Linguistic Characteristics of Social Media Messages Spreading across Geographic and Linguistic Boundaries

Xinchen Yu[1], Jeremy Boy[2], Rene Clausen Nielsen[3] and Lingzi Hong[1]
[1] College of Information, University of North Texas, Denton, TX, USA
[2] UNDP Accelerator Labs, USA
[3] UN Global Pulse
Xinchen.Yu@my.unt.edu
jeremy.boy@undp.org
rene@unglobalpulse.org
lingzi.hong@unt.edu

Abstract: Social media enable messages to be exchanged beyond geographic constraints. Some of the messages could be shared and forwarded by people with different cultural backgrounds across different geographical regions. Studying the content of messages that can reach diverse populations is important for practices such as movement propagation and global marketing. Existing studies mainly investigated the characteristics of messages that are popular, i.e., shared or forwarded by more users. As the diffusion of information is prone to be echoed inside certain geographical and linguistic boundaries, popular messages are not always to be shared and spread across geographical and linguistic boundaries. We investigated the linguistic characteristics of social media messages that can reach and be disseminated by people across nations, and across geographic and linguistic boundaries in the MeToo movement. Specifically, we analyze the diffusion paths of messages according to the geolocation of tweets and conducted statistical analysis to compare the linguistic characteristic of tweets that spread across geographical or linguistic boundaries with those that do not. We focus on the linguistic characteristics from three aspects: 'emotions', 'social relations', and 'economics, politics, and religion'. Our findings reveal that popular messages tend to contain more negative emotions, however, messages with negative emotions are unlikely to be disseminated across geographical or linguistic boundaries. On the other hand, messages on economic topics or non-adults' issues are more probable to be disseminated universally. The findings provide insights on the content that is more probable to be shared and disseminated by people with different cultural backgrounds across geographical regions.

Keywords: Information Diffusion; Twitter; Transnational Movement; Content Analysis

1. Introduction

Social media enable the spread of information beyond geographic constraints, which facilitate cultural and social globalization (Choi *et al.,* 2013) in the forms such as transnational movements (Della *et al.,* 2015). Marketing and propagation projects are implemented on social media platforms, some of which aim to achieve transnational impact (Sirkeci, 2013). The diffusion of information on social networks has been studied a lot. Many focus on the identification or characterization of posts that are more likely to be shared by users, i.e., popular (Firdaus *et al.,* 2018). As people tend to communicate with those who share similar backgrounds or interests, the diffusion of information is prone to be echoed inside certain geographic and language boundaries (De Choudhurry *et al.,* 2010). Therefore, popular information does not always have a transnational impact. It has been found only a small proportion of popular tweets have transnational diffusion (Yu *et al.,* 2020). In this study, we investigate characteristics of information that is more probable to be disseminated by people from different geographical regions with diverse cultural backgrounds.

The diffusion of information on social networks is through posting and sharing (Suh *et al.,* 2010). Taking posts on Twitter as an example, the popularity and spatial diffusion of a tweet can be measured based on the number of retweets and the geographic distribution of its retweets. We aim to address two questions:
RQ1: For tweets of similar popularity, what linguistic characteristics differentiate their potentials to disseminate to regions with different geographic proximity, starting from a transnational level to a trans-continent level?

RQ2: For tweets of similar popularity, what linguistic characteristics differentiate their potentials to disseminate to regions with or without linguistic proximity?

We conducted a study on the spatial diffusion of information in the transnational MeToo movement. Many factors could influence the diffusion of a message. The popularity of topics changes over time with occurrences of events (Hu *et al.,* 2016). We set the study in the context of the MeToo movement on Twitter to control factors such as the time and topics, to focus on the relations between linguistic characteristics of messages and their spatial diffusion. Hashtag MeToo (#MeToo) has been used as a symbol of the exposure and condemnation of gender-based violence on social networks since October 2017, which forms the uptake of the MeToo movements across many countries in the world. Transnational movements such as the MeToo movement are one of the scenarios where we can observe information spread globally. In this study, we examine the spatial diffusion of the MeToo tweets and conduct statistical tests to identify the linguistic characteristics of tweets with different levels of spatial diffusion. The study has three main contributions:

- The diffusion of information is investigated from a spatial perspective. Specifically, geographic proximity and linguistic proximity are used to measure the diffusion across geographic and linguistic boundaries.
- The findings imply that the linguistic characteristics that lead a message to be popular such as negative emotions may retrain it from disseminating across geographic and language boundaries.
- Through analysis of #MeToo tweets, this study sheds light on the linguistic characteristics of information that disseminate across geographic and language boundaries, which may assist content-generating in social movement propagation that aims for global impact and provides insights about factors that relate to the development of transnational social movements.

2. Literature review

2.1 Sentiment and Topics on the Diffusion of Information

Recent studies have examined whether emotions affect the diffusion and what roles emotions play in the 'virality' of messages. Stieglitz and Dang-Xuan (2013) found that emotionally charged tweets are more widely transmitted than neutral messages. There is a positive correlation between the number of retweets and the proportion of words indicating positive and negative emotions of tweets on political events. Several researchers have investigated the effect of different emotions or polarity of sentiment on diffusion. Ferrara and Yang (2015) found that negative messages spread faster than positive ones, but positive messages finally can reach more audiences. Tsugawa and Ohsaki (2017) found that negative information is likely to be retweeted more frequently and rapidly than positive or neutral information. Fan *et al.* (2016) categorized emotions into joy, anger, disgust, and sadness, and found anger is more contagious than joy given that anger could trigger more follow-up tweets with anger emotions. Similarly, Chen *et al*. (2017) found that angry tweets spread more rapidly than happy tweets. These studies are conducted based on short social media posts like tweets. Emotions are also found to relate to the diffusion of long texts such as news articles. Berger and Milkman (2012) coded the sentiment of New York Times articles and found that 'virality' was partially driven by physiological arousal. Content that triggered positive or negative emotions was more viral. Kim (2015) analyzed the sharing and viewing of New York Times news articles on health and found that samples with positive sentiments had more views.

Most research focused on how sentiments or emotions relate to information diffusion. Little has been done to understand the effect of structural components such as indicators of social relations or cognition on diffusion. Arnaboldi *et al.* (2016) used direct contacts between users and their friends to measure social ties and found tie strength and the number of retweets is highly correlated. In this work, we explore the emotions, social relations, and economics, politics, and religion components on the spatial diffusion of information.

2.2 Information Diffusion and Geographical Proximity

Although the information on social networks has the potential to spread globally across geographic boundaries (Takhteyev *et al.,* 2012), studies showed that diffusion is mostly constrained in a certain geographic scope (Wang *et al.,* 2012). The diffusion of information partially relies on network structures of users (Guille *et al.,* 2013), while most social-network friendships are dependent on geographic proximity (Liben-Nowell *et al.,* 2005). Several studies have revealed that users are more inclined to become friends with those who live close (Kulshrestha *et al.,* 2012; Scellato *et al.,* 2010). The geographic proximity of users can predict the diffusion paths in the networks (Wang *et al.*, 2013). De Choudhury et al. (2010) further showed that information from users who have similar locations tend to have similar information diffusion patterns in terms of network topology-based reach and spread. These studies imply that popular messages on social media may not reach a wide range of audiences or have transnational diffusion. A few studies have characterized the spatial diffusion of information. Kwon et al. (2015) has developed a spatiotemporal model to examine the spatial diffusion of different types of messages during the Egyptian Revolution 2011. They defined four types of proximity, i.e., physical, diasporic, economic, and ideological proximity, to measure the relations of other countries to Egypt, and examined the diffusion of ad-hoc reporting, situation verifying, and action supportive messages in these countries. Results showed that ad-hoc reporting and action-supportive messages were widely exchanged across geographical boundaries. Here, we aim to understand the linguistic components of 'emotions', 'social relations', and 'economics, politics and religion' in messages on their spatial diffusion.

