

A Graph-based Approach to Explore Relationship between Hashtags and Images

Zhiqiang Zhong¹, Yang Zhang², and Jun Pang^{1,3}

¹ Faculty of Science, Technology and Communication,
University of Luxembourg, Esch-sur-Alzette, Luxembourg

² CISA Helmholtz Center for Information Security,
Saarland Informatics Campus, Saarbrücken, Germany

³ Interdisciplinary Centre for Security, Reliability and Trust,
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Abstract. Online social networks are playing a great role in our daily life by providing a platform for users to present themselves, articulate their social circles, and interact with each other. Posting image is one of the most popular online activities, through which people could share experiences and express their emotions. Intuitively, there must exist a connection between images and their associated hashtags. In this paper, we focus on systematically describing this relationship and using it to improve downstream tasks. First, we use a two-sample *Kolmogorov-Smirnov* test on an Instagram dataset to show the existence of the relationship at a significance level of $\alpha = 0.001$. Second, in order to comprehensively explore the relationship and quantitatively analyse it, we adopt a graph-based approach, utilising the semantic information of hashtags and graph structure among images, to mine meaningful features for both hashtags and images. At last, we apply the extracted features about the relationship to improve an image multi-label classification task. Compared to a state-of-the-art method, we achieve a 12.0% overall precision gain.

1 Introduction

The last decade has witnessed the rapid development of online social networks (OSNs). To certain extent, OSNs have mirrored our society: people perform various activities in OSNs as they do in the offline world, such as establishing social relations, interacting with their friends, sharing life moments, and expressing opinions about various topics.

Image is one of the most popular information being shared in OSNs. For instance, 300 million photos are uploaded to Facebook on a daily base.⁴ Moreover, there exist several popular OSNs dedicated to image sharing, including Instagram and Flickr. Images themselves are a rich source of information. Previously, researchers have studied images in OSNs from various perspectives [6, 23, 26]. These works mainly concentrate on the contents of the images, thus adopting computer vision techniques as the main instrument. Different from images hosted

⁴ <https://zephoria.com/top-15-valuable-facebook-statistics/>

		
Hashtags	#wintersport #skiing #piste #travelgram #snowboarding #mountain #blue_sky #travel #snow #tyrol #winter #outdoors #austria #snowboard #scenery #lechtal	#skiing #utah #snow #equipment #feedtheyouth #findyourgreatest
Labels	helicopter, piste, mode of transport, mountain, snow, geological phenomenon, winter, cable car, mountainous landforms, mountain range	fir, snow, winter, geological phenomenon, mountain range, winter sport, ski equipment, ski, fun, tree

Table 1. Two example images from Instagram. Hashtags are generated by users, and labels are given by Google’s Cloud Vision API.

on other platforms, images in OSNs are often affiliated with other types of user-shared information, such as image captions and hashtags. Such information can contribute to understanding OSN images as well. However, the relationship between images and user-shared information has been left mostly unexplored. We aim to fill this gap by analysing the relationship between images and hashtags.

A hashtag is a single word or short phrase prefixed by the “#” symbol [6]; it is initially created to serve as a metadata tag for people to efficiently search for information in OSNs. Interestingly, hashtags themselves have evolved to convey far richer information than expected and provide an incredibly varied and nuanced method for describing images. Some hashtags describe precise objects in the images, e.g., #glass, #window, #building, and #sky; some are related to the feelings and intent of the users, such as #lovelyday, #whyme, and #celebrating; others refer to some event or geographic position, e.g., #paris, #rio, and #newyork [25]. Besides, users also create many hashtags to convey meanings which previously did not exist in natural languages, e.g., #tbt (an abbreviation for “Throw Back Thursday”), or hashtags without specific meanings, e.g., #ig-photo. Therefore, how to accurately describe and understand the relationship between hashtags and image contents is a significant issue.

Contributions. In this paper, we perform an empirical study on the relationship between hashtags and image contents. Our experiments are conducted on a real-world dataset collected from Instagram. It is worth noting that as it is time-consuming to tag the image contents for all the images in our dataset manually (148,106 images), we use the image *labels* obtained from an automatic image detection tool, i.e., Google’s Cloud Vision API, to represent the image contents.

