International Journal of Semantic Computing
Vol. 16, No. 3 (2022) 1–24

DOI: 10.1142/S1793351X22500052

Automated Hashtag Hierarchy Generation Using Community Detection and the Shannon Diversity Index, with Applications to Twitter and Parler

Spencer Torene*, Andrew Follmann, Thomas Teague,
Peter Chang and Blake Howald†

Thomson Reuters Special Services, LLC
1410 Spring Hill Road Suite 125
McLean, VA 22102, USA
**spencer.torene@trssllc.com*
†blake.howald@trssllc.com

Published

Developing semantic hierarchies from user-created hashtags in social media can provide useful organizational structure to large volumes of data. However, construction of these hierarchies is difficult using established ontologies (e.g. WordNet [C. Fellbaum (ed.), WordNet: An Electronic Lexical Database (MIT Press, Cambridge, MA, 1998)]) due to the differences in the semantic and pragmatic use of words versus hashtags in social media. While alternative construction methods based on hashtag frequency are relatively straightforward, these methods can be susceptible to the dynamic nature of social media, such as hashtags with brief surges in popularity. We drew inspiration from the ecologically based Shannon Diversity Index (SDI) [J. L. Wilhm, Use of biomass units in Shannon's formula, *Ecology* **49**(1) (1968) 153–156] to create a more representative and resilient method of semantic hierarchy construction that relies upon network-based community detection and a novel, entropy-based ensemble diversity index (EDI) score. The EDI quantifies the contextual diversity of each hashtag, resulting in thousands of semantically related groups of hashtags organized along a general-to-specific spectrum. Through an application of EDI to social media data (Twitter and Parler) and a comparison of our results to prior approaches, we demonstrate our method's ability to create semantically consistent hierarchies that can be flexibly applied and adapted to a range of use cases.

Keywords: Information entropy; semantics; ontology; social computing.

1. Introduction

Given the volume of users and content on social media (Instagram, Facebook, Twitter, etc.), it is necessary to rely on strategies to organize data relative to different analytical use cases (e.g. identifying networks, detecting trends). Organization around hashtags is one common method — for example, *#ai* in the following tweets indicates a topic:

> *US has announced The National Artificial Intelligence Initiative Office to regulate **#ai** research and policy.*

Rich is a Python library for rich text and beautiful formatting in the terminal. ***#AI*** *#DataScience #MachineLearning #DeepLearning.*

A community "topic" can be created around the *#ai* hashtag based on the frequency of co-occurring hashtags. For example, Fig. 1 shows a general-to-specific hierarchy of semantically similar relationships to *#ai* within the *#data* community — e.g. internet of things ("iot") is a specific field within AI, Microsoft is a company contributing to AI, and cloud technology is heavily used in machine learning.

Fig. 1. Top 10 most diverse hashtags in the Twitter #data community hierarchy — hashtags closer to the top are conceptually more general and hashtags closer to the bottom are conceptually more specific. The horizontal position is meaningless, see Sec. 2 for more information.

The actual processing of hashtag data (e.g. segmentation and normalization) is well understood (*see generally* [3–7]), but using hashtag-based structures (e.g. clustering [8]) in a range of NLP tasks (e.g. text classification by hashtag [9, 10]) has proven challenging. This is due in part to the inherent variation in the semantics, pragmatics, and users of hashtags [11]. There is no "wrong" time, place, or context to use a hashtag and user intent can vary — in addition to conveying topics [12], hashtags can be used for expression of emotions [13], sentiment [14], and named entities [15]. At best, hashtags are part of a *folksonomy* which does not have an established curated reference [16].

While a folksonomy bears some resemblance to a logical, semantically consistent taxonomy, it is an ephemeral reflection of how people use language within the context of social media. These reflections are too dynamic to establish a static, semantically coherent hierarchy that can be consistently relied upon for different analytical purposes. For example, #minnesota was a geographical reference to one of the 50 United States prior to the murder of George Floyd on May 25, 2020 in Minneapolis which sparked massive protests. After that date, #minnesota became more associated with the Black Lives Matter movement, rather than the geographic location, *per se*. In combination with automated tracking of semantic associations, determining the relative "generality" of each hashtag would also help establish a reference for a folksonomy through the creation of a hashtag hierarchy.

The state-of-the-art generality score for hashtags is the degree of co-occurrence with other hashtags [17, 18], defined as the number of unique hashtags co-occurring in the same tweets. Several studies show that simple measures such as degree centrality are an effective method of determining generality or abstractness, as compared to gold standard hierarchies such as WordNet [19] or probabilistic models [20]. While intuitive and useful, degree centrality can be overly influenced by popularity in social media. In our data, for example, #coronavirus is used at a rate 40 times higher than #virus and has 32 times higher degree centrality, even though *coronavirus* is a hyponym of *virus* (*see also* [21, 22]).

We expand on our previous work in [23] which re-framed and broadened the idea behind hashtag co-occurrence entropy [24] to account for additional features that contribute to the idea of generality, such as users, posting times, and semantic overlap. We rely on the same ecologically inspired Shannon Diversity Index (SDI) [2] as in [23], but with improved community detection metrics, additional measures of diversity, and an application of the method to Parler data.

As discussed in Sec. 2, we calculated the SDI (i.e. Shannon entropy) of each hashtag as a measurable proxy of generality for eight different features involving time, users, words/tokens, hashtags, and hashtag communities to help address how event-driven or community-specific that hashtag may be [25]. The resulting hierarchy is organized by "topics" (i.e. hashtag communities) and ordered by hashtag diversity. This hierarchy maintains all edges and does not join nodes (contra parent–child hierarchies), in order to preserve complex semantic structure. In Sec. 3, we demonstrate through quantitative and qualitative analyses that a sensible hierarchy can be automatically generated through community detection and the linear combination of eight measures of entropy (i.e. diversity) as one ensemble diversity index (EDI). We apply the method to Parler data in Sec. 4, and discuss the method and its limitations in Sec. 5.

2. Methods

Using Twitter data, we created a hierarchy of English language hashtags by first constructing an undirected hashtag network. Next, we performed community detection using the hashtag co-occurrence edge weights. We then calculated the SDI for the eight diversity measures of Hashtag Co-occurrence, Community, Month-of-Year, Year-and-Month, Day-of-Week, User, Hour, and Word/Token. Lastly, we calculated our novel EDI by linear weighted combination of the eight diversity measures.

August 24, 2022 9:41:53am WSPC/214-IJSC 2250005 ISSN: 1793-351X 2nd Reading

2.1. *Data*

We collected the random 1% sample Twitter "Spritzer" streams (a.k.a. "Sample" streams) from the Internet Archive (archive.org/details/twitterstream) for 52 months from October 2016–December 2021, with some monthly and daily exceptions due to archive.org availability or expedience.

We limited our analyses to English language tweets as indicated by metadata. We also rejected hashtags with non-Latin characters, which means we did not capture Unicode hashtags (e.g. emojis and many non-Latin language hashtags). Before any data cleaning, there were 92,037,572 English language tweets with 2–5 hashtags (we need at least two hashtags to measure co-occurrences and we limited to five hashtags to increase the judiciousness of hashtag selection). We performed several data cleaning steps to reduce the effect of duplicate and non-human data. First, we removed duplicate tweets (i.e. "retweets") and retrieved the user account and tweet text from the original tweet. Second we removed tweets from 95,918 purported bot accounts [26], to end up with 46,060,194 eligible tweets across 10,760,353 unique user accounts.

2.2. *Network*

We represented the hashtag data as an undirected network using the NetworkX Python package (v.2.6.3) [27]. Nodes of the network represent the unique, lower-cased hashtags encountered in the data set, and edges represent co-occurrences, with the edge weights representing the count of co-occurrences. Each hashtag node recorded the following data: the total number of tweets with the hashtag; the number of uses for each year-and-month combination; the number of uses by month of the year, the number of uses by each user account; the number of uses by day of the week; the number of uses by hour of the day, and the number of uses per word/token. These seven node metrics plus one edge metric are the bases for the eight measures of diversity discussed in Sec. 2.4.

