When countries become the talking point in microblogs: Study on country hashtags in Twitter

Sesagiri Raamkumar, Aravind; Pang, Natalie; Foo, Schubert

2016

Sesagiri Raamkumar, A., Pang, N., & Foo, S. (2016). When countries become the talking point in microblogs: Study on country hashtags in Twitter. *First Monday*, 21(1).

<https://hdl.handle.net/10356/82016>

<https://doi.org/10.5210/fm.v21i1.6101>

© 2016 First Monday. This paper was published in First Monday and is made available as an electronic reprint (preprint) with permission of First Monday. The published version is available at: [<http://dx.doi.org/10.5210/fm.v21i1.6101>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 23:13:14 SGT

When countries become the talking point in microblogs: Study on country hashtags in Twitter

by Aravind Sesagiri Raamkumar, Natalie Pang, and Schubert Foo

Abstract

Hashtags are placeholder features for capturing the underlying themes in microblog posts. Prior studies have investigated the conversation dynamics, interplay with other media platforms and communication patterns between users for specific event-based hashtags. Commonplace hashtags have been largely ignored, albeit the utility of these hashtags is the main reason behind their continued usage. This study aims to understand the rationale behind the usage of a particular type of commonplace hashtags, namely country hashtags. Manual tweet classification was performed on twitter extracts, to identify the themes of tweets containing three country hashtags. Eleven categories were identified with varied rankings conditioned by factors such as national interest, tourist attractions and cross-media sharing, across the three countries. Network analysis was employed to identify the underlying network types. Broadcast networks and tight crowd networks were identified as the prominent types. Findings will inform subsequent studies on national topics oriented discussions in social media.

Contents

- [1. Introduction](#)
- [2. Related work](#)
- [3. Methodology](#)
- [4. Findings](#)
- [5. Discussion](#)
- [6. Limitations](#)
- [7. Conclusion and future work](#)

1. Introduction

Twitter has become one of the most popular online social networks (OSN) in recent times (Zhao and Rosson, 2009). It has popularized the concept of micro-blogging (Java, *et al.*, 2007) and brought about its relevance in both interpersonal and public communication spheres. Twitter, a microblogging Web site, was launched in 2006. It has been the focus of research studies from 2008 (Krishnamurthy, *et al.*, 2008). People use tweets for multiple purposes such as status updates, conversations with other users, endorsing opinions ('retweet' and 'favorite' options), promotions and spamming (Benevenuto, *et al.*, 2010). Hashtags (words starting with '#' symbol) are used in twitter as a placeholder with multiple purposes. Due to the 140 character limit in tweets, it is important to have an indicator in the tweet to show its representativeness to an idea or concept. Hashtags are prominently used to show the inclusive nature of tweets to a particular topic of conversation. For instance, the hashtag #Occupy was used by Twitter users during the famous Occupy Movement in 2011 (Orcutt, 2011). Hashtags can refer to places such as country names (e.g., #singapore, #india) used on a daily basis. Hashtags are also used as reference to people (e.g., #Obama, #stevejobs). Apart from aiding users in conversation, hashtags are used to indicate the main theme of a tweet. For example, the hashtag #review is used in tweets indicating that the tweet text is related to a movie review. Therefore, hashtags are ideal candidates for indexing so as to speed up information retrieval.

Hashtag studies have taken two approaches, by either concentrating on event-oriented hashtags (Lin, *et al.*, 2012) or they have used a hashtag agnostic approach where a random extract of Twitter data is used for analysis with no particular focus on specific hashtags (Pöschko, 2011). Commonplace trending hashtags have not been studied extensively. General discussions in Twitter are about common topics such as politics, entertainment, sports and local events (Duan, *et al.*, 2012). These topics shape the discussion at both national and international level with the former being more prevalent. Therefore, it would be interesting to study the rationale behind the usage of

country and city hashtags and their contribution to conversations at a holistic level in microblogs. It is observed that some of the popular and frequently used hashtags refer to location names and people names (Trendinalia, 2015). Particularly, a location name can refer to a particular site, town, city or a country. Country hashtags are most often used, to associate a tweet with a country.

This study employed an explorative approach to understand the dynamics around country hashtags at both content and structural levels by using content analysis based classification of tweets and network analysis techniques. Three Asia related hashtags, #singapore, #india and #indonesia, were selected for this study. Tweets containing these hashtags were extracted for a week's duration for conducting the analyses. The contributions of this study can be summarized as follows: Eleven categories are identified through manual tweet classification, pertaining to the tweets' content. Out of the 11 categories, four categories are novel categories that provide the potential to be used in subsequent studies. The prevalent network types have been identified as broadcast and tight crowd networks for the three hashtags. These network types corroborate with the allocation percentage of the categories for the three hashtags, thereby validating the results of the classification exercise.

2. Related work

2.1. Research agendas in Twitter

Twitter was launched in 2006 as a microblogging platform that facilitated users in sharing and consuming information about day-to-day happenings and opinions on topics. It was a unique product during the time of its introduction due to its character limit on user posts (users are allowed to post tweet messages within 140 characters limit). Twitter has become immensely popular (Rank 9 in Alexa Web Rankings (Alexa, 2015)) which has to lead to regional spinoffs, such as Sina Weibo. Noticing the dynamics around the interactions in Twitter, academic research in Twitter started in 2008 (Krishnamurthy, *et al.*, 2008). Twitter research has been surveyed and summarized in (Cheong and Lee, 2010; Cheong and Ray, 2011; Williams, *et al.*, 2013). Research has furthered in different directions with varied focus such as organizing information (Sriram, *et al.*, 2010), understanding trends and convergence events from a communications perspective (Lin, *et al.*, 2012), usage of Twitter data in practical applications (e.g., governments, activism) (Bruns and Burgess, 2011), cross-application of Twitter data (in cross-platform recommendations (Abel, *et al.*, 2013)) and traditional computer science oriented focus on information retrieval (Magnani, *et al.*, 2011) and semantics (Abel, *et al.*, 2011). In the context of information organisation, tweet categories are identified using both manual content analysis based classification and machine learning classification. The studies in tweet classification are discussed in the next sub-section.