2.3 Information Diffusion and Linguistic Proximity

Users who use the same language are more likely to build connections on social media platforms (Kulshrestha *et al.,* 2012). Information is mainly disseminated among people who speak or understand the same language. Through an analysis of the social relations based on retweets and mentions on a global scale, Hale (2014) found most of the interactions are in the same language. Linguistic proximity is usually indicative of the economic relations of countries (Hutchinson, 2005) or communities that share similar interests (Samoilenko *et al.,* 2016). However, there is always information disseminated across linguistic boundaries, as there are multilingual users who play the bridging roles to extend the information diffusion chain (Hale, 2014). In addition to the transnational diffusion, we study the linguistic characteristics of tweets that spread to countries without linguistic proximity.

3. Data Collection and Methods

3.1 Twitter Data set

We acquired a list of hashtags used in different countries related to the MeToo movement. These include #MeToo, #YoTambien, #SendeAnlat, #WithYou, #WeToo, #QuellaVoltaChe, #AnaKaman, #RiceBunny, #BalanceTonPorc, #Cuéntalo, #TimesUp, #TimeisNow, #MeQueer #немолчи, #СегаКажувам, #米兔, and #미투. English tweets that contain at least one of the hashtags were retrieved. We only used English tweets as the translation of different languages to one may generate variances in the tones and sentiment that affect the consistency in content analysis. English is the most used language in the MeToo movement (Yu *et al.,* 2020). In total, we gathered 609,476 tweets through the Crimson Hexagon platform (now Brandwatch), which includes original posts and retweets from 209 nations from October 15, 2018, to January 15, 2019. The retrieved Twitter messages are associated with provenance information at the country level, which is derived either by the geo-tagged information or by the location indicated in user profiles. Each post includes the author, country of origin, text, and type—whether it is a "tweet" (original posts), a "reply", or a "retweet". Table 1 shows the top 10 countries with the highest number of retweets.

Table 1: Number of retweets and original posts for selected 10 countries

Country	Retweets	Original Posts
United States of America	15908	2294
India	9954	1339
Japan	2481	225
United Kingdom	1989	426
Canada	1244	396
Argentina	723	103
Australia	711	140
France	556	122
Indonesia	356	27
Spain	308	120

All the retrieved retweets contain the author and content information of the original posts. Through content matching, we can detect all retweets and sharing of an original post to identify the geographic diffusion of a message. We got a total of 41,864 original posts that have been retweeted, the retweet frequency varies from 1 to 4,004. The retweet periods of original posts vary from 1 day to 3 months. More than 95% of retweets happen in fewer than 10 days. Therefore, we take the retweets in the first 10 days to compute the number and geographic distribution of retweets for all the original posts. The spatial diffusion of original posts varies from 1 country to 69 countries.

3.2 Methods

3.2.1 Grouping Messages according to Diffusion Outcomes

The spatial diffusion of a message is measured based on the geographic distribution of its retweets at a country level. The popularity of a message is defined as the total number of retweets it gets. We are interested in the spatial diffusion reaching countries that have different levels of geographic and language proximity. Previous studies have defined geographic proximity as two countries that are on the same continent (Java *et al.,* 2007;

Kulshrestha *et al.,* 2012); and linguistic proximity as two countries share the same languages (Kulshrestha *et al.,* 2012; Samoilenko *et al.,* 2016).

We separated the original posts into three groups according to the level of diffusion across geographical boundaries: i) retweets only exist in the country where the original post was created (OneCountry); ii) retweets exist in other countries but only in those that have geographic proximity (on the same continent) to the origin country (Geo-I); iii) retweets exist in countries that do not have geographic proximity to the origin country (Geo-II). Based on the linguistic proximity of countries, we again divided the original posts into three groups: i) retweets only exist in the country where the original post was created (OneCountry); ii) retweets exist in other countries that share the same official language with the origin country (Lang-I); iii) retweets exist in countries that do not have linguistic proximity to the origin country (Lang-II). Figure 1 illustrates the three types of tweet diffusion of tweets originating from the United States (US).