Relationship verification & quantification. We first verify the relationship between hashtags and image contents (represented by their labels) using the two-

sample *Kolmogorov-Smirnov* (KS) test. Experiments demonstrate that hashtags are indeed related to image contents with a significance level $\alpha = 0.001$.

Furthermore, we model the relationship between hashtags and images (i.e., their labels) as bi-directional prediction tasks, i.e., using an image’s associated hashtags to predict the image’s labels (H2L) and using an image’s labels to predict its hashtags (L2H). The prediction performance is then used to describe the strength of the relationship between images and hashtags. For the H2L task, a straightforward approach is to use word embedding methods [10] to transform hashtags into continuous vectors, representing hashtag semantics, which are later used as features to train a machine learning classifier. A similar approach can be applied to the L2H task as well, namely, to use the obtained label vectors from word embedding methods to predict image’s hashtags. However, this approach only considers the semantic meaning of hashtags (and labels), while neglecting connections among the images. As demonstrated by the example in Table 1, if two images share a few hashtags (i.e., #skiing, #snow), then their contents may have certain similarity as well (i.e., both are about outdoor winter sports in the mountain). To this end, we propose a graph-based approach, which can explore both semantic information of hashtags (and labels) and the graph structure among the images, to measure the relationship between hashtags and images.

Through extensive experiments, we show that our approach has better prediction performance – 34.85% overall precision (O-P) for the H2L task and 23.88% O-P for the L2H task on our Instagram dataset. Compared with the approach based on word embedding, it achieves 70.1% and 17.9% O-P gain for the two tasks, respectively.

Application. After verifying and quantifying the relationship between hashtags and images, we further explore this relationship to improve one downstream task – image multi-label classification. Experiments on the *NUS-WIDE* dataset [4] show that we can achieve a 12.0% O-P gain over a state-of-the-art method. This result further shows that there is indeed a significant relationship between hashtags and image contents.

Overall our current paper makes the following two contributions:

1. We statistically demonstrate and quantify the relationship between hashtags and images. In particular, we propose a new graph-based approach which can extract comprehensive information from both hashtags and images.
2. We further apply the above-identified relationship with our new approach to improve the performance of image multi-label classification.

2 Image-Hashtag Relationship Verification

Instagram is one of the most popular OSNs and a major platform for hashtag- and image-sharing. Therefore, we resort to Instagram to collect our dataset relying on its public API.⁵ Our data collection follows a similar strategy as the

⁵ The dataset was collected in January 2016 when Instagram’s API was still publicly available.

nyc	21553	manner	3341
new_york	7918	nofilter	3265
love	6695	summer	3107
brooklyn	4454	food	2889
instagood	4224	photooftheday	2857
travel	3754	foodporn	2717
newyorkcity	3617	latergram	2689
manhattan	3610	sunset	2427
tbt	3608	picooftheday	2369
art	3505	friend	2336

Table 2. The set of most frequent hashtags in our Instagram dataset.

one proposed by Zhang et al. [25]. Concretely, we sample users from New York by their geo-tagged posts. Then, for each user, we collect all her/his images. In total, we obtain 10,605,399 images from 25,658 users. Then, we perform some pre-processing filtering out those images with less than 3 hashtags. Table 2 gives the top 20 most frequent hashtags together with their frequencies.

As mentioned before, we represent image contents as labels. Manual labelling can be an option but not scalable. Instead, we adopt Google’s Cloud Vision API⁶ to label images. The Cloud Vision API is supported by pre-trained machine learning models; it describes an image’s content as a list of labels. The detected labels cover various aspects of an image ranging from the contained objects to personal feelings as well, e.g., happiness. It is worth noticing that this API has been already used in social media image analysis before [15]. Table 1 depicts two images labelled by Google’s Cloud Vision API.

In total, we have spent 227\$ on labelling 148,106 images. There are 255,298 different hashtags associated with these images. On average, each image has 6.46 hashtags, and each hashtag can appear in 4.19 images. Figure 1(a) presents the distribution of the number of hashtags associated with each image. We can see that the images with 3 hashtags have the largest count, and most images have less than 10 hashtags.