2.2. Tweet classification

The two high level entities in Twitter are user and message (Cheong and Lee, 2010). Recent research has introduced two additional entities technology and concept (the central topic being addressed in the tweet) (Williams, *et al.*, 2013). This classification scheme has been used in studying Twitter data. On the topic of tweet classification, past research has identified many categories which differ based on factors such as method of classification, amount of data, period of data and frame of reference. The categories identified by earlier research studies are presented in [Table 1](#) (a & b). Categories such as news, information sharing, events, opinions and promotions appear to be common across the schemes. The variation in schemes is mainly due to the vocabulary used for naming the categories and the purpose of classification. There has been a lack of consolidation across studies, except for the work of Dann (2010) where four earlier classification schemes have been combined to form a new scheme with six generic categories. It is to be noted that all these classification attempts have not used hashtag as the frame of reference, albeit hashtags convey the central themes in the tweet content.

Table 1(a): Examples of classification schemes from previous Twitter studies (2007–2010).

<i>Java, et al.</i> (2007)	<i>Jansen, et al.</i> (2009)	<i>Honeycutt and Herring</i> (2009)	<i>Pear Analytics</i> (2009)	<i>Horn</i> (2010)
Conversations URL sharing News	Info seeking Info providing	About addressee Advertise Exhort Info for others Info for self Meta- commentary Media use	Mainstream news Spam Self- promotion of businesses	C1: News, events, company

reporting Daily chatter	Comment/sentiment	Express opinion Other's experience Self experience Solicit info Other miscellaneous	Babble Conversations Pass-along messages (retweets)	C2: Factual, opinionated
----------------------------	-------------------	---	---	--------------------------------

Table 1(b): Examples of Classification schemes from previous Twitter studies (2010–).

<i>Sriram, et al. (2010)</i>	<i>Dann (2010)</i>	<i>Sandra, et al. (2010)</i>	<i>Naaman, et al. (2010)</i>	<i>Rosa, et al. (2011)</i>	<i>Duan, et al. (2012)</i>	<i>Huang, et al. (2013)</i>
News Opinions Deals Events Private messages	Conversational Pass along News Status Phatic Spam	Movies Books Music Apps Games	Info sharing Self-promotion Opinions, complaints Statements and random thoughts Me now Question to followers Presence maintenance Anecdote	News Sports Science and technology Entertainment Money, Business Just for fun	Entertainment Politics Science and technology Lifestyle Business and products Sports	Complaint Promotion Compliment Mention

2.3. Hashtag studies

Hashtag is a keyword which starts with the symbol '#'. It is mainly used for categorizing content and joining conversations on various topics (Huang, *et al.*, 2010). Hashtags serve the same purpose as tags made famous by Web 2.0 services such as Flickr and Delicious. Even though users need not necessarily add hashtags to their tweets, it is generally observed that regular users add hashtags to most of their tweets. Global political events are represented in twitter through hashtags, some of the popular ones include #occupy, #OWS and #Syria.

The analysis of behavior around hashtags was done as part of an earlier study on conversational tagging (Huang, *et al.*, 2010). The authors used statistical measures such as standard deviation, skewness and kurtosis to study the popularity of hashtags and the scenarios in which hashtags gain traction. Pöschko (2011) performed an exploratory study on hashtags by analyzing 29 million tweets which involved tweet classification, studying hashtag co-occurrences, part-of-speech tagging and SNA based clustering thereby highlighting different ways of dissecting the tweet data to gain insights.

Yang, *et al.* (2012) developed a machine learning model to predict the future adoption of hashtags by users, by combining the two cases under which a hashtag is used by users. The two cases are content organisation and community participation. Measures such as relevance, preference, prestige and influence were used as the main features for the machine learning model. A similar, albeit technically focused, approach by Tsur and Rappoport (2012) used more number of features to predict the spread of ideas (hashtag) in the Twitter environment. Bruns and Burgess (2011) used social network analysis to study the growth and decline of conversations happening around hashtags at different points of time and raise the need for a detailed catalogue to better understand the patterns of interaction. Lin, *et al.* (2012) did a broader study by analyzing 256 hashtags related to the U.S. presidential elections for understanding the growth, survival and context of their usage. They put forth a two-way classification of hashtags with the categories 'Winners' and 'Also-rans' and introduced a theoretical framework to understand the adoption behavior of user-generated content. As seen from these studies, the focus has been largely on event-based hashtags. The dynamics around commonplace hashtags are yet to be explored even though these hashtags shape the discussions in microblogs.

From the previous studies, it can be ascertained that a variety of analysis techniques could be used to study hashtags from user, message and conversation dynamics viewpoints. The works of Yang, *et al.* (2012) and Bruns and Burgess (2011) are important as they could be used in any hashtag study for: 1) Predicting the actual utility of the hashtag on whether it is beneficial in either content organisation or community participation; and 2) Studying the growth and decline of hashtags in a given period of time.

3. Methodology

3.1. Research questions

It is apparent from the earlier studies that hashtags play a focal role in directing conversations in Twitter. Explorative hashtag studies (Pöschko, 2011) have taken a generalized approach by not looking at a particular type of hashtags. In-depth studies on hashtags so far have focused on political events which are of periodic nature (Bruns and Burgess, 2011; Lin, *et al.*, 2012). Therefore, there is a necessity to explore the dynamics around commonplace hashtags that are used on a regular basis. Hashtags which are about a place (location) are quite common trending topics in Twitter. Not much is known about the rationale behind their usage. In this study, hashtags with country name are studied. The overarching research question for the current study is "Why do users make use of the country hashtag?" Specific research questions that are explored are:

RQ1a: What are the categories that represent the tweets containing country hashtags?

RQ1b: Does the new set of categories identified using the country hashtag as a frame of reference differ from the existing tweet classification schemes and why?

RQ2: What are the prevalent network types in user communication data extracted from the tweets?

RQ3: Does the provenance data of the tweets and other tweet statistics provide any new insights?