After all hashtag data were inserted, pruning occurred in the following order: edges of weight less than 8 (approximately two co-occurrences per year), hashtag nodes used by only one user, and all disconnected nodes (i.e. nodes with degree 0). Pruning was intended to decrease processing time and increase folksonomy user agreement [24]. The final network contained 361,644 hashtag nodes and 1,363,566 undirected co-occurrence edges.

We report several statistics about the networks constructed from Twitter in Table 1. We computed each of these statistics on the undirected hashtag network using the corresponding functions as implemented in the NetworkX Python library.

2.3. *Community detection*

To organize the hashtags into semantically similar groups, we explored several community detection methods, including Louvain modularity, Greedy modularity, and asynchronous label propagation [28]. A difficult question in network community

Table 1. Twitter network statistics.

Statistic	Value
Number of connected components	16,813
Number of nodes	361,644
Number of edges	1,363,566
Number of weighted edges	78,521,400
Average degree	7.54
Average weighted degree	434.25
Average clustering coefficient	0.39
Density	2.09E−05

Notes: We considered the *undirected* version of the Twitter network for these statistics.

detection is evaluation — how do we know if the communities are good, let alone "correct"? Despite the fact that modularity scores indicated the Louvain and Greedy modularity algorithms outperformed the asynchronous label propagation algorithm, the two former methods resulted in extremely large, counterintuitive hashtag communities — the 10 largest communities for Louvain and Greedy modularity contained 186,169 (51.5%) and 219,770 (60.8%) of the hashtags, respectively. In contrast, asynchronous label propagation resulted in 34,824 communities, of which the 10 largest contained only 66,391 (18.4%) of the hashtags. We therefore investigated additional measures of community detection performance using an external validation set based on the synsets in WordNet [1] to determine whether the use of asynchronous label propagation was justified.

2.3.1. *Closed synsets and evaluation*

There are several methods for internal comparison between network communities, but external community evaluation is difficult without ground truth communities to compare against. To create an external validation set to view community detection, we leveraged the synsets in WordNet [1] using the NLTK Python library (v.3.6.5) [29] and created *closed synsets*, which consist of semantically consistent and unambiguous terms.

The process to create the closed synsets was: for each synset in WordNet, if that synset had two or more lemmas which could only be found within that synset, then extract and group those lemmas into a *closed synset*. We reduced the original 117,659 synsets and 148,730 lemma to 30,151 closed synsets containing 73,121 lemmas. We then intersected these lemmas with our hashtag set to come up with our testing set, which consisted of 1178 synset pairs, derived from 976 synsets containing 1965 lemmas.

We consider the closed synset pairs to contain terms semantically similar enough to warrant their grouping within the same communities as each other, and because the lemmas selected were not shared across synsets, there is less ambiguity whether any matching hashtags *should* be contained within the same community as each other. If the consistent and unambiguous terms in each synset are allocated to the same community in our Twitter network, we can be confident that the community detection algorithm is confirming some external index of similarity. To evaluate the community detection performance, we looked at the F-score [30] and Bootstrapped Community Assignment (BCA) ratio of the closed synset pairs and their community assignments.

For the F-score, we calculated True Positives as closed synsets where all lemmas were assigned to the same community, False Positives as a lemma assigned to the same community as another lemma which is not in the same closed synset, and False Negatives as closed synset pairs in which at least one lemma was in a different community. The F-score was then calculated as defined in [30], and provides a balance between joining semantically similar terms and separation of semantically different terms.

For the BCA ratio, we compared how many closed synset pairs were matched into the same communities, against a randomly generated partition of the hashtags into community sets, where each hashtag is randomly assigned to a community according to a probability distribution corresponding to the distribution of community sizes. We generated the random community sets 100 times for a stable estimate of the baseline score and then used the ratio of the actual proportion of correctly matched closed synset pairs to that random baseline proportion of correctly matched scores.

Based on the F-score and BCA ratio (see Table 2), we determined that asynchronous label propagation created the most semantically consistent communities of reasonable size — large enough to contain synset pairs and small enough to provide appropriate specificity. We used the *asyn_lpa_communities*() function from the NetworkX Python package to generate communities and provided a seed value of 1 for repeatable community assignments for our analyses.

Table 2. Twitter community evaluation of closed synsets.

Algorithm	Weight metric	n_{comm}	F-score	BCA ratio
Asyn-lpa	co-occur	34,824	**0.0060**	**95.8**
Louvain	co-occur	17,166	0.0015	14.8
Greedy modularity	co-occur	18,215	0.0003	10.6

Notes: The asynchronous label propagation (asyn-lpa) method produced the best F-score and BCA ratio for the closed synsets. Best scores are in bold.

2.4. *Shannon diversity index*

The SDI [2] for each of the six hashtag contexts is calculated via Shannon entropy:

$$H_s(X) = -\sum_{i=1}^{n} P(X_i)\log_2 P(X_i), \tag{1}$$

where i represents the context diversity feature h, c, m, y, d, u, r, w, where $P(X_i)$ is the probability of a hashtag co-occurrence with a specified feature:

(1) Hashtag Co-occurrence Diversity (h), co-occurring with another hashtag i in the same tweet. Higher values indicate a hashtag co-occurs with many other hashtags.

(2) Community Diversity (c), co-occurring with hashtags from community i. Higher values indicate a hashtag is more likely to occur with hashtags outside of its community. This could mean a hashtag is applicable to more topics.
(3) Month-of-Year Diversity (m), occurring during month i. Higher values indicate a hashtag is used year-round. Lower values mean a hashtag is applicable to fewer months of the year and could indicate seasonality.
(4) Year-and-Month Diversity (y), occurring during a specific Year-and-Month i. Higher values indicate a hashtag is consistently used, despite ongoing events in the world. Lower values could indicate an association with specific events.
(5) Day-of-Week Diversity (d), occurring on day of the week i. Higher values indicate a hashtag is applicable throughout the week. Lower values indicate a hashtag is more applicable to fewer days of the week.
(6) User Diversity (u), being used by user i. Higher values indicate a hashtag has been adopted by more users and has wider popularity and acceptance.
(7) Hour Diversity (r), occurring during hour i. Higher values indicate a hashtag is used throughout the day. Lower values indicate a hashtag is applicable to fewer hours of the day.
(8) Word/Token Diversity (w), occurring with word/token i in the same tweet. Higher values indicate a hashtag co-occurs with many other words/tokens.

2.5. *Ensemble diversity index*

The EDI is a linear combination of the eight SDI measures. Determining the appropriate weights of each SDI presents a challenge. We first attempted to learn appropriate weights for each SDI through relationship matching between hashtags and formal ontologies. This process yielded poor results, so we instead determined weights based on the amount of information provided by the SDI distributions through an Entropy Weight Method (EWM)-inspired process.

2.5.1. *Ontology-derived SDI weights*

We first attempted to learn appropriate weights for each SDI through gradient descent learning of same-community hashtags with corresponding hypernym–hyponym pair ranks found in four ontologies: ACM [31], Microsoft Concept Graph [32, 33], DBpedia [34], and WordNet [1]. The number of hashtag pairs found was 26, 3669, 2314, and 13,781, respectively. In all four cases, Day-of-Week Diversity was calculated as the highest weight (1.0), with other diversity measures considerably lower: Month-of-Year (mean 0.17, range 0.03–0.51), Hour-of-Day (mean 0.11, range 0.0–0.25), Year-and-Month (mean 0.02, range 0.0–0.07), and Word/Token (mean 0.004, range 0.0–0.02), with the remaining SDIs approaching 0.

The folly of a heavy reliance on Day-of-Week Diversity is demonstrated in Table 3: #havenlust was used only 14 times in our data — 2 times each day of the week — and does not represent a common or general idea. Moreover, the distribution of ontology-derived EDIs indicates relatively few specific hashtags and many

Table 3. Twitter hashtag diversity measures.