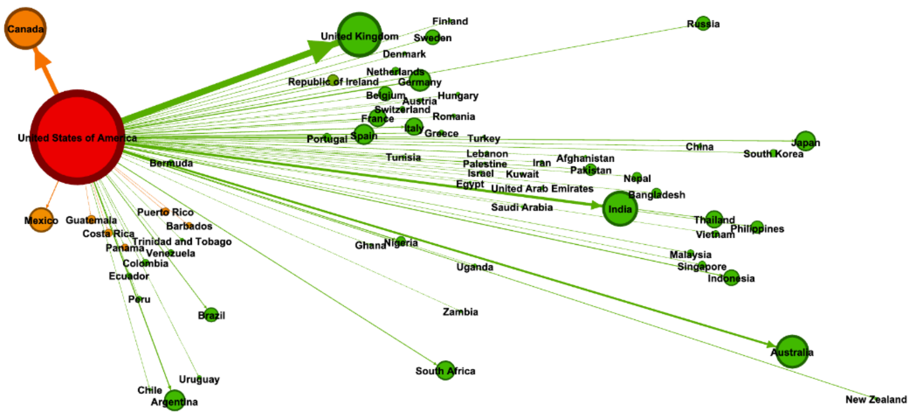

Figure 1: Three types of diffusion for tweets from US. The nodes represent countries. Orange arrows represent diffusion from the US to other countries in North America. Green arrows represent information diffusion to countries not in North America. Node size reflects the number of retweets to these countries.

3.2.2 Control Popularity

Tweets that spread across geographic boundaries usually have a higher number of retweets, i.e., more popular, than those that spread only in one country. Therefore, tweets in the three groups may present certain linguistic features due to the popularity level of tweets, as the linguistic characteristics of information, such as emotions, can affect the popularity in diffusion (Ferrara and Yang*,* 2015; Stieglitz and Dang-Xuan*,* 2013). To focus on the characteristics that are related to spatial diffusion, we control the popularity of tweets in different groups. Then, we measure the Cohen's D value (Durlak*,* 2009) to compare the popularity of tweets in four pairwise comparisons: OneCountry with Geo-I, OneCountry with Geo-II, OneCountry with Lang-I, and OneCountry with Lang-II (Table 2). An effect size under 0.2 can be considered a small difference between the two groups. Since all Cohen's D values of pairwise comparisons in Table 2 are smaller than 0.2, we can conclude that the popularity level of tweets is similar between groups.

Table 2: Effect size of pairwise comparisons for four sets of groups

Groups	Pairwise Comparisons (#tweets)	Cohen's D
Geographical	OneCountry (377) and Geo-I (136)	0.191
	OneCountry (377) and Geo-II (720)	0.102
Linguistic	OneCountry (347) and Lang-I (410)	0.088
	OneCountry (347) and Lang-II (112)	0.193

3.2.3 Content Analysis and Statistical Test

We apply SEANCE to analyze the linguistic components of tweets (Crossley *et al.,* 2017). SEANCE contains predefined word vectors under categories of sentiment, cognition, and social order. It can be used to measure multiple linguistic dimensions of English text. Each tweet is analyzed from three aspects: i) *Emotion*, which includes dimensions such as positive emotion, negative emotion, sadness, hostile, and anger; ii) *Economy, Politics, and Religion*, which measures the proportion of words relevant to the topics of economic, wealth, and religion, etc.; and iii) *social relation*, which measures the proportion of linguistic indicators of social roles and social bonds. 66 indicators are belonging to these three categories. Each indicator measures one linguistic dimension of the text. The score of an indicator is calculated by the ratio of indicators in the sentence. For example, in the message "We all know they are trying hard to win over women voter's using fear mongering #metoo and identity politics. My question is what are they going to do for men?", one out of 30 words is identified to exist in the word list of Fear_Emolex, so the Fear_Emolex is 0.033 for this tweet.