For all our images, Google’s Cloud Vision API provides 6,327 different labels. On average, each image contains 8.27 labels. Figure 1(b) presents the distribution of the number of labels for each image. Google’s Cloud Vision API gives at most 10 labels for one image, thus the amount of images with 10 labels is much more than the amount of images with other numbers of labels (< 10).

From the example in Table 1, we can confirm that the labels given by Google’s Cloud Vision API can sufficiently describe the image contents. It can find the objects (e.g., cable car, piste, ski equipment) in the images, and detect the subject (e.g., winter sport) and feeling (e.g., fun) of images. Besides, we also find out that some hashtags have a close relationship with image contents, e.g., #snowboard, #piste, #skiing, and some of them describe additional information, e.g., #utah, #austria, #travelgram. However, there are also some other hashtags which do not have too much relation with the image’s contents, e.g., #findyourgreatest.

⁶ <https://cloud.google.com/vision/>

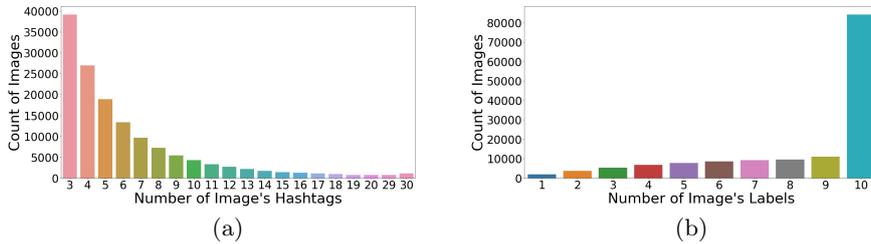


Fig. 1. (a): Distribution of the number of hashtags associated with each image in our Instagram dataset. (b): Distribution of the number of labels for each image in our Instagram dataset.

To verify the existence of the relationship between hashtags and image labels, we perform a two-sample KS test. We construct two vectors hc_c and hc_d with equal number of elements, where each element in hc_c is obtained by calculating the appear ratios of labels in images that have one specific hashtag and similarly each element in h_d is the appear ratio score of labels in images that don't have this hashtag. We perform a two-sample KS test on vectors hc_c and hc_d . The null hypothesis here is that the appear ratio of labels in images with one specific hashtag does not differ from images without this hashtag, i.e., these two vectors are the same, $H_0 : hc_c = hc_d$. Another hypothesis is that the appear ratio of labels in images with one specific hashtag differs from images without this hashtag. Therefore, we have the following two-sample KS test:

$$H_0 : hc_c = hc_d, H_1 : hc_c \neq hc_d$$

The two-sample KS test result suggests a strong evidence with a significance level $\alpha = 0.001$ (p-value = $1e - 91$) to reject the null hypothesis. As a result, we confirm that there exists a relationship between hashtags and image contents.

3 Quantifying Image-Hashtag Relationship

In the previous section, we have demonstrated the existence of the relationship between hashtags and image contents (through examples and a statistical test). In this section, we will systematically quantify this relationship.

Our idea for quantification is to model the relationship between hashtags and images as bi-directional prediction tasks, i.e., using an image's associated hashtags to predict the image's labels (H2L) and using an image's labels to predict its hashtags (L2H). The prediction results can be used to quantify the relationship strength – higher prediction performance indicates a stronger relationship.

In the rest of the section, we first discuss how to use word embedding methods to extract semantic meaning for hashtags and labels for our prediction tasks (Section 3.1). Then, we present a graph embedding based approach in Section 3.2. The experimental results are presented in the end (Section 3.3). For presentation purposes, we use H2L as an example task, similar approaches can be described for the L2H task as well.

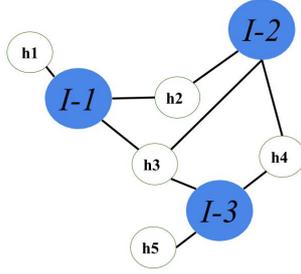


Fig. 2. The example graph of \mathcal{G} . Graph consists of two types of nodes: image (I) and hashtag (h), where each image node connects with the hashtag nodes that appear with the image, and each hashtag node connects with those image nodes with which the hashtag is tagged.