Diversity measure	Hashtag co-occurrence	Community	Month-of-year	Year-and-month
Max SDI	17.06	6.81	3.58	5.58
Hashtags with highest diversity	job	fanart	bootworship	twitch
	nsfw	cosplay	communist	webcomic
	love	netflix	carbonfootprint	birding
	free	rip	rocketleague	comic
	art	newprofilepic	laughter	ukhousing
Diversity measure	**User**	**Day-of-week**	**Hour-of-day**	**Word/token**
Max SDI	17.45	2.81	4.58	11.34
Hashtags with highest diversity	iheartawards	havenlust	listenlive	nonsenseengine
	bestfanarmy	question	onair	nonsense
	teenchoice	streetart	decoration	vss365
	mamavote	gay	hits	mvrp
	bbmas	nipples	nowplaying	feedly

Notes: The diversity measure row indicates the diversity feature, with the maximum value for any hashtag in that context shown in the max SDI row. Hashtags with highest diversity shows the hashtags with the top 5 SDI measures in that context, in descending order. Day-of-week has the smallest domain with seven possible options and user has the largest domain with 10,760,353 possible options.

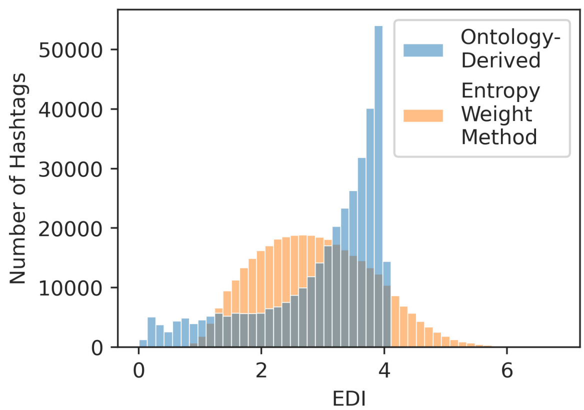

Fig. 2. (Color online) Ontology-derived SDI weights lead to a counter intuitive distribution of Twitter hashtag EDIs. The distribution of ontology-derived EDIs (blue) suggests relatively few specific hashtags (i.e. lower EDIs) and relatively many general hashtags (i.e. higher EDIs), with a sharp drop off above EDI = 4. In contrast, EDIs derived through the EWM (orange) form a smooth progression from relatively many specific hashtags to a longer tail of increasingly more general hashtags.

general hashtags (Fig. 2, blue). Finally, the top 10 most diverse hashtags in the ontology-derived hierarchy were counterintuitive: #openrp, #sixwordstory, #haiku, #bbc, #anime, #poem, #fastcast4u, #au, #lewdrp, and #poetry.

We therefore conclude that folksonomy hashtag usage patterns do not correspond well to the expected diversity of general, formal ontological terms and that SDI weights derived through relation matching folksonomies and ontologies do not generate reasonable hashtag hierarchies.

2.5.2. *Entropy weight method*

Because there is no mathematically rigorous way of learning appropriate weights for each diversity measure that does not include human judgment, we determined feature weightings via a process similar in spirit to the EWM [35–37]. In cases where feature weights would be derived subjectively, EWM can be used to objectively derive feature weights based on the reliability of the information provided by each feature. In our case, we used the Jensen–Shannon (JS) divergence of the distribution of each SDI to determine how informative the SDI would be. First, we transformed the SDI distribution via kernel density estimate, with rule-of-thumb bandwidth estimator $1.06\hat{\sigma}n^{\frac{-1}{5}}$, with n the number of samples and $\hat{\sigma}$ the empirical standard deviation [38]. Second, we calculated the JS divergence of the kernel density function from a uniform distribution (with support $[\min(H_i), \max(H_i)]$) and subtracted that value from 1. Finally, we divided each of these eight values by the sum of those eight values to derive the final SDI weightings (see Table 3).

The EWM weights (rounded for brevity) result in the following equation for the EDI:

$$\begin{aligned} H_\Sigma(i) = {} & 0.08H_h(i) + 0.08H_c(i) + 0.16H_m(i) + 0.17H_y(i) \\ & + 0.14H_d(i) + 0.12H_u(i) + 0.13H_r(i) + 0.11H_w(i). \end{aligned} \tag{2}$$

2.6. *Hierarchy evaluation*

There is no objective measure of the correctness of a folksonomy hierarchy, but our approach withstands scrutiny through theoretical argument, and qualitative and quantitative analyses. *Theoretically*, when a hashtag has a higher diversity measure for one of the eight entropy features, then by definition it was applied in more diverse circumstances than a hashtag with a lower diversity measure. *Qualitatively*, we visualize the resulting hierarchy for the #data hashtag, as well as its immediate community neighborhood. *Quantitatively*, we compare differences between the highest rank hashtags between the EDI hierarchy and a ranked degree co-occurrence hierarchy in a community-agnostic manner for more direct comparison.

Additional perspectives on a hashtag's generality include the character length of the hashtag and whether the hashtag has an English language counterpart. These two metrics are a good proxy of hashtag simplicity and acceptance, respectively, and we compare these statistics between the degree co-occurrence rankings and the EDI rankings. First, we created two rank-ordered sets of hashtags, sorted by degree co-occurrence and by EDI. Then we counted the number of characters in each hashtag and determined whether each word was a dictionary word.

To determine whether a hashtag was a dictionary word, it needed at least one synset representation in WordNet [1] using the NLTK Python library (v.3.6.5) [29]. Then we looked at the proportion of dictionary word hashtags in rank-decreasing order. Because many hashtags share identical, low degrees of co-occurrence, combining them into one rank would not adequately reveal their relative weight as compared to the fewer hashtags with higher degrees of co-occurrence. Therefore, we used a tumbling average of 1000 hashtags for proportion of degree rank dictionary word hashtags, which also represented the continuous rank representation of EDI with fidelity. We then determined the mean and median rank of English language dictionary words for EDI rank, and then of degree co-occurrence by performing 20 iterations (with mean values) of randomization of the degree rank among same-degree hashtags. In other cases where a direct comparison between the degree and EDI ranking was necessary, we were able to assign equivalent degree rankings to hashtags ordered by their EDI ranking. To accomplish this, we mapped the EDI-ordered list of hashtags to the ordered list of degree rankings and assigned the degree rank from the second list to the hashtags in the first list, so that there are an equal number of hashtags of each rank for each paradigm.

3. Results

361,644 hashtags were extracted from 46,060,194 tweets and assigned to 34,824 hashtag communities. By the EDI method, we find the top 10 most diverse hashtags to be #love, #art, #fanart, #india, #usa, #trump, #free, #youtube, #twitch, and #music, in decreasing order. We discuss the application of the EDI method to this Twitter data through the following quantitative and qualitative assessments: (1) the closeness of created communities at the individual and group level of community; (2) the expression of diversity in the hashtags; (3) comparisons of EDI to degree rank; and (4) other intrinsic observations.

3.1. *Semantic consistency*

Term co-occurrence is an indication of semantic "closeness" [39] and has proven useful in the context of hashtags [40]. We tested three community detection algorithms and found that asynchronous label propagation [28] outperformed both the Louvain [41] and Greedy modularity [42] algorithms using the co-occurrence edge weights (see Table 2). Asynchronous label propagation kept more of the closed synset lemmas from WordNet (see Sec. 2.3.1) within the same communities as each other, while balancing semantic separation between potentially incompatible inter-synset lemmas. The resulting EDI hashtag communities appear semantically consistent as demonstrated by the top (most diverse) and bottom (least diverse) five hashtags for four communities containing the seed hashtags #ai, #beer, #coffee, and #dogs (Table 4) — ostensibly, few hashtags are out of place.

Table 4. Twitter hashtag community examples.