3.2. Research methods

The exploratory nature of this study demanded the employment of content and structural analysis methods to gain deeper understanding of the data. The two main methods used in the study are content analysis based tweet classification and social network analysis (Wasserman and Faust, 1994). In the tweet classification exercise, categories were identified by keeping the hashtags as the frame of reference. Content analysis was performed on tweets from the manually classified set. Social network analysis (SNA) techniques were used to analyze the user-mentions data extracted from the tweets. In tweets, users can tag other users so as to meaningfully direct the tweet's content. This feature is called as 'user-mentions'. Directed graphs for user-mentions data were built using the visualization tool Gephi. The unconnected nodes were filtered out from the graphs since the focus was to analyze the connectivity between the users and hence the presence of unconnected nodes doesn't provide any contribution to the analysis. A community detection algorithm (Newman, 2006) was run to identify the underlying communities in the graphs. A force atlas algorithm (Bastian, *et al.*, 2009) was used for re-arranging the graph in a readable layout with authoritative nodes clearly differentiated from normal nodes. For generating the graphs, tweets from the original extract were used.

3.3. Data collection

Three country hashtags, #singapore, #india and #indonesia, were chosen as the candidate hashtags for this study. Three hashtags were shortlisted for this study as it is important to validate the results from three different geographic samples. These countries were selected for this study as the intent was to generalize findings for Asian countries. Samples for other geographies will be utilized in future studies. Note that in the case of Singapore, the country and city name are the same. The Twitter data extraction service TweetArchivist was used to extract data for the hashtags #singapore, #india and #indonesia for the period between 26 August and 1 September 2013. Tweets with English as the system language were shortlisted from the original extract. Since the current study employed a manual classification approach, the objective was to have a sample size close to 5,000 tweets for manual classification. This is higher than the average sample size of 2,000–3,000 tweets used in some earlier studies (Dann, 2010; Pear Analytics, 2009; Naaman, *et al.*, 2010), and the sample was developed with tweets selected randomly from the three extracts so that each hashtag gets sufficiently represented in the sample set. [Table 2](#) provides the tweet count for the three hashtags in the Twitter extracts along with the distinct users count in each extract.

Table 2: Statistics for #singapore, #indonesia and #india Twitter extracts.

Note: Column A refers to the number of tweets in TweetArchivist extract with the filter setting *lang=en* and *period=Aug 26'13 to Sep 01'13*. Column B refers to number of final tweets shortlisted for manual classification. Column C refers to number of distinct

users in shortlisted extract (B).			
Hashtag\stat	A	B	C
#singapore	12,303	1,757	980
#india	59,552	1,691	1,022
#indonesia	11,702	1,673	1,369

4. Findings

4.1. RQ1a and RQ1b: What are the categories that represent the tweets and does the new set of categories identified using the country hashtag as a frame of reference differ from the existing tweet classification schemes and why?

Ten categories were identified after the manual classification process. Initially, 23 subcategories were identified during a pilot classification exercise with about 800 tweets. The aggregation of the 23 subcategories to the ten categories was performed before the start of the actual classification exercise, to reduce sparsity in the assignment of categories to the tweets in the sample sets. The 11 categories are *Local Events (LE)*, *Local News (LN)*, *Current Location and Landmarks (CLL)*, *Asia Related (AR)*, *Unrelated*, *Commercial Deals (CD)*, *Tourism and Travel Related (TTR)*, *National Identity (NI)*, *Group Reference (GR)*, *National Group Reference (NGR)* and *Personal Events and Rants (PER)*. These categories have varied allocation percentages across the three samples. [Table 3](#) provides the tweet count and percentage for the categories from the three samples. In [Figure 1](#) (a, b and c), the categories' performance is illustrated with radar charts. Each of these categories is described subsequently along with sample tweets. Twitter account names and personal URLs (Uniform Resource Locators) have been removed for security reasons.

Table 3: Category count and percentages for three hashtag Twitter samples.

Category	#singapore extract		#india extract		#indonesia extract	
	Tweet count	Tweet percentage	Tweet count	Tweet percentage	Tweet count	Tweet percentage
Asia Related (AR)	29	1.65	30	1.77	13	0.78
Commercial Deals (CD)	329	18.73	69	4.08	74	4.43
Current Location and Landmark (CLL)	288	16.39	36	2.13	435	26.02
Group References (GR)	23	1.31	87	5.14	81	4.84
Local Events (LE)	114	6.49	62	3.67	70	4.19
Local News (LN)	312	17.76	663	39.21	127	7.60
National Group References (NGR)	32	1.82	76	4.49	18	1.08
National Identity (NI)	12	0.68	131	7.75	65	3.89
Personal Events and Rants (PER)	349	19.41	219	12.95	536	32.06
Travel and Tourism Related (TTR)	78	4.44	19	1.12	39	2.33

Unrelated	199	11.33	299	17.68	214	12.80
-----------	-----	-------	-----	-------	-----	-------

Figure 1: Radar charts with category counts for the three hashtag extracts. Charts a, b and c are for #singapore, #india and #indonesia respectively.

4.1.1. Asia Related (AR)

This category with a tweet percentage of less than two percent corresponds to the topics related to Asian countries. It mainly covers the tweets that are posted by local news agencies and selective users when the topic is related to Asia. Music celebrities who tour Asian countries while performing live concerts (e.g., Pitbull) use the country hashtag in their tweets to indicate their presence in these countries. In the case of India, which has a contentious cross-border history with some of its neighboring nations, this category finds a presence in tweets when citizens voice their opinions about politics (e.g., Pakistan nationals talk about Indian politics and vice versa).

"Made in India — tags on designer wears and wallets makes you proud! #Singapore"
"Looks like another 1997 Asia crisis on the cards #India"
"... what about #indonesia and #malaysia and the #oramgutans #rhinos #tigers rainforests destroyed"

4.1.2. Commercial Deals (CD)

This category corresponds to tweets posted by commercial bodies with the intention of marketing and promoting their products to their Twitter followers. The activity can be seen as an alternative/compliment to the RSS based push services provided by online portals. This category is quite popular in Singapore (18.73 percent) compared to India (4.08 percent) and Indonesia (4.43 percent). From the sample set, book stores (e.g., sgbookstore, singaporebook) and job portals (e.g., StanChartJobs) use Twitter to push new offerings to the public. The usage of the country hashtag in these tweets is of redundant nature as it can be assumed that the tweets by nature are related to the country's context. One of the main reasons that commercial bodies persist in the continual usage of these hashtags is to leave a digital imprint and capture the mind share of users.

