportunities now available to citizens of Egypt, Libya, and other countries in the region.

References

Anderson, C. (2006) *The Long Tail: How Endless Choice Is Creating Unlimited*

Demand, Random House Business Books: London.

Bruns, A. (2005) *Gatewatching: Collaborative Online News Production*, Peter Lang: New York.

Bruns, A. (2011, June 21) Switching from *Twapperkeeper* to *yourTwapperkeeper*. *Mapping Online Publics*. Retrieved from <http://www.mappingonlinepublics.net/2011/06/21/switching-from-twapperkeeper-to-yourtwapperkeeper/>

Bruns, A., & Burgess, J. (2011) 'The use of twitter hashtags in the formation of ad hoc publics', in *European Consortium for Political Research conference, Reykjavík, 25-27 Aug. 2011*. Reykjavík. Retrieved from <http://eprints.qut.edu.au/46515/>

Bruns, A., & Highfield, T. (2012) 'Blogs, Twitter, and Breaking News: The Producers of Citizen Journalism', in R. A. Lind (Ed.), *Producing Theory in a Digital World: The Intersection of Audiences and Production* (pp. 15–32), Peter Lang: New York.

Bruns, A., & Stieglitz, S. (2013) 'Towards More Systematic Twitter Analysis: Metrics for Tweeting Activities', in *International Journal of Social Research Methodology*, 16(2), 91–108. doi:10.1080/13645579.2012.756095

Dahlgren, P. (2009) *Media and political engagement: Citizens, communication, and democracy*, Cambridge University Press: New York, NY.

González-Bailón, S., Wang, N., & Rivero, A. (2012) 'Assessing the Bias in Communication Networks Sampled from Twitter Assessing the Bias in Communication Networks Sampled from Twitter'. Retrieved from <http://dx.doi.org/10.2139/ssrn.2185134>

- Habermas, J. (2006) 'Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research', in *Communication Theory* 16(4): 411-26.
- Hermida, A. (2010) 'From TV to Twitter: How Ambient News Became Ambient Journalism', in *M/C Journal*,13(2). Retrieved from <http://journal.media-culture.org.au/index.php/mcjournal/article/view/220>
- Kubitschko, S. (2011) 'Networked Civic Life: Issue Publics during Federal Elections in Germany', in *Communication, Politics & Culture*, 44(2), 112–132.
- Marres, N. (2006) 'Net-work is format work: Issue networks and the sites of civil society politics', in J. Dean, J. W. Anderson, & G. Lovink (Eds.), *Reformatting politics: Information technology and global civil society* (pp. 3–17). CRC Press: Hoboken, NJ.
- Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013) 'Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose', paper at International Conference on Weblogs and Social Media, Boston, July 2013. Retrieved from <http://www.public.asu.edu/~fmorstat/paperpdfs/icwsm2013.pdf>
- Phelan, J. (2012) 'Twitter launches in Arabic, Farsi, Hebrew, Urdu', in *Global Post*, 7 Mar. 2012. <http://www.globalpost.com/dispatch/news/business-tech/technology-news/120307/twitter-launches-arabic-farsi-hebrew-urdu>