Seed hashtag	#ai ($n = 2327$)	#beer ($n = 185$)	#coffee ($n = 189$)	#dogs ($n = 700$)
5 most diverse hashtags	data	**beer**	**coffee**	dog
	ai	craftbeer	tea	**dogs**
	microsoft	ipa	cafe	dogsoftwitter
	cloud	homebrew	coffeetime	pets
	learning	cider	mug	puppy
5 least diverse hashtags	ipisolutionsng	takecraftback	internationalteaday2021	boxerfirstlook
	5gultra	gabf2018	twoforme	germanshepherdtwitter
	instasharemod	50thandfrance	coffeeschool	bravewinstonrip
	wtmistakes	gabf2016	imsharing	findwombat
	memoriesizone	lagerlove	safetyrazors	zsparade2016

Notes: The seed hashtag was used to select the communities for display. n is the number of hashtags in the community. Each of the top five most diverse hashtags in these communities (top) appear related to the seed hashtag. Note the relative brevity of the most diverse hashtags, in contrast to the length of the least diverse hashtags (bottom).

We additionally relied on visualization to support this perspective on consistency.[a] In particular, Fig. 1 illustrates the #data hashtag community using the EDI scores. The most diverse hashtag in this community is #data, followed by #ai, #microsoft, #cloud, and #learning. At the very bottom of the #data community hierarchy (see bottom of Table 4) are extremely specific terms including #snatchword, #ipisolutionsng, #5gultra, #instasharemod, and #wtmistakes, which appear semantically relevant but narrowly applicable.

We also look at the #data hashtag community in the wider context of strongly connected neighbor communities (Fig. 3). The #data community is most strongly connected to the #art, #love, #usa, and #india hashtag communities, although it is also connected to many other communities (not shown, for clarity).

3.2. *Hashtag diversity*

We next looked at how the measures of diversity were represented by hashtags — Table 3 displays the top five most diverse hashtags for each feature:

(1) *job* is the hashtag that co-occurs most uniformly with all other 361,644 hashtags.
(2) *fanart* is the most uniformly co-occurring hashtag with all 34,824 hashtag communities.
(3) *bootworship* is the most uniformly distributed hashtag across all 12 months of the year.

[a] We wrote a custom anti-gravity + spring simulation visualization which allowed free movement on the x-axis, but locked nodes on the y-axis according to their EDI. For each community, nodes were placed randomly on the x-axis. The anti-gravity + spring simulation then moved the nodes along the x-axis only, according to the anti-gravity and spring forces generated by neighboring nodes within the same community.

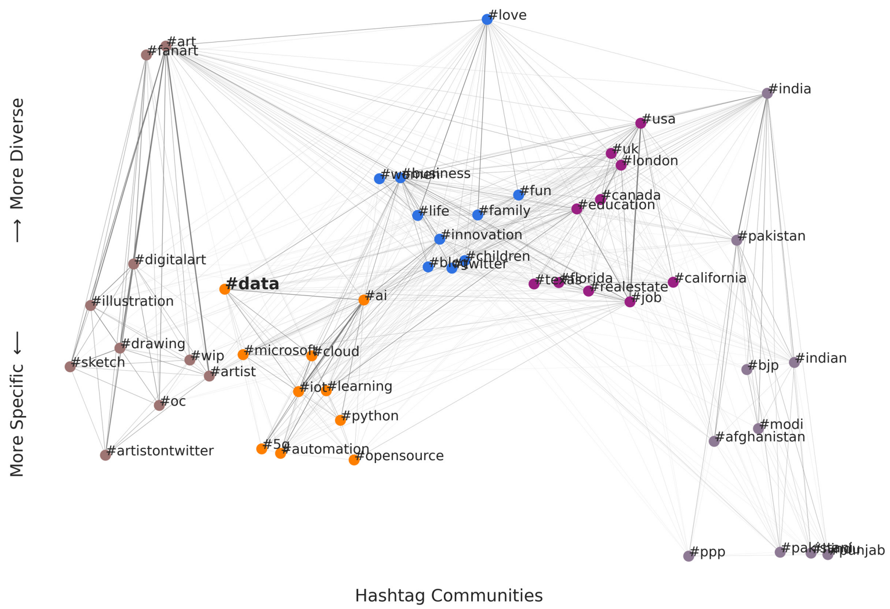

Fig. 3. Twitter #data community neighborhood. Shown are the top 10 most diverse hashtags in the communities most strongly connected to the #data community, as determined by co-occurrence weights. Similar to Fig. 1, hashtags closer to the top are more diverse and lower hashtags are less diverse. The horizontal position is meaningless, except for a hashtag's proximity to other hashtags within the same community.

(4) *twitch* is the most uniformly distributed hashtag across the 52 year-and-month combinations.
(5) *iheartawards* is the most uniformly distributed hashtag across the 10,686,214 users.
(6) *havenlust* is the most uniformly distributed hashtag across the 7 days of the week.
(7) *listenlive* is the most uniformly distributed hashtag across the 24 h of the day.
(8) *nonsenseengine* is the most uniformly distributed hashtag across all 2,856,232 words/tokens.

Broadly, the more uniform a distribution is over a larger space of possibilities, the higher the diversity measure. Hashtag popularity can certainly increase the diversity measure across all the eight measures due to more frequent usage. The eight features were specifically included, however, to balance popularity, trendiness, and topicality.

3.3. *Degree rank comparisons*

Although the ordering of hashtags by EDI appears reasonable, we investigated the comparison of the top 20 most diverse hashtags to the top 20 hashtags from the degree ranking method and found an overlap of 9, indicating some agreement with

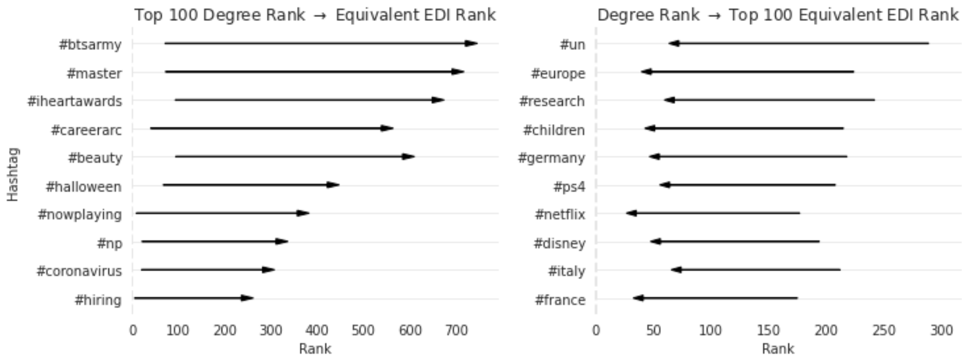

Fig. 4. Largest rank changes *out of* the top 100 degree ranks and *into* the top 100 EDI ranks for Twitter. Left: The hashtags with the largest drops *out of* the top 100 degree ranks are more associated with seasonal and newsworthy events, and consist of multiple words. Right: The hashtags with the largest rises *into* the top 100 EDI ranks.

the state of the art. For rank changes involving significant hashtags as a result of using the EDI, we looked at the largest Degree-to-EDI moves of the top 100 Degree Rank and the top 100 EDI Rank hashtags. First, we found the largest rank changes from the top 100 Degree Rank hashtags to their EDI Rank (Fig. 4, left). Second, we found the largest rank changes of hashtags in the top 100 EDI Rank hashtags from their Degree Rank (Fig. 4, right). As statistically expected, the biggest movers of the top 100 Degree Rankings were all decreases in rank and the biggest movers of the top 100 EDI Rankings were all increases.

Many of the largest Top 100 Degree Rank decreases are either compound terms associated with narrow — though noteworthy — entities or events that will gradually fade from public discourse. On the other hand, many of the largest Degree Rank increases into the top 100 EDI Ranks were shorter terms or potentially have more permanence.

An interesting feature of these two plots is that the character lengths of the rank-decreasing hashtags is higher than the character lengths of the rank-increasing hashtags (Fig. 4, left versus right, mean 7.8 versus 5.8). We wondered if the movement of longer hashtags to lower EDI Ranks and shorter hashtags to higher EDI Ranks was a broader trend. To examine this possibility, we took the mean character length of windows of 1000 hashtags in decreasing rank order and found that the mean character length of higher EDI Rank hashtags was lower than higher Degree Rank hashtags (Fig. 5, left). Conversely, the mean character length of lower EDI Rank hashtags was higher than lower Degree Rank hashtags. This suggests a more natural progression of the complexity of character combinations for EDI rankings than degree rankings.