"#Singapore #Books: Encyclopedia of Sleep (Academic Press) — Encyclopedia of Sleep (Academic Press) Edited by..."
"2BHK flat @ 35 lakhs. PS: B in BHK is bathroom. #property #India #Pune"
"Ready stock distortion merch at #firsthandstore #samarinda #limited #indonesia ..."

4.1.3. Current Location and Landmark (CLL)

This category corresponds to the tweets that are about the current location of the user and also references to landmarks in the locality. This category is most popular in both Singapore (16.39

percent) and Indonesia (26.02 percent) compared to India (2.13 percent), mainly due to the higher tourist rankings of the former two countries. Almost all the tweets of this category are posted from Instagram where users share their visual content to Twitter for reaching out to their Twitter followers. This is a case of cross-media sharing done for reaching out to broader set of social media users. The absence of this category from most of the previous classifications is due to its unique association with location based hashtags.

"Photo: Sky walk, #singapore #trip #nature (at Gardens By The Bay) ..."

"Sunrise 🌅 #sunrise #goa #india #sunrays #bridge #drive #instapic #instalikes #nature #amazing ..."

"Half Lombok already Explored, View of Lombok ... #travel #backpacker #travelblog #indonesia #ttot #ITB"

4.1.4. Group Reference (GR)

This category corresponds to the tweets where the users address a group of people. Greeting messages (e.g., good morning wishes) and general messages to fellow countrymen fall under this category. There is a separate subcategory 'National Group References' which will be discussed separately. This category's presence is more significant in India (5.14 percent) and Indonesia (4.84 percent) compared to Singapore (1.31 percent), perhaps indicating the use of Twitter for group messaging in this study's context.

"new to #singapore and looking for some good #startup and #tech events, any pointers?"

"People who pushing for dialogue with #India, even after #LOC killings, are same that want dialogue with #BLA but want operation against #TTP"

"My #family with ustadz sholehmahmoed #latepost #vacation #airport #jakarta #indonesia #ramadhan ..."

4.1.5. National Group References (NGR)

This category corresponds to tweets that are posted by users as references to fellow citizens in order to convey information or opinion related to the image of the country. These are the only set of tweets that are directly addressed to the country from either a geographic or geopolitical viewpoint. This category is dominated with normal user accounts unlike other categories which are mainly represented by group accounts (e.g., personaSingapore, sgbroadcast). Socio-cultural messages fall under this category. This category is famous in India (4.49 percent) where there is consistent discussion about corruption and politics while its presence is low in Indonesia (1.08 percent) and Singapore (1.82 percent).

"#Singapore!! We called this home for a year! Maybe we'll see #SEAsia again when we become ..."

"31 children are born in #India per minute ... and 62 women are raped. Are we still proud Indians?"

"The babies say: take my hand, not my life. #indonesia 2,5 million babies every year scream and Die!! ... #StopAbortion"

4.1.6. Local Events (LE)

This category corresponds to the tweets that are about local events happening around the country. Users tweet about musical events, sports events, festivals, conferences and other gatherings often in Twitter. The country hashtag is not necessary in these tweets, with the exception of sports events. The hashtag is added to indicate the importance of the particular events to the country. Its popularity is higher in Singapore (6.49 percent) compared to Indonesia (4.19 percent) and India (3.67 percent) as the former conducts more international events on a regular basis with users taking advantage of Twitter to indicate their attendance.

"#Singapore's #Formula 1 Grand Prix 2013, Lets go for the Race ..."

"#Exhibition: 'Oriënteren' in #denhaag — August 31st | ViaTerra via #photography #iran #india #easternmysticism"

"Seek n Destroy #metallica #jakarta #indonesia #swag #music #concert #84.000 #crowds #instamood ..."

4.1.7. Local News (LN)

This category corresponds to tweets about news that are mainly posted by news agencies and commercial bodies. It is quite evident that news agencies use the hashtag more than any other type of user (e.g., user accounts india_breaking, personaSingapore, AsiaPacNews to name a few). There are two reasons for this behavior, the first to gain attention by the use of a easily relatable hashtag and secondly, the hashtag is added to indicate that the tweet content is to be interpreted within the context of country. This category subsumes news about sports, weather, entertainment and business. The count of these tweets is higher than any other category in the case of India (39.21 percent) and a close second for Singapore (17.76 percent).

"The Strike That Rattled #Singapore: A WSJ Investigation. The first of a 5-part series this week. <http://t.co/nAKrimR4OG> #China #migrants"
"#India Second test-firing of Agni-V next month <http://t.co/nQkjUuMVE4>"
"#Indonesia Is Trying to Attract More #Hollywood Films <http://t.co/aHV66dOqL8> #cinema"

4.1.8. National Identity (NI)

This special category corresponds to tweets where users express opinion about the identity of a given country. Based on the tweets, both optimistic and pejorative opinions are conveyed by users in this category. For example, Singapore is praised for its tourism while India is blamed for its corruption and inefficient government, with patriotic comments about Indonesia. This category is prevalent in India (7.75 percent) as it corroborates with the other related categories Local News (LN) and National Group References (NGR). Its presence is of decent percentage in Indonesia (3.89 percent) while it is almost negligible for Singapore (0.68 percent).

"NOTICE: Please note that our gov has made 107 requests to #Facebook for information on 117 user accounts. #Singapore"
"The kind of secularism being practiced by political parties in #India is utterly farce n vote-centric. It is highly dangerous & regressive."
"And I'd be happy to show you around, how beautiful my #indonesia is! Keep up, dev! ☺ ..."

4.1.9. Personal Events and Rants (PER)

This category corresponds to tweets that are entirely user specific, referring to a personal event or a personal rant (opinion) about an entity. Frustrations about traffic or a personal communication with other users are candidates for this theme. This category has the highest allocation amongst all the categories for Singapore and Indonesia. The key reasons are cross-media content sharing from Instagram (including only posts that are not about current location or landmark) and secondarily, users use Twitter to voice their individual opinions. Even though this category is a primary category, a secondary category has been assigned to the tweets where there is a presence of an additional theme. The bar chart in [Figure 2](#) shows the mix of secondary themes in tweets. The data label 'Primary' refers to tweets which have PER as the sole category. The presence of secondary themes is more than 50 percent for India and Singapore while it is just 3.2 percent for Indonesia since users post their personal images from Instagram and other image-sharing sources without much information about location.