We checked the homoscedasticity of all the indicators in the four pairwise groups by examining if their variances are the same. Tests showed that none of the 66 indicators are homoscedastic. We then performed Welch's Anova tests to examine the differences between the means of groups (McDonald*,* 2009). Due to the unequal sample sizes and unequal variances of the indicators, we conducted Games-Howell tests for post-hoc pairwise comparisons (Schlegel, 2016). This test is based on Welch correction and uses Tukey's studentized range statistic, which is widely used for testing the differences between means of two groups when their variances are unequal.

4. Results

4.1 Tweets Disseminating across Geographical Boundaries

Table 3 shows the statistical test results for the indexes that are significantly different. We list the pairwise comparison between OneCountry and Geo-I, and between OneCountry and Geo-II.

Table 3: Statistic test results for popular tweets disseminating in one nation versus across nations in one/many continents

Category		Indicators	OneCountry and Geo-I		OneCountry and Geo-II	
			MD	Upper/lower limit	MD	Upper/lower limit
Emotion	Negative emotion	Negative_EmoLex	-0.003	0.011/-0.016	0.010**	0.018/0.002
		hu_liu_neg_nwords	0.003	0.015/-0.009	0.009**	0.017/0.002
		hu_liu_neg_perc	0.047	0.145/-0.051	0.071*	0.134/0.008
		vader_negative	0.008	0.036/-0.021	0.018*	0.035/0.001
Economics, politics, legal & power	Economics& Wealth	Econ_2_GI	-0.007	0.001/-0.016	-0.005*	0.000/-0.010
		Wlttot_Lasswell	-0.007	0.001/-0.014	-0.004*	0.000/-0.008
	Legal & Rectitude	Legal_GI	0.007	0.015/-0.001	0.006*	0.012/0.001
		Rcloss_Lasswell	0.000	0.003/-0.003	0.002*	0.004/0.000
		Rctot_Lasswell	-0.001	0.009/-0.011	0.007*	0.013/0.001
Social relations & communication	Communication	Com_GI	-0.008	0.002/-0.019	-0.008**	-0.002/-0.015
		Comform_GI	-0.012*	0.000/-0.024	-0.004	0.003/-0.011
	Social relations	Nonadlt_GI	0.000	0.001/0.000	-0.003***	-0.001/-0.005
		Race_GI	0.001*	0.003/0.000	0.001	0.002/0.000

Observation 1. Most of the significant indexes in the Geographic set are not significant between the pair of OneCountry and Geo-I (except 'Comform_GI' and 'Race_GI'), which reflects that the linguistic characteristics of messages that spread in one country are very similar to those spread in countries with close geographic proximity. Geographically close countries may tend to share similar economic or cultural backgrounds. The spatial distance in diffusion may not be affected by the characteristics of the content if the diffusion is in countries with close geographic proximity. Nevertheless, messages that spread to countries in other continents present different characteristics, which are to be presented in the following observations.

Observation 2. Negative tweets are less probable to be disseminated across geographical boundaries. The mean deviations (MD) in negative emotion indexes (Negative EmoLex, hu_liu_neg_nwords, hu_liu_neg_perc, vader_negative) are all positive in the first pairwise comparison (OneCountry - Geo-II), which shows that messages that spread across continents generally carry fewer negative emotions. Messages that trigger negative emotions may be viral (retweeted by many users), but the diffusion tends to be constrained to countries that share close geographic proximity. This finding confirms that popularity in diffusion is not equivalent to the wider geographic scope of diffusion. Messages with less negative emotions are more probable to reach diverse and geographically far apart populations.

Observation 3. Tweets that contain economy-related linguistic components tend to spread to countries without geographic proximity, while tweets with legal or rectitude components are more likely to spread within a country or to countries in close geographic proximity. The proportions of economic and wealth elements (Econ_2_GI, Wlttot_Lasswell) are significantly higher in tweets that spread across continents, while the index values of legal (Legal_GI) and rectitude (Rcloss_Lasswell, Rctot_Lasswell) elements are significantly smaller comparing tweets in OneCountry and Geo-II. The economy is the most popular topic in online news articles according to the Project for Excellence in Journalism (Purcell *et al.,* 2010). On the other hand, as legal systems of countries differ, tweets addressing legal issues or advocating for rectitude can be more limited to national or closely located countries.