We also wondered if a higher frequency of dictionary word hashtags (see Sec. 2.6) would be an outcome of the shorter hashtags, in case the increased length of certain hashtags was due to modifications of a base hashtag such as dating (e.g. #nbafinals

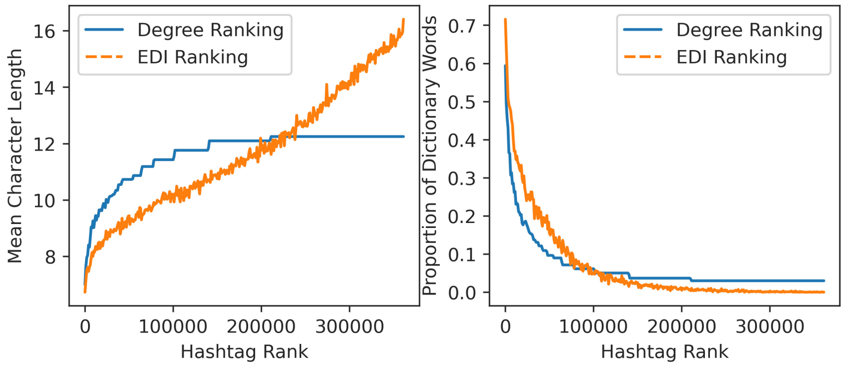

Fig. 5. (Color online) Emergent properties due to rank, for Twitter. Left: The mean character length of higher degree ranked hashtags (blue line) rises much faster and levels out much sooner than EDI ranked hashtags (orange line). The mean character length of the degree ranked hashtags flattens out around rank 150,000, while the mean character length of EDI ranked hashtags continues to increase almost throughout the entire 361,644 term folksonomy. Right: The proportion of dictionary word hashtags is larger for higher EDI ranked hashtags than higher degree ranked hashtags.

versus #nbafinals2019) or concatenation (e.g. #art versus #digitalart). About 6.38% of all hashtags in the network were dictionary word hashtags. The mean and median ranks of dictionary word hashtags for EDI (mean = 46,883.9, median = 31,229) and degree co-occurrence (mean $\sim$ 105,951.4, median $\sim$ 63,721) indicate that dictionary word hashtags are skewed towards higher (i.e. more general) ranks with EDI than with degree co-occurrence. This suggests a more natural progression of dictionary to non-dictionary word hashtags for EDI rankings.

Hashtags not found in the dictionary are more likely to consist of phrases or invented words that are applicable to a narrower set of situations. This is in contrast with dictionary word hashtags, which consist of words useful enough to be found in a dictionary. That these two trends of character length and dictionary word progression of the EDI hierarchy emerge despite not being accounted for in the generation process is additional evidence that the EDI-based hierarchy is reasonable.

4. Application to Parler

We claim that our hashtag hierarchy (i.e. a hashtag network with EDI measures) is not restricted to Twitter, and it can be applied to other multi-label data, especially social tagging data. We identified Parler and its Post data collected by [43] to demonstrate this generalizability, as it is a similarly structured data set with a few key differences in the topic community topology and hashtag usage patterns.

The Parler micro-blogging social network was explicitly created as an alternative to Twitter, appealing to unmoderated free speech for those users who either were de-platformed from Twitter, or found the content moderation on Twitter to be too restrictive. This self-selection narrows the scope of Twitter in users, communities, and content.

4.1. *Parler data*

Our Parler data extends from August 2018 through January 2021. We focused on the unique Posts ("tweets" in Twitter parlance) for our analysis, excluding the Echos ("retweets") and Comments ("threads") to mimic the Twitter dataset. We restricted the data in a similar logic to that on Twitter: we kept posts with between two and five hashtags, hashtags with four or more uses (at least two uses per year of data), and excluded hashtags used by only one user account. From the original 14,344,415 posts, we ended with 832,426 posts from 80,543 unique user accounts, with most of the reduction in size due to the imposed 2–5 hashtag limit (6.5% of total).

The Parler hashtag network statistics (Table 5) were computed similar to Twitter (Table 1). We observe that the average clustering coefficient and average degree measures are similar between the two networks, but that the Parler network is an order of magnitude more dense. Each of the Parler and Twitter networks consist of a single, large connected component and many smaller connected components. We believe this phenomenon is a result of our preprocessing and culling operations.

The frequency of hashtag use on Twitter versus Parler is substantially different (see Fig. 6), with hashtags deployed relatively sparsely on Twitter. Some of this difference can be clearly linked to platform level differences across the two microblogging sites, including the substantially higher character limitation of Parler ($3.5\times$ Twitter's limit) and the lack of full-text post search [44]. The character limit could allow "more space for more topics" and be a relatively linear shift in usage pattern. However, the primacy of hashtags for search would likely induce some nonlinear behavioral changes in hashtag use, as the marginal utility for a hashtag is much higher on Parler than on Twitter. Parler is also known to contain a high number of automated accounts and accounts that promote commercial off-site content [45], which could lead to a further increase in the length of and number of hashtags used in a typical post when combined with the platform's relatively permissive moderation policy.

Table 5. Parler network statistics.

Statistic	Value
Number of connected components	21
Number of nodes	42,309
Number of edges	1,663,986
Number of weighted edges	5,053,070
Average degree	78.66
Average weighted degree	238.87
Average clustering coefficient	0.42
Density	9.30E−04

Notes: We considered the *directed* version of the Parler network for these statistics. As noted in a previous section, we computed statistics for the *undirected* version of the Twitter network (Table 1). These choices reflect the nature of the algorithms we applied to each network.

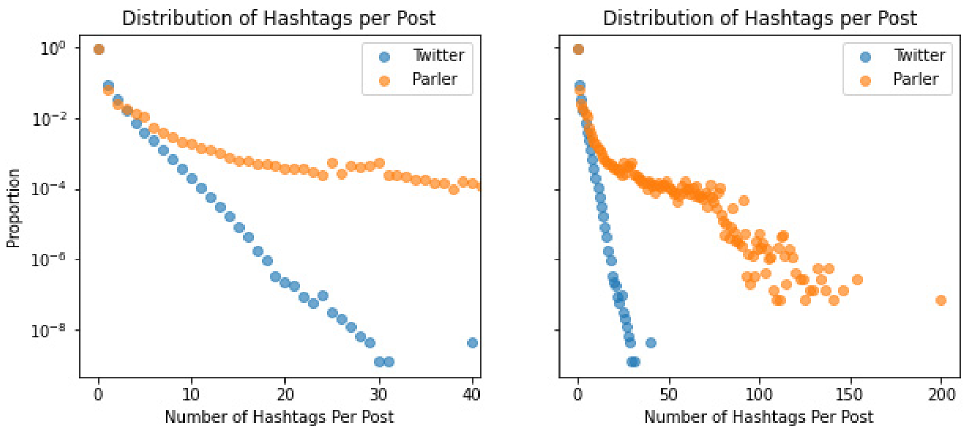

Fig. 6. Note the y-axis is log10-transformed to illustrate the difference in the long tail of hashtag usage. Left: The x-axis is constrained for detail. Right: Complete distribution from our data.

4.2. *Parler community detection*

Using the same community detection process as with Twitter (i.e. asynchronous label propagation using the co-occurrence edge weights) for the Parler network presented additional challenges. Possibly due to the increased density of the Parler network over the Twitter network (compare Table 5 to Table 1), 378 communities were generated, including one community with 39,662 of the 42,309 hashtags (93.7%). To create smaller, more meaningful communities, we conducted a community detection grid search of algorithms and weighting schemes.