"Sexual transmitted diseases r increasing in #singapore #std #nsc"
"@<usermention> Have you ever been here? #India, #Andhrapradesh ?"
"Join me on #Path: #me #mine #instabeauty #bandung #indonesia #instaplace #asian #asianboy #asiangirl ..."

Figure 2: Allocation of secondary themes in PER tweets in the three hashtag samples.

4.1.10. Tourism and Travel Related (TTR)

This category corresponds to tweets that are related to tourism and travel related information sharing by users. It is closely related to the Current Location and Landmark (CLL) category. However, the difference is notable with users posting tweets indicating their travel in and out of a given country or posting tweets to promote tourism for a particular locality. The usage of the country hashtag in the context of these tweets is very specific as it is directly related to the main topic of the tweet. This category does not have peers in the classifications of previous studies due to its specific nature. This category has a comparatively higher presence in Singapore (4.44 percent) when compared to Indonesia (2.33 percent) and India (1.12 percent).

"Just a flashback from 2012 Singapore Visit.. :-) #Singapore
#vacation #Merlion #throwback #malaysia ..."
"#newdelhi airport. Welcome to #India ..."
"TODAY's TOP SHOT #INDONESIA, NUSA DUA : Foreign #tourists
enjoy water hitting...: INDONESIA, NUSA DUA : Foreign ..."

4.1.11. Unrelated

This category corresponds to tweets where the presence of the country has no connection with the content of the tweets. A majority of the tweets in this category are spam posts. The presence of spam posts in popular social media platforms is a known phenomenon (Agichtein, *et al.*, 2008). Spammers make use of the hashtag in their tweets to increase their reach as country hashtags are trending topics quite often in Twitter. This category has a sizeable presence in all three samples with India (17.68 percent) being the highest followed by Indonesia (12.80 percent) and Singapore (11.33 percent).

"Diabetes by arithmetic 22 WCMS 4 #law #Singapore #Jakarta
#China #Police #Navy #auspol #Catholic #USA #Islam"
"HARRY'S WAR on # Amazon! #USA #CANADA #SPAIN
#GERMANY #FRANCE #ITALY #JAPAN #BRAZIL #UK #INDIA."
"♥RETWEET TO GET MORE FOLLOWERS♥
#MustFollow#TeamFollowBack #RT2Follow#JFB #FFFB
#IFollowBack#MentionToFollow#INDONESIA #72"

4.2. RQ2: What are the prevalent network types in user communication data extracted from the tweets?

The purpose of using network analysis in this study is twofold. The first objective is to identify the prevalent network types and secondly, to learn about communication pattern(s) of users in Twitter

with these hashtags. The directed graphs formed with the data from the original extracts are illustrated in [Figure 3](#) (a, b & c). [Figure 4](#) (a, b & c) provides the cumulative frequencies of the degree metric from the three graphs. Statistics of the three graphs are provided in [Table 4](#).

The original extracts for the three hashtags have differences in terms of number of nodes and edges due to the varying number of English tweets generated during the sampling period. The other major difference is the reduction in the number of nodes and edges when the graphs are filtered to form giant component graphs. The percentage of reduction in number of nodes from the initial graph (original extract) to the connected graph is 89.39 percent (#singapore), 71.20 percent (#india) and 96.24 percent (#indonesia) respectively. The #singapore and #indonesia extracts have significant reduction in size which signifies the tweeting styles of its users. These users tend to post tweets with just textual content, without directing the tweets to other Twitter users. An average degree less than 2 indicates the long tail of nodes with just a single interaction with other user nodes in the sample set. There is a semblance of a power-law distribution in the degree frequencies of #singapore and #india ([Figure 4](#) a & b).

The prevalent network types in the graphs are similar in the case of #singapore and #indonesia ([Figure 3](#) a & c). Broadcast networks are common in these two graphs as Twitter users retweet or reply to the tweets made by celebrities (e.g., Pitbull) and new agencies (e.g., strait times). These broadcast networks are connected to each other by a few nodes which are part of different groups. In these networks, users are mostly connected to news hubs and prominent personalities (Smith, *et al.*, 2014). The #singapore graph represents some characteristics of 'community clusters' with many interconnections between the clusters. The network structure of the #india graph is similar to that of a 'tight crowd' network where the users are closer to each other, even though they may be part of different communities. In this network type, mutual sharing and support are the key characteristics. Even in this graph, there are broadcast networks centered on few nodes with high in-degree values, however the overall network is tightly knit which indicates the presence of common discourse among users.

Figure 3: Directed graphs built with 'User-mentions' data from the three hashtag Twitter samples.

Figure 4: Cumulative degree distributions of users in the three hashtag samples. Line graphs a, b and c are for #singapore, #india and #indonesia respectively.

Table 4: Graph statistics from three directed user-mentions graphs.			
<i>Statistic</i>	<i>#singapore</i>	<i>#india</i>	<i>#indonesia</i>
Nodes (from original extract)	5,251	21,214	7,917
Edges (from original extract)	1,684	11,144	2,049
Nodes (only connected nodes)	557	6,110	297
Edges (only connected nodes)	647	8,209	433
Average degree	1.162	1.344	1.458
Average path length	2.071	1.806	1.066

4.3. RQ3: Does the provenance data of the tweets and other tweet statistics provide any new insights?

Instagram is the most used source in tweets that contain #singapore and #indonesia ([Table 5](#)) which directly translates to the high number of tweets in the CLL category and PER category for these two hashtags. This demonstrates the popularity of Instagram as a media sharing platform and also the intent of users to promote visual content through Twitter. However, Instagram is not in the top five sources for #india since the focus of users who tweet with #india leans towards national topics and less towards self-promotion. This observation is backed by the highest percentage of tweets in the Local News (LN) category (39.21 percent). The syndication service 'dlvr.it' is the most used source in #india tweets. This finding corresponds to the high number of tweets by users who broadcast news articles to their followers.

The other major sources are the Web and smart phones. [Table 6](#) shows that a high percentage of tweets from the sample set contain URLs, with #indonesia topping the list with 81.61 percent. This findings illustrates that Twitter is used to share content, like other social media sites. URL Sharing URLs is a stamp of approval in some cases, a statement that "I as a user have gone through this Web site and I feel this will be worth reading for you too". An earlier study by Liu (2013) suggested this finding. The major presence of URLs in tweets is due to content sharing from Instagram, where a given link points to an image or video. Hence, provenance data can be generalized to other studies involving location hashtags.