Observation 4. Messages mentioning children or youth are more retweeted by diverse and geographically far apart populations. A social role is one dimension that belongs to the category of social relations, which includes indexes of 'Female', 'Male', 'Humans', and 'Non-Adults' etc. The index of Nonadlt_GI refers to words associated with infants through adolescents. Tweets that spread in one nation have significantly lower Nonadlt_GI values compared to those spread across continents (MD= -0.003***). The transnational #MeToo movement has provided a new avenue for adolescent survivors to disclose sexual abuse or assault (Alaggia and Wang, 2020), and non-adult issues may be a universal concern and trigger information to be disseminated across national and continental boundaries.

4.1 Tweets Disseminating across Linguistic Boundaries

We performed two pairwise comparisons: OneCountry with Lang-I and OneCountry with Lang-II to investigate the factors that relate to dissemination across linguistic boundaries. Table 4 shows the indexes that are significantly different in either of the two pairwise comparisons.

Table 4: Statistic test results for popular tweets disseminating in one nation versus across nations with/without linguistic proximity.

Category		Indicators	OneCountry and Lang-I		OneCountry and Lang-II	
			MD	Upper/lower limit	MD	Upper/lower limit
Emotion	Negative	No_GI	0.000	0.002/-0.002	0.003***	0.002/0.004
		Negativ_GI	0.005	0.015/-0.005	0.022***	0.009/0.034
		vader_negative	0.008	0.028/-0.012	0.047***	0.019/0.074
		Negative_EmoLex	0.004	0.014/-0.006	0.022**	0.008/0.035
		Hostile_GI	-0.001	0.008/-0.009	0.016**	0.005/0.027
		Anger_EmoLex	-0.001	0.006/-0.009	0.014**	0.004/0.024
		Fail_GI	0.001	0.004/-0.002	0.004**	0.001/0.007
		hu_liu_neg_nwords	0.005	0.013/-0.004	0.015**	0.004/0.027
		Ngtv_GI	-0.001	0.008/-0.009	0.013*	0.001/0.024
		hu_liu_neg_perc	0.031	0.103/-0.041	0.118*	0.010/0.226
	Positive	hu_liu_pos_perc	-0.017	0.046/-0.079	-0.133**	-0.236/-0.029
Economics, politics & legal	Economics& Wealth	Econ_2_GI	-0.005	0.001/-0.010	-0.013*	-0.025/-0.002
		Wlttot_Lasswell	-0.003	0.001/-0.008	-0.010*	-0.020/0.000
	Legal & Rectitude	Legal_GI	0.004	0.011/-0.002	0.013***	0.005/0.021
		Rcgain_Lasswell	0.001	0.004/-0.001	0.004**	0.001/0.007
		Rctot_Lasswell	0.005	0.012/-0.002	0.012**	0.004/-0.020
Social relations & communica tion	Communication	Com_GI	-0.011**	-0.004/-0.018	-0.004	-0.015/0.007
		formlw_Lasswell	-0.008*	-0.001/-0.015	-0.006	-0.017/0.006
	Social relations	Nonadlt_GI	-0.003**	-0.001/-0.005	-0.005**	-0.010/-0.001
		Ptlw_Lasswell	0.000	0.005/-0.005	0.008*	0.002/0.014

Observation 5. Tweets that spread to countries without linguistic proximity contain fewer negative emotions. The mean difference (MD) of negative emotions, such as negative (Negative_EmoLex) and hostile (Hostile_GI) are both positive in the comparison between OneCountry and Lang-II. This finding is similar to Observation 2 in Section 4.1. While nearly half of the contents in the MeToo movement are angry-related (Yu et al., 2020), the anger might rise empathy for people who live far away (Berger and Milkman, 2012), such tweets are less likely to spread to countries with no linguistic proximity. In addition, Table 4 shows tweets that disseminate across linguistic boundaries contain more positive emotions (hu_liu_pos_perc).