We explored several alternative edge weighting schemes to co-occurrence on the Parler network, detailed as follows:

(1) **Co-occurrence** is the number of co-occurrences of hashtags within posts, identical to the Twitter community detection process. [Undirected edge weight].
(2) **Co-occurrence proportion** is the proportion of a given node's co-occurrence edge weights, defined as $P(Y|X)$. [Directed].
(3) **Sqrt co-occurrence** is the square root of the co-occurrence weights. [Undirected].
(4) **$\log_2$ co-occurrence** is the $\log_2$ of the co-occurrence weights. [Undirected].
(5) **Mutual information** is a measure of the dependence between two hashtags, defined as $P(X, Y) * \log_2(P(X, Y)/(P(X) * P(Y)))$. [Undirected].
(6) **Support** is the Bayesian support of one hashtag for another, defined as $\frac{P(Y|X)}{P(Y)}$, representing the support that Y provides for X. [Directed].

Overall, we find the Bayesian *support* weight to perform the best according to the F-score and BCA ratio (see Sec. 2.3.1 for a description of these methods, and Table 6 for performance metrics). This is likely due to the many small communities that are

Table 6. Parler community detection performance.

Algorithm	Weight metric	n_{comm}	F-score	BCA ratio
Asyn-lpa	co-occur	378	0.0050	1.1161
	co-occur_prop.	376	0.0050	1.1188
	sqrt(co-occur)	201	0.0048	1.0211
	$\log_2$(co-occur)	2045	0.0053	1.2206
	mutual_information	613	0.0051	1.1411
	support	9269	**0.1975**	**727.2727**
Louvain	co-occur	181	0.0161	4.7204
	co-occur_prop.	52	0.0284	5.3945
	sqrt(co-occur)	34	0.0219	4.3288
	$\log_2$(co-occur)	10,818	0.0379	8.9198
	mutual_information	140	0.0334	8.5185
	support	272	0.0948	29.7125
Greedy modularity	co-occur	204	0.0197	4.5917
	co-occur_prop.	182	0.0301	5.9777
	sqrt(co-occur)	193	0.0120	2.7153
	$\log_2$(co-occur)	11,129	0.0202	5.4848
	mutual_information	207	0.0313	7.7966
	support	292	0.0855	26.0352

Notes: Metrics for Parler community detection performance using three algorithms and six different weighting schemes. The best scores were reached with asynchronous label propagation using the Bayesian support weighting scheme (scores in bold). Notably, the Bayesian support weights repeatedly produced the highest scores within each of the three community detection algorithms.

formed versus the fewer, badly skewed community sizes generated with the straightforward co-occurrence weights.

4.3. *Parler results*

We calculated the SDI measures and EDI weights identically to the Twitter data. Shown in Table 7 are the five hashtags with the highest SDI values for each of the eight SDIs. #leftists is the hashtag that co-occurs most evenly with other hashtags, #democrats co-occurs most evenly with the 9269 hashtag communities, #justice is used most evenly for each month of the year, #alexjonesshow is used most evenly for the 26 Year-and-Month combinations of our data, #trump2020 is used most evenly by the 80,543 user accounts, #letthemlive is used most evenly for each day of the week, #saudi is used most evenly for each hour of the day, and #parlerksa is used most evenly with the 326,103 words/tokens in the Parler data.

Calculating weights for the SDI measures using the same EWM-inspired method as for Twitter gives the following EDI equation for the Parler hashtags (rounded for brevity):

$$\begin{aligned} H_{\Sigma}(i) = {} & 0.12H_h(i) + 0.13H_c(i) + 0.14H_m(i) + 0.14H_y(i) \\ & + 0.13H_d(i) + 0.11H_u(i) + 0.14H_r(i) + 0.10H_w(i). \end{aligned} \tag{3}$$

Table 7. Parler hashtag diversity measures.

Diversity measure	Hashtag co-occurrence	Community	Month-of-year	Year-and-month
Max SDI	20.95	9.07	3.55	4.29
Hashtags with highest diversity	leftists	democrats	justice	alexjonesshow
	family	america	egardwatches	illegalalien
	america	leftists	jesusfollower	dem
	democrats	california	news	davidknightshow
	usa	christmas	maryamrajavi	warroomshow
Diversity measure	**User**	**Day-of-Week**	**Hour-of-Day**	**Word/Token**
Max SDI	12.12	2.81	4.58	11.48
Hashtags with highest diversity	trump2020	letthemlive	saudi	parlerksa
	stopthesteal	teamfollowback	parlerksa	ksa
	trump	covidfarce	ksa	saudi
	covid19	high	saudiparler	ccp_is_terrorist
	twitter	psychedelics	saudiarabia	parler

Notes: The diversity measure row indicates the diversity feature, with the maximum value for any hashtag in that context shown in the max SDI row. Hashtags with highest diversity shows the hashtags with the top five SDI measures in that context, in descending order. Day-of-week has the smallest domain with seven possible options and word/token has the largest domain with 326,103 possible options.

Qualitative inspection reveals that the communities generated by the Bayesian *support* mechanism are reasonable (see Table 8). Within each of four communities seeded by the #covid19, #data, #ml, and #nyc hashtags, we see that the majority of the most and least diverse hashtags appear semantically relevant to each other.

Using a similar process to the Twitter hierarchy generation (except for community detection, see Sec. 4.2 for details), we again find a reasonable hierarchy of hashtags. The top 10 most diverse hashtags in Parler are #america, #democrats,

Table 8. Parler hashtag community examples.

Seed hashtag	#covid19 ($n = 10$)	#data ($n = 11$)	#ml ($n = 40$)	#nyc ($n = 23$)
5 most diverse hashtags	**covid19**	**data**	blockchain	newyork
	pandemic	datascience	programming	**nyc**
	covidlockdowns	impeachthis	website	cuomo
	coronalockdown	trumppencelandslide2020	digitalmarketing	newyorkcity
	pandemicpanic	chrismatthews	wordpress	deblasio
5 least diverse hashtags	covidpanic	dataleak	webdev	freedomtoworship
	coronalockdowns	ironcurtain	benford	comradedeblasio
	covidpolice	dataharvesting	backend	newyorkexodus
	tiers	analytics	html	trumpfacts
	pandemiclockdowns	chatbots	vuejs	billdeblasiomustgo

Notes: The seed hashtag was used to select the communities for display. We chose seeds in communities with at least 10 hashtags. n is the number of hashtags in the community. Most hashtags appear semantically related within their communities.

#usa, #trump, #fakenews, #covid19, #deepstate, #truth, #democrat, and #trump2020, in decreasing order.

While Twitter certainly has a significant contingent of political tweets, Parler's hashtags appear proportionally much more political, and despite our efforts not to highlight this (as it was not the purpose of our work), we see how political the community neighborhood around the #data hashtag is (cf. Fig. 3). Nonetheless, we again see that familiar, broader terms were considered more diverse (i.e. at the top of the hierarchy) and that less familiar, more specific terms were considered less diverse (i.e. at the bottom of the hierarchy). Separating the hashtags into specific, meaningful communities was difficult, given the increased frequency of hashtag use and co-occurrence, and the increased focus on politics — with a relatively consistent viewpoint. Our use of the Bayesian *support* edge weighting scheme, however, appears to have performed reasonably well and communities appear semantically consistent (see Fig. 7).