Table 5: Top five Twitter sources from three hashtag samples.		
<i>#singapore</i>	<i>#india</i>	<i>#indonesia</i>
Instagram	dlvr.it	Instagram
Web	Web	Web
twitterfeed	twitterfeed	Twitter for BlackBerry
Twitter for iPhone	Twitter for Android	Write Longer
dlvr.it	TweetDeck	Twitter for iPhone

Table 6: Additional statistics on tweets in three hashtag samples.			
<i>Extract</i>	<i>Percentage in English</i>	<i>Percentage of retweets</i>	<i>Percentage with URLs</i>
#singapore	84.07	29.44	68.59
#india	25.90	35.33	78.00
#indonesia	85.74	31.71	81.61

5. Discussion

The 11 categories identified during this classification exercise have both similarities and differences with categories from previous studies. The common categories include News (LN) (Horn, 2010; Java, *et al.*, 2007; Rosa, *et al.*, 2011; Sriram, *et al.*, 2010), Current Location (CLL) (Naaman, *et al.*, 2010), Commercial Deals (CD) (Honeycutt and Herring, 2009; Jansen, *et al.*, 2009; Naaman, *et al.*, 2010; Pear Analytics, 2009; Sriram, *et al.*, 2010), Spams (Unrelated) (Dann, 2010; Pear Analytics, 2009) and Group References (GR) (Dann, 2010; Honeycutt and Herring, 2009; Naaman, *et al.*, 2010). Hence, these findings validate the pervasive nature of these categories. The novel categories Asia Related (AR), Personal Events (PER), Tourism (TTR) and National Identity (NI) have been newly identified mainly due to the specific nature of a location-based hashtag, such as the country hashtag and usage of hashtag centric frame of reference in the course of classifying tweets. An all-encompassing classification method needs to have sub-categories to capture the themes of tweets or the classification has to be set at some abstract level. One could argue that the new set of categories will be useful for some future studies.

The varying allocation percentages of categories across the three countries can be attributed to tourist popularity for each country. Singapore and Indonesia are prominent tourist attractions in Asia compared to India (United Nations World Tourism Organization (UNWTO), 2015). Accordingly, the categories CLL, PER and TTR, which are about the physical location of the user, are prominent in these two countries. In this context, since users mostly post pictures of their whereabouts, Instagram is the top source. When tweets were analyzed at a structural level by forming graphs based on user mentions data, broadcast networks were found. There is not a great deal of correlation between content and structural analysis for Singapore and Indonesia mainly because most of the users are not part of communication networks. Users who generate a given country hashtag often don't direct a specific tweet to some other user or reply to other users. However, broadcast networks were found because of the retweeting behavior of users in the special case of tweets from popular personalities being retweeted. Based on the findings, the key features for Singapore and Indonesia were *self-promotion* and *user-based orientation* since the focus of the hashtag usage in the tweets was directed towards users themselves.

On the other hand, most of the tweets with the India hashtag fell into the categories LN, PER and NI, indicating interest on common national topics. Correspondingly, the syndication service dlvr.it was the most used source as news articles were posted, retweeted and used for to initiate discussions. Retweeting was a major trend with the popularity of news related Twitter accounts. At a structural level, a tight crowd network was found in user mentions. This observation correlated with the identified top categories, since retweeting and group discussions were common in the tweets using India hashtag. The key features for India was *news-promotion* and *content-based orientation* since the focus of hashtag use was directed mainly at content posted in the medium.

The findings from this study add to existing literature on country/nation based narratives in online social networks. Secondly, the identified categories could be useful in future hashtag recommendation studies (Godin, *et al.*, 2013; Xiao, *et al.*, 2012) particularly for rule-based recommenders. Hashtags could be recommended based on the tweet content compliance to certain pre-set rules.

6. Limitations

The findings from this study were based on an in-depth analysis of tweets extracted for three country hashtags, collected for a duration of a week. Since the three countries are from Asia, the category Asia-Related will not be suitable for tweets for non-Asian countries. Tweets posted in English were considered in this study; therefore there is a possibility of missing other findings based on non-English tweets. It is expected that the same set of categories would be identified with different Twitter extracts, although there is a possibility of changes in the allocation percentages of tweets in each of the categories. The categories assigned to the tweets do not necessarily represent a singular theme in each tweet but instead the assignment was based on the most prominent theme in a tweet. Thereby, the categories should not to be considered as mutually exclusive.

7. Conclusion and future work

The objective of the study was to identify the rationale behind the usage of a particular type of commonplace hashtag: the country hashtag. Twitter data extracts containing three country hashtags were manually classified to observe content level characteristics of tweets. Eleven categories were identified. The hashtags were found to be prominent in tweets about personal rants, local events, local news, users' current location and landmark related information sharing.

In the case of #singapore and #indonesia, users who shared content from social media sites, such as Instagram, used the hashtag more prominently than users who posted textual content. News sharing was the most prevalent activity in #India tweets. News agencies and commercial bodies made use of the hashtag more than common individuals in all three extracts. Similarities and differences with existing tweet classifications were identified along with the justifications for novel categories. This classification scheme could be used in future studies involving nation/country level analysis in microblogs, as a part of practical implications. Network analysis was performed to identify network types and communication patterns between users in extracts. Broadcast network was the commonly found network type in user communication networks generated for tweets, indicating prominence given to news hubs and eminent personalities in Twitter. These broadcast networks combine to form interconnected communication clusters for #singapore and #indonesia, while a tight crowd network was evident for #india, indicating the existence of common discourse and sharing across users.

As a part of future work, the manually classified tweets will be used for identifying appropriate features for machine learning classification algorithms so that manual effort will be reduced in future studies. A similar study with hashtags from other geographical zones will be conducted to validate the findings from this study. Cross-media validation will be performed by extracting similar data from platforms such as Google Plus and Facebook as the hashtag has become a common feature across many social media platforms. It would be interesting to see if users make use of commonplace hashtags with similar intentions across other platforms.

About the authors

Aravind Sesagiri Raamkumar is a doctoral student in the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He received his Master's of Science in knowledge management from Nanyang Technological University. His research interests include recommender systems, information retrieval, social media and linked data.