Observation 6. Tweets with more economic, and wealth-related components are more probable to spread to countries without linguistic proximity, while tweets on legal matters and expressing rectitude are less likely to go beyond national or linguistic boundaries. The economic and wealth indexes (Econ_2_GI, Wlttot_Lasswell) have significantly negative MD in the comparison between OneCountry and Lang-II, while legal and rectitude indexes (Legal_GI, Rcgain_Lasswell) have significantly positive MD. These results are like the findings in observation 3 in 4.1. Meanwhile, the MD of Nonadlt_GI is both negative at the significance level of $p<0.01$, which complies with Observation 4 that messages on non-adult issues in the MeToo movement can especially trigger people from diverse countries to share.

5. Conclusion and Future Work

We investigated the characteristics of social media messages disseminating across geographic and linguistic boundaries. For RQ1, results showed that tweets containing more negative emotions are unlikely to spread across continents, while those focusing more on economic topics or non-adults' issues can travel a long distance across geographic boundaries. Regarding RQ2, there are similar patterns: tweets that disseminate to countries without linguistic proximity contain more positive, fewer negative emotions, and are less focused on legal or rectitude matters. There are several directions to explore in our future work. First, we will include more cases to figure out if the findings are specific to the diffusion of #MeToo messages and if the findings can generalize to other cases. Second, we would like to figure out who and what are the roles of users in the diffusion of messages across geographic and linguistic boundaries.

References

Alaggia, R. and Wang, S., 2020, "I never told anyone until the# metoo movement": What can we learn from sexual abuse and sexual assault disclosures made through social media?. *Child Abuse & Neglect*, Vol. 103 No. 104312, pp.1-10.

Arnaboldi, V., Conti, M., La Gala, M., Passarella, A. and Pezzoni, F., 2016. Ego network structure in online social networks and its impact on information diffusion. *Computer Communications*, 76, pp.26-41.

Berger, J. and Milkman, K. L., 2012, What makes online content viral?. *Journal of marketing research*, Vol. 49 No. 2, pp. 192-205.

Chen, J., Liu, Y. and Zou, M., 2017, User emotion for modeling retweeting behaviors. *Neural Networks,* Vol. 96, pp. 11-21.

Choi, J., Jung, J. and Lee, S. W., 2013, What causes users to switch from a local to a global social network site? The cultural, social, economic, and motivational factors of Facebook's globalization. *Computers in Human Behavior*, Vol. 29 No. 6, pp. 2665-2673.

Crossley, S. A., Kyle, K. and McNamara, D. S., 2017, Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior research methods*, Vol. 49 No.3, pp. 803-821.

De Choudhury, M., Sundaram, H., John, A., Seligmann, D. D. and Kelliher, A., 2010, " Birds of a Feather": Does User Homophily Impact Information Diffusion in social media? available at: https://arxiv.org/abs/1006.1702/.

Della Porta, D., Andretta, M., Calle, A., Combes, H., Eggert, N., Giugni, M. G., and Marchetti, R. (Ed.s), 2015, *Global justice movement: Cross-national and transnational perspectives*. Routledge, Abingdon, UK.

Durlak, J. A., 2009, How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology,* Vol. 34 No. 9, pp. 917-928.

Fan, R., Xu, K. and Zhao, J., 2016, Higher contagion and weaker ties mean anger spreads faster than joy in social media, available at: https://arxiv.org/abs/1608.03656.

Ferrara, E. and Yang, Z., 2015, Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, Vol. 1 No. e26, pp. 1-15, doi: 10.7717/peerj-cs.26.

Firdaus, S. N., Ding, C. and Sadeghian, A., 2018, Retweet: A popular information diffusion mechanism–A survey paper. *Online Social Networks and Media*, Vol. 6, pp. 26-40.

Guille, A., Hacid, H., Favre, C. and Zighed, D. A., 2013, Information diffusion in online social networks: A survey. *ACM Sigmod Record,* Vol. 42 No. 2, pp. 17-28.

Hale, S. A., 2014, Global connectivity and multilinguals in the Twitter network. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2014,* Toronto, Canada, pp. 833-842.