As with Twitter, we compared the relative rankings of dictionary words between EDI and degree co-occurrence. We again found that the mean and median positions of dictionary word hashtags for EDI (mean = 16,658.9, median = 14,670) were skewed to higher ranks than degree co-occurrence (mean ∼ 17,258.5, median ∼ 15,493.5), suggesting a more natural progression of dictionary to non-dictionary word hashtags for EDI rankings. Parler has a much higher proportion of dictionary

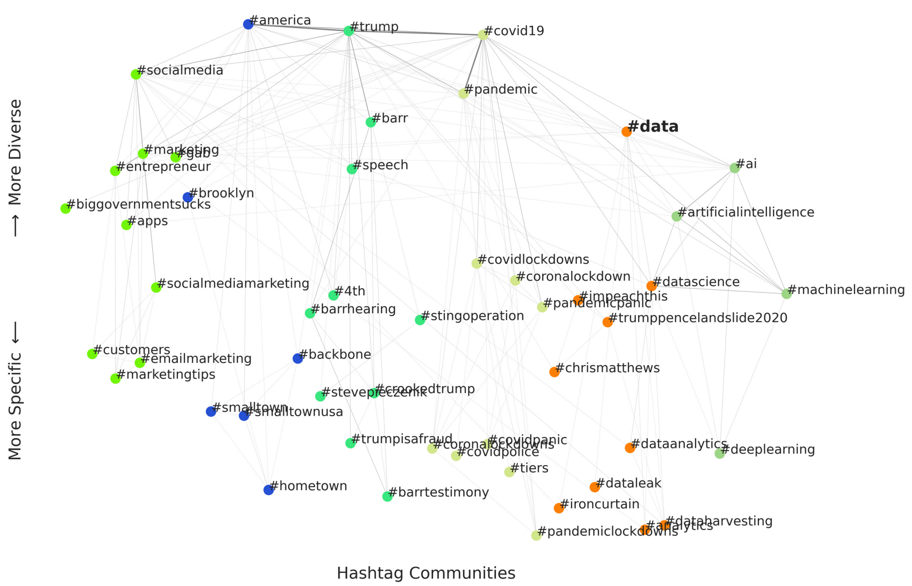

Fig. 7. Parler #data community neighborhood. Shown are the top 10 most diverse hashtags in the communities most strongly connected to the #data community, as determined by co-occurrence weights. Hashtags closer to the top are more diverse and lower hashtags are less diverse. The horizontal position is meaningless, except for a hashtag's proximity to other hashtags within the same community.

word hashtags than Twitter (22.5% versus 6.38%), but what downstream effect this has on either hierarchy construction or evaluation is unclear.

Conducting the same automated process as with Twitter, with only one modification to account for the increased density of the hashtag network, appeared to produce a reasonable hashtag hierarchy and demonstrates the generality of the method.

5. Discussion and Limitations

Our conception of Shannon entropy as a marker of generality is similar to previous work [24, 46], and the approach of linearly combining measures of entropy like our EDI measure has been considered before [47]. We do not make any assumptions about the relationships represented through co-occurrences, but rather rely on the predictability of a hashtag's context. The basic question being answered with our process is "given a hashtag, what can we guess about other features of the tweet?" More diverse (i.e. general) hashtags will have a larger space of possibilities, while less diverse (i.e. specific) hashtags will have a smaller space of possibilities. Given this question, another possible, subtly different perspective on hashtag generality could be through actual prediction of the source context using (e.g.) deep learning — more general hashtags would be less predictive than more specific hashtags.

Measures of entropy prevent infrequent hashtags from being considered diverse. Notably, this could mean that a general English language word used infrequently as a hashtag would not be considered diverse. Words have multiple definitions, and not all of those definitions may be adopted by users in a folksonomy. Definitions and associations may even be generated or altered on a social platform (for example, #minnesota, as discussed in Sec. 1). What might be considered general in the official language may not be general in the folksonomy, and vice versa. For example, the #rt hashtag is prevalent on Twitter and should be considered broadly applicable to many different contexts, despite its narrow semantics — it represents a request of its reader to retweet, which is a specific action that has a very limited use case in the English language.

5.1. *Limitations*

There are a few important limitations of our study. The first limitation concerns our formulation of the EDI. We chose to linearly combine the eight SDI measures, because two issues arose during consideration of a more rigorous calculation of their joint entropy. First, given that we only have a 1% sample of the data feed, we are likely to have mostly unique date + user + hashtag combinations. The average probability $\hat{P}(x_1, \ldots, x_8)$ for each combination of hashtag, year/month, month, user, weekday, community, hour, and tokens is very likely to approach the uniform distribution $\frac{1}{N}$, where N is the number of instances of a hashtag. Second, under a joint probability calculation, the distinction between hashtag use by Year-and-Month (i.e. an "event"), month of the year (i.e. "seasonality"), and day of the week would be

August 24, 2022 9:42:51am WSPC/214-IJSC 2250005 ISSN: 1793-351X 2nd Reading

removed. For example, it would have been impossible to separate the month from the year-and-month in a joint entropy calculation. This would have eliminated the possibility of understanding the seasonality of a hashtag mostly recurring during the same few months every year, as distinct from an event-related hashtag that appeared over the same number of months in one continuous block.

The second limitation concerns choices that affect large aspects of the generated hierarchy, including which measures of diversity to include, how to weight the measures, the method of community detection, and the time frame of data used. If any of these choices were different, the resulting hierarchy would change. Which measures of diversity we include and how we weight them would have obvious influence on the hierarchy. Additional measures could include the geographic location in which a hashtag was used and the use of hashtags within *user* communities (in contrast to the existing *hashtag* communities). Geographic location might distinguish between different dialects and regional terms from the same language, and user communities could distinguish between broad and narrow popularity.

Additional limiting observations include the fact that our weighting is not the One True Weighting; we present a modification to an objective weighting method used in cases when objectivity is difficult [35–37]. We cannot learn weights from a true (unknown) hierarchy, nor from curated semantic ontologies (see Sec. 2.5.1). Further, human judgement of hashtag generality could be wrong unless the person is an extremely heavy user of Twitter and is fluent in Twitterspeak [39], but there would be no guarantee of this given the highly dynamic nature of social media.

We found asynchronous label propagation created the most semantically consistent hashtag communities, but it is not a deterministic process and other methods of community detection would produce different — and potentially more semantically consistent — communities. We inspected many other communities than presented here, and while the vast majority of communities appear semantically consistent, the largest communities (with thousands of hashtags) contain many of the most frequent hashtags, due to strong co-occurrences, and can appear less topical. While we provided a weighting scheme that reduced the size of huge, non-specific communities, other methods may also reduce the impact of the strong co-occurrence of the most popular hashtags, which could lead to more semantic consistency.

The time frame of the data used to construct the hierarchy also affects the outcome, as term semantics in the folksonomy drift over time. This illustrates the advantage of a fully automated hierarchy generation method like EDI, compared to any process with manual effort. As events unfold, hashtags will assume a variety of different semantics and it is difficult and time consuming for experts to track those changes and appropriately place them within a hierarchy containing hundreds of thousands or millions of hashtags.

6. Conclusion

We demonstrated an automatic hashtag hierarchy by translating Shannon's Diversity Index into a mathematical definition of hashtag diversity and applied the method to both Twitter and Parler with reasonable results. Our EDI considers eight different measures of diversity which are linearly combined for a more holistic view of a hashtag's diversity and how applicable it is to different contexts. While hashtag hierarchies based on co-occurrence alone may be simpler to compute, they are more a representation of the data as it is at a given point in time, and any alignment with established semantic hierarchies is likely to be coincidental and require additional processing (e.g. using cosine similarity to collapse related nodes). Our method provides a way to not only form a more resilient hierarchy, but a framework for adaptation either through adjustment of weights, or the inclusion (or subtraction) of different diversity measures.

Further research will focus on refining the measures associated with community detection and working through more quantifiable comparisons to track improvement. We will also apply hashtag hierarchies to time-series analyses to understand change over time and training data augmentation for tasks such as membership prediction and classification of text by hashtag.

Acknowledgments

We thank Berk Ekmekci and Ian Coffman for extensive feedback on early drafts, and Temidayo Akinsanya for helpful conversations. Thanks are also due to anonymous reviewers from the Association for Computational Linguistics and IEEE-ICSC.

References

[1] C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database* (MIT Press, Cambridge, MA, 1998).
[2] J. L. Wilhm, Use of biomass units in Shannon's formula, *Ecology* **49**(1) (1968) 153–156.
[3] T. Declerck and P. Lendvai, Processing and normalizing hashtags, in *Proc. Int. Conf. Recent Advances in Natural Language Processing*, 2015, pp. 104–109.
[4] M. Maddela, W. Xu and D. Preoţiuc-Pietro, Multi-task pairwise neural ranking for hashtag segmentation, in *Proc. 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2538–2549.
[5] A. Çelebi and A. Özgür, Segmenting hashtags using automatically created training data, in *Proc. Tenth Int. Conf. Language Resources and Evaluation*, 2016, pp. 2981–2985.
[6] X. Wang, F. Wei, X. Liu, M. Zhou and M. Zhang, Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach, in *Proc. 20th ACM Int. Conf. Information and Knowledge Management*, 2011, p. 1031–1040.
[7] C. Simeon and R. Hilderman, Evaluating the effectiveness of hashtags as predictors of the sentiment of tweets, in *Discovery Science*, eds. N. Japkowicz and S. Matwin (Springer International Publishing, Cham, 2015), pp. 251–265.
[8] D. Gromann and T. Declerck, Hashtag processing for enhanced clustering of tweets, in *Proc. Int. Conf. Recent Advances in Natural Language Processing*, 2017, pp. 277–283.