Direct comments to aravind002 [at] ntu [dot] edu [dot] sg

Natalie Pang is an Assistant Professor at the Wee Kim Wee School of Communication and Information, and Principal Investigator at the Centre of Social Media Innovations for Communities (COSMIC) at Nanyang Technological University (NTU). She specialises in social informatics, with social media and information behaviour in stochastic and crises contexts being her main area of research. She also studies structurational models of technology in marginalised communities such as older adults and people with disabilities.

E-mail: nlspang [at] ntu [dot] edu [dot] sg

Schubert Foo is Professor of Information Science at the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He received his B.Sc.(Hons), M.B.A. and Ph.D. from the University of Strathclyde, U.K. He is a Chartered Engineer, Chartered IT Professional, Fellow of the Institution of Mechanical Engineers and Fellow of the British Computer Society. He has authored more than 250 publications in the areas of multimedia technology, Internet technology, multilingual information retrieval, digital libraries, information literacy and social media research. He is the current Director of the Centre of Social Media Innovations for Communities at NTU. E-mail: sfoo [at] ntu [dot] edu [dot] sg

Acknowledgments

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office.

References

Fabian Abel, Eelco Herder, Geert-Jan Houben Nicola Henze and Daniel Krause, 2013. "Cross-system user modeling and personalization on the social Web," *User Modeling and User-Adapted Interaction*, volume 23, number 2, pp. 169–209.

doi: <http://dx.doi.org/10.1007/s11257-012-9131-2>, accessed 11 January 2016.

Fabian Abel, Ilknur Celik, Geert-Jan Houben and Patrick Siehdnel, 2011. "Leveraging the semantics of tweets for adaptive faceted search on Twitter," In: Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Noy and Eva Blomqvist (editors). *The Semantic Web — ISWC 2011. Lecture Notes in Computer Science*, volume 7031. Berlin: Springer, pp. 1–17.

doi: http://dx.doi.org/10.1007/978-3-642-25073-6_1, accessed 11 January 2016.

Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis and Gilad Mishne, 2008. "Finding high-quality content in social media," *WSDM '08: Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 183–194.

doi: <http://dx.doi.org/10.1145/1341531.1341557>, accessed 11 January 2016.

Alexa, 2015. "Twitter.com: Site overview," at <http://www.alexa.com/siteinfo/twitter.com>, accessed 12 June 2015.

Mathieu Bastian, Sebastien Heymann and Mathieu Jacomy, 2009. "Gephi: An open source software for exploring and manipulating networks," *Third International AAAI Conference on Weblogs and Social Media*, at <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>, accessed 11 January 2016.

Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues and Virgilio Almeida, 2010. "Detecting spammers on Twitter," *CEAS 2010: Proceedings of the Seventh Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*. Red Hook, N.Y.: Curran Associates, pp. 75–83.

Axel Bruns and Jean E. Burgess, 2011. "The use of Twitter hashtags in the formation of ad hoc publics," *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*, at <http://eprints.qut.edu.au/46515/>, accessed 11 January 2016.

Marc Cheong and Sid Ray, 2011. "A literature review of recent microblogging developments," at <http://www.csse.monash.edu.au/publications/2011/tr-2011-263-full.pdf>, accessed 11 January 2016.

Marc Cheong and Vincent Lee, 2010. "Dissecting Twitter: A review on current microblogging research and lessons from related fields," In: Nasrullah Memon and Reda Alhajj (editors). *From sociology to computing in social networks: Theory, foundations and applications*. Wien: Springer-Verlag, pp. 343–362.

doi: http://dx.doi.org/10.1007/978-3-7091-0294-7_18, accessed 11 January 2016.

Stephen Dann, 2010. "Twitter content classification," *First Monday*, volume 15, number 12, at <http://firstmonday.org/article/view/2745/2681>, accessed 11 January 2016.

Yajuan Duan, Furu Wei, Ming Zhou and Heung-Yeung Shum, 2012. "Graph-based collective classification for tweets," *CIKM '12: Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pp. 2,323–2,326.

doi: <http://dx.doi.org/10.1145/2396761.2398631>, accessed 11 January 2016.

Frédéric Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen and Rik Van De Walle, 2013. "Using topic models for Twitter hashtag recommendation," *WWW '13: Companion Proceedings of the 22nd International Conference on World Wide Web*, pp. 593–596.

Courtenay Honeycutt and Susan C. Herring, 2009. "Beyond microblogging: Conversation and collaboration via Twitter," *HICSS '09: 42nd Hawaii International Conference on System Sciences*, pp. 1–10.

doi: <http://dx.doi.org/10.1109/HICSS.2009.89>, accessed 11 January 2016.

Christopher Horn, 2010. "Analysis and classification of Twitter messages," Master's thesis, Graz University of Technology, at <https://www.yumpu.com/en/document/view/4019685/analysis-and-classification-of-twitter-messages-know-center->, accessed 11 January 2016.

Jeff Huang, Katherine M. Thornton and Efthimis N. Efthimiadis, 2010. "Conversational tagging in Twitter," *HT '10: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pp. 173–178.

doi: <http://dx.doi.org/10.1145/1810617.1810647>, accessed 11 January 2016.

Shu Huang, Wei Peng, Jingxuan Li and Dongwon Lee, 2013. "Sentiment and topic analysis on social media: A multi-task multi-label classification approach," *WebSci '13: Proceedings of the Fifth Annual ACM Web Science Conference*, pp. 172–181.

doi: <http://dx.doi.org/10.1145/2464464.2464512>, accessed 11 January 2016.

Bernard J. Jansen, Mimi Zhang, Kate Sobel and Abdur Chowdury, 2009. "Twitter power: Tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology*, volume 60, number 11, pp. 2,169–2,188.

doi: <http://dx.doi.org/10.1002/asi.21149>, accessed 11 January 2016.

Akshay Java, Xiaodan Song, Tim Finin and Belle Tseng, 2007. "Why we Twitter: Understanding microblogging," *WebKDD/SNA-KDD '07: Proceedings of the Ninth WebKDD and First SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pp. 56–65.

doi: <http://dx.doi.org/10.1145/1348549.1348556>, accessed 11 January 2016.