Hu, Y., Hu, C., Fu, S., Shi, P. and Ning, B., 2016, Predicting the popularity of viral topics based on time series forecasting. *Neurocomputing*, Vol. 210, pp. 55-65.

Hutchinson, W. K., 2005, "Linguistic distance" as a determinant of bilateral trade. *Southern Economic Journal*, Vol. 72 No. 1, pp. 1-15.

Java, A., Song, X., Finin, T. and Tseng, B., 2007, Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis,* San Jose, California, pp. 56-65.

Kim, H. S., 2015, Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. *Journal of Communication,* Vol. 65 No. 3, pp. 512-534.

Kulshrestha, J., Kooti, F., Nikravesh, A. and Gummadi, K., 2012, Geographic dissection of the Twitter network. *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 6 No. 1, pp. 202-209.

Kwon, K. H., Wang, H., Raymond, R. and Xu, W. W., 2015, A spatiotemporal model of twitter information diffusion: An example of egyptian revolution 2011. *Proceedings of the 2015 International Conference on Social Media & Society, SMSociety 2015, Toronto, Canada,* pp. 1-7.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P. and Tomkins, A., 2005, Geographic routing in social networks. *Proceedings of the National Academy of Sciences,* Vol. 103 No. 33, pp.11623-11628.

McDonald, J. H., 2009, *Handbook of biological statistic*. Sparky house publishing, Baltimore, MD, Vol. 2, pp. 6-59.
Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T. and Olmstead, K., 2010, Understanding the participatory news consumer. *Pew Internet and American Life Project*, Vol.1, pp. 19-21, available at: https://assets.pewresearch.org/wp-content/uploads/sites/13/legacy/Participatory_News_Consumer.pdf .
Samoilenko, A., Karimi, F., Edler, D., Kunegis, J. and Strohmaier, M., 2016, Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity. *EPJ data science,* Vol. 5 No. 1, pp. 1-20.
Scellato, S., Mascolo, C., Musolesi, M. and Latora, V., 2010, Distance Matters: Geo-social Metrics for Online Social Networks. *Proceedings of the 3rd conference on online social networks*, pp. 8-8.
Schlegel, A., 2016, Games-Howell Post-Hoc Test. Rpubs, available at: https://rpubs.com/aaronsc32/games-howell-test .
Sirkeci, I., 2013, *Transnational marketing and transnational consumers*, Springer, New York, NY.
Stieglitz, S. and Dang-Xuan, L., 2013, Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*, Vol. 29 No. 4, pp. 217-248.
Suh, B., Hong, L., Pirolli, P. and Chi, E. H., 2010, Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. *2010 IEEE Second International Conference on Social Computing,* pp. 177-184.
Takhteyev, Y., Gruzd, A. and Wellman, B., 2012, Geography of Twitter networks. *Social networks,* Vol. 34 No. 1, pp. 73-81.
Tsugawa, S. and Ohsaki, H., 2017, On the relation between message sentiment and its virality on social media. *Social Network Analysis and Mining,* Vol. 7 No. 1, pp. 1-14.
Wang, F., Wang, H. and Xu, K., 2012, Diffusive logistic model towards predicting information diffusion in online social networks. *2012 32nd International Conference on Distributed Computing Systems Workshops*, pp. 133-139.
Wang, F., Wang, H., Xu, K., Wu, J. and Jia, X., 2013, Characterizing information diffusion in online social networks with linear diffusive model. *2013 IEEE 33rd international conference on Distributed Computing Systems*, pp. 307-316.
Yu, X., Reddy, D. S., Bentula, L., & Hong, L., 2020. Characteristics of information spreading across nations. *Proceedings of the Association for Information Science and Technology*, Vol. 57 No. 1, pp. 1-6.
Yu, X., Daida, S. R., Boy, J., & Hong, L., 2020, The Effect of Structural Affinity on the Diffusion of a Transnational Online Movement: The Case of# MeToo. *International Conference on Social Informatics*. Springer, Cham, pp. 447-460.