[9] Y. Gong, Q. Zhang and X. Huang, Hashtag recommendation using Dirichlet process mixture models incorporating types of hashtags, in *Proc. 2015 Conf. Empirical Methods in Natural Language Processing*, 2015, pp. 401–410.
[10] Y. Li, T. Liu, J. Jiang and L. Zhang, Hashtag recommendation with topical attention-based LSTM, in *Proc. COLING 2016, 26th Int. Conf. Computational Linguistics: Technical Papers*, 2016, pp. 3019–3029.
[11] M. Ames and M. Naaman, Why we tag: Motivations for annotation in mobile and online media, in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 2007, pp. 971–980.
[12] O. Ozdikis, P. Senkul and H. Oguztuzun, Semantic expansion of hashtags for enhanced event detection in twitter, in *Proc. VLDB 2012 Workshop on Online Social Systems*, 2012.
[13] M. Abdul-Mageed and L. Ungar, EmoNet: Fine-grained emotion detection with gated recurrent neural networks, in *Proc. 55th Annual Meeting of the Association for Computational Linguistics: Long Papers*, Vol. 1, 2017, pp. 718–728.
[14] S. Mohammad, S. Kiritchenko and X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets, in *Second Joint Conf. Lexical and Computational Semantics (*SEM), Proc. Seventh Int. Workshop on Semantic Evaluation*, Vol. 2, 2013, pp. 321–327.
[15] A. Ritter, S. Clark, Mausam and O. Etzioni, Named entity recognition in tweets: An experimental study, in *Proc. 2011 Conf. Empirical Methods in Natural Language Processing*, 2011, pp. 1524–1534.
[16] T. Vander Wal, vanderwal.net (2007).
[17] D. Benz, A. Hotho and G. Stumme, Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge, in *Proc. 2nd Web Science Conf.*, 2010.
[18] P. Heymann and H. Garcia-Molina, Collaborative creation of communal hierarchical taxonomies in social tagging systems, Stanford InfoLab Technical Report (2006).
[19] D. Benz, C. Körner, A. Hotho, G. Stumme and M. Strohmaier, One tag to bind them all: Measuring term abstractness in social metadata, in *The Semantic Web: Research and Applications*, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 6643 (Part 2), 2011, pp. 360–374.
[20] M. Strohmaier, C. Körner and R. Kern, Understanding why users tag: A survey of tagging motivation literature and results from an empirical study, *J. Web Semant.* **17** (2012) 1–11.
[21] F. Almoqhim, D. E. Millard and N. Shadbolt, Improving on popularity as a proxy for generality when building tag hierarchies from folksonomies, in *5th Int. Conf. Social Informatics*, 2014, pp. 95–111.
[22] L. Nie, Y. Zhao, X. Wang, J. Shen and T.-S. Chua, Learning to recommend descriptive tags for questions in social forums, *ACM Trans. Inf. Syst.* **32**(1) (2014) 1–23.
[23] S. Torene and B. Howald, Automated hashtag hierarchy generation using community detection and the Shannon diversity index, in *2022 IEEE 16th Int. Conf. Semantic Computing*, 2022, pp. 59–66.
[24] P. Heymann, D. Ramage and H. Garcia-Molina, Social tag prediction, in *Proc. 31st Annual Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2008, pp. 531–538.
[25] W. Feng and J. Wang, We can learn your #hashtags: Connecting tweets to explicit topics, in *Proc. Int. Conf. Data Engineering*, 2014, pp. 856–867.
[26] M. Sayyadiharikandeh, O. Varol, K. Yang, A. Flammini and F. Menczer, Detection of novel social bots by ensembles of specialized classifiers, *CoRR*, arXiv:abs/2006.06867.
[27] A. A. Hagberg, D. A. Schult and P. J. Swart, Exploring network structure, dynamics, and function using NetworkX, in *Proc. 7th Python in Science Conf.*, 2008, pp. 11–15.

[28] U. N. Raghavan, R. Albert and S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Phys. Rev. E*, *Stat. Nonlinear Soft Matter Phys.* **76**(3) (2007) 036106.
[29] S. Bird, E. Loper and E. Klein, *Natural Language Processing with Python* (O'Reilly Media Inc., 2009).
[30] X. Liu, H.-M. Cheng and Z.-Y. Zhang, Evaluation of community detection methods, *IEEE Trans. Knowl. Data Eng.* **32** (2018) 1736–1746.
[31] The 2012 ACM computing classification system, http://www.acm.org/about/class/2012.
[32] Z. Wang, H. Wang, J.-R. Wen and Y. Xiao, An inference approach to basic level of categorization, in *Proc. 24th ACM Int. Conf. Information and Knowledge Management*, 2015, pp. 653–662.
[33] W. Wu, H. Li, H. Wang and K. Zhu, Probase: A probabilistic taxonomy for text understanding, in *Proc. 2012 ACM SIGMOD Int. Conf. Management of Data*, 2012, pp. 481–492.
[34] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer and C. Bizer, DBpedia — A large-scale, multilingual knowledge base extracted from wikipedia, *Semantic Web* **6**(2) (2015) 167–195.
[35] L. Liu, J. Zhou, X. An, Y. Zhang and L. Yang, Using fuzzy theory and information entropy for water quality assessment in three Gorges region, China, *Expert Syst. Appl.* **37**(3) (2010) 2517–2521.
[36] X. Li, K. Wang, L. Liu, J. Xin, H. Yang and C. Gao, Application of the entropy weight and TOPSIS method in safety evaluation of coal mines, *Procedia Eng.* **26** (2011) 2085–2091.
[37] Y. He, H. Guo, M. Jin and P. Ren, A linguistic entropy weight method and its application in linguistic multi-attribute group decision making, *Nonlinear Dyn.* **84**(1) (2016) 399–404.
[38] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Routledge, 2018).
[39] W. L. Woon and S. E. Madnick, Asymmetric information distances for automated taxonomy construction, *Knowl. Inf. Syst.* **21**(1) (2009) 91–111.
[40] V. S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini and F. Menczer, The DARPA twitter bot challenge, *Computer* **49**(6) (2016) 38–46.
[41] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech.*, *Theory Exp.* **2008** (2008) P10008.
[42] A. Clauset, M. E. J. Newman and C. Moore, Finding community structure in very large networks, *Phys. Rev. E*, *Stat. Nonlinear Soft Matter Phys.* **70** (2004) 066111.
[43] M. Aliapoulios, E. Bevensee, J. Blackburn, E. D. Cristofaro, G. Stringhini and S. Zannettou, An early look at the Parler online social network, *CoRR*, arXiv:abs/2101.03820.
[44] Hitkul, A. Prabhu, D. Guhathakurta, J. Jain, M. Subramanian, M. Reddy, S. Sehgal, T. Karandikar, A. Gulati, U. Arora, R. R. Shah and P. Kumaraguru, Capitol (Pat)riots: A comparative study of twitter and parler, arXiv:2101.06914.
[45] D. Thiel, R. DiResta, S. Grossman and E. Cryst, Contours and controversies of parler, technical report, Stanford Internet Observatory (2021).
[46] E. Khabiri, J. Caverlee and K. Y. Kamath, Predicting semantic annotations on the real-time web, in *Proc. 23rd ACM Conf. Hypertext and Social Media*, 2012, pp. 219–228.
[47] Y. Qian, Y. Li, M. Zhang, G. Ma and F. Lu, Quantifying edge significance on maintaining global connectivity, *Sci. Rep.* **7** (2017) 45380.