Balachander Krishnamurthy, Phillipa Gill and Martin Arlitt, 2008. "A few chirps about Twitter," *WOSN '08: Proceedings of the First Workshop on Online Social Networks*, pp. 19–24.

doi: <http://dx.doi.org/10.1145/1397735.1397741>, accessed 11 January 2016.

Yu-Ru Lin, Drew Margolin, Brian Keegan, Andrea Baronchelli and David Lazer, 2012. "#Bigbirds never die: Understanding social dynamics of emergent hashtags," *Proceedings of the Seventh AAAI Conference on Weblogs and Social Media*, pp. 370–379, and at

<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6083>, accessed 11 January

2016.

Wenlin Liu, 2013. "How Twitter connects to information sources: A network analysis of the sourcing structure of the OWS tweets," paper presented at the Annual Conference of International Communication Association (London).

Matteo Magnani, Danilo Montesi, Gabriele Nunziante and Luca Rossi, 2011. "Conversation retrieval from Twitter," In: Paul Clough, Colum Foley, Cathal Gurrin, Gareth J. F. Jones, Wessel Kraaij, Hyowon Lee and Vanessa Mudoch (editors). *Advances in information retrieval. Lecture Notes in Computer Science*, volume 6611. Berlin: Springer, pp. 780–783.
doi: http://dx.doi.org/10.1007/978-3-642-20161-5_93, accessed 11 January 2016.

Mor Naaman, Jeffrey Boase and Chih-hui Lai, 2010. "Is it really about me? Message content in social awareness streams," *CSCW '10: Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work*, pp. 189–192.
doi: <http://dx.doi.org/10.1145/1718918.1718953>, accessed 11 January 2016.

M.E.J Newman, 2006. "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, volume 103, number 23, pp. 8,577–8,582.
doi: <http://dx.doi.org/10.1073/pnas.0601602103>, accessed 11 January 2016.

Mike Orcutt, 2011. "How Occupy Wall Street occupied Twitter, too," *MIT Technology Review* (9 November), at <http://www.technologyreview.com/view/426079/how-occupy-wall-street-occupied-twitter-too/>, accessed 12 June 2015.

Pear Analytics, 2009. "Twitter study — August 2009," at <http://www.pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August-2009.pdf>, accessed 11 January 2016.

Jan Pöschko, 2011. "Exploring Twitter hashtags," *arXiv.org* (28 November), at <http://arxiv.org/abs/1111.6553>, accessed 11 January 2016.

Kevin Dela Rosa, Rushin Shah, Bo Lin, Anatole Gershman and Robert Frederking, 2011. "Topical clustering of tweets," *SWSM'11*, at <http://www.cs.cmu.edu/~kdelaros/sigir-sws-2011.pdf>, accessed 11 January 2016.

Garcia Esparza Sandra, Michael P. O'Mahony and Barry Smyth, 2010. "Towards tagging and categorization for micro-blogs," paper presented at the 21st National Conference on Artificial Intelligence and Cognitive Science (AICS 2010); version at <http://researchrepository.ucd.ie/handle/10197/2517>, accessed 11 January 2016.

Marc A. Smith, Lee Rainie, Ben Shneiderman and Itai Himelboim, 2014. "Mapping Twitter topic networks: From polarized crowds to community clusters," *Pew Research Center* (20 February), at <http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/>, accessed 12 June 2015.

Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu and Murat Demirbas, 2010. "Short text classification in twitter to improve information filtering," *SIGIR '10: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 841–842.
doi: <http://dx.doi.org/10.1145/1835449.1835643>, accessed 11 January 2016.

Trendinalia, 2015. "Trending topics Singapore," at <http://www.trendinalia.com/twitter-trending-topics/singapore/singapore-150121.html>, accessed 12 June 2015.

Oren Tsur and Ari Rappoport, 2012. "What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities," *WSDM '12: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pp. 643–652.
doi: <http://dx.doi.org/10.1145/2124295.2124320>, accessed 11 January 2016.

United Nations World Tourism Organization (UNWTO), 2015. "UNWTO tourism highlights," at <http://www.e-unwto.org/doi/pdf/10.18111/9789284416899>, accessed 23 September 2015.

Stanley Wasserman and Katherine Faust, 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Shirley Ann Williams, Melissa M. Terras and Claire Warwick, 2013. "What people study when they study Twitter? Classifying Twitter related academic papers," *Journal of Documentation*, volume 69, number 3, pp. 384–410.
doi: <http://dx.doi.org/10.1108/JD-03-2012-0027>, accessed 11 January 2016.

Feng Xiao, Tomoya Noro and Takehiro Tokuda, 2012. "News-topic oriented hashtag recommendation in Twitter based on characteristic co-occurrence," In: Marco Brambilla, Takehiro Tokuda and Robert Tolksdorf (editors). *Web engineering. Lecture Notes in Computer Science*, volume 7387. Berlin: Springer, pp. 16–30.
doi: http://dx.doi.org/10.1007/978-3-642-31753-8_2, accessed 11 January 2016.

Lei Yang, Tao Sun, Ming Zhang and Qiaozhu Mei, 2012. "We know what @you #tag: Does the dual

role affect hashtag adoption?" *WWW '12: Proceedings of the 21st international conference on World Wide Web*, pp. 261–270.

doi: <http://dx.doi.org/10.1145/2187836.2187872>, accessed 11 January 2016.

Dejin Zhao and Mary Beth Rosson, 2009. "How and why people Twitter: The role that micro-blogging plays in informal communication at work," *GROUP '09: Proceedings of the ACM 2009 International Conference on Supporting Group Work*, pp. 243–252.

doi: <http://dx.doi.org/10.1145/1531674.1531710>, accessed 11 January 2016.

Editorial history

Received 5 July 2015; revised 25 September 2015; accepted 12 January 2016.

Copyright © 2016, *First Monday*.

Copyright © 2016, Aravind Sesagiri Raamkumar, Natalie Pang, and Schubert Foo.

When countries become the talking point in microblogs: Study on country hashtags in Twitter by Aravind Sesagiri Raamkumar, Natalie Pang, and Schubert Foo.

First Monday, Volume 21, Number 1 - 4 January 2016

<http://firstmonday.org/ojs/index.php/fm/rt/prINTERfriendly/6101/5193>

doi: <http://dx.doi.org/10.5210/fm.v21i1>.