3.1 Word embedding based approach

We use \mathcal{I} to represent the set of images. Each image i is associated with a list of hashtags $H_i = \{h_1, h_2, \dots, h_{m_i}\}$ and a list of labels $L_i = \{\ell_1, \ell_2, \dots, \ell_{n_i}\}$. We use m_i and n_i to denote the number of hashtags and labels in an image i , respectively. Furthermore, we use \mathcal{H} to represent the set of all the hashtags and \mathcal{L} to represent the set of all the labels.

For our H2L task, one intuitive approach is to use hashtags’ semantic meaning as the features to train a machine learning classifier to predict image labels. We apply word embedding to transform each hashtag into a continuous vector, and average the vectors of all hashtags of an image as its feature. To train hashtag embedding, we adopt the Word2vec model [10], meaning that we treat each image’s associated hashtags as a “phrase”, and all these phrases form a “corpus”. The learning process follows the same objective function as Skip-Gram, by applying stochastic gradient descent.

3.2 Graph embedding based approach

The above word embedding based approach only considers the semantic meaning of hashtags (and labels) while neglecting connections among the images. In the example depicted in Table 1, if two images share some hashtags, then their contents share certain similarities as well. We hypothesise that connections among images also possess a strong signal for our prediction task, thus we aim for a method to summarise this relationship as new features.

Our idea of feature extraction is to organise images in a graph according to the connections among them and extract images’ connection information represented in the graph. The graph we construct is $\mathcal{G} = (\mathcal{H}, \mathcal{I}, \mathcal{E}_{HI})$. \mathcal{G} contains two types of nodes: hashtag (\mathcal{H}) and image (\mathcal{I}), each image node connects with its hashtags and each hashtag node connects with images that it appears with (edges in \mathcal{E}_{HI}). The graph in Figure 2 depicts an example of \mathcal{G} .

The state-of-the-art method to extract information from a graph is graph embedding, which aims to learn a mapping that embeds nodes as points in a

low-dimensional vector space [8]. Through optimising this mapping, geometric relation in this learned space reflects the attributes of the original graph.

The graph embedding method we adopt is DeepWalk [13], it is inspired by the idea of word embedding. We treat a graph as a “document” and sample sequence of nodes by random walk on the graph as a “phrase”. Then, word embedding methods can be applied to these phrases as a traditional document task to return us the feature vectors of image nodes. The main reason for adopting this method is that it is relatively efficient and suitable for a large dataset, and its idea has been successfully used in other hashtag-related work [2, 24].

3.3 Experiments

We evaluate the two approaches proposed in Sections 3.1 and 3.2 on the bi-directional prediction tasks (H2L and L2H) on our Instagram dataset to quantify the relationship between hashtags and images.

Evaluation metrics. We adopt those overall evaluation metrics that are widely used in multi-label image classification fields [20], including overall precision (O-P), overall recall (O-R) and overall F1 score (O-F1).

The precision is the number of correctly predicted labels (or hashtags) divided by the number of predicted labels (or hashtags); the recall is the number of correctly predicted labels (or hashtags) divided by the number of ground-truth labels (or hashtags); the F1 score is the geometrical average of the precision and recall scores. Overall means the average is taken over all testing examples. Moreover, we only consider the top 3 predictions for both tasks in our evaluation.

Preprocessing. We adopt the following steps to prepare our dataset. We first convert hashtags into lowercase and delete punctuation. Second, as multiple hashtags may refer to the same underlying concept, we apply a simple process that utilises WordNet [11] synsets to merge some hashtags into a single canonical form, such as “coffeehouse” and “coffeeshop” to “cafe”. Third, for the H2L task: we select the most frequent 100 labels from the dataset and keep images with these labels. For the L2H task: we similarly select 100 most frequent hashtags from the dataset and keep images with these hashtags. To study the influence of the number of hashtags (or labels) of images for these two tasks respectively, we set the minimum number of hashtags (or labels) of the image as a hyperparameter n , then we can filter images with different n . After the preprocessing, we randomly select 20,000 images with different settings.

Implementation details. For fairness, the default embedding dimension d in this paper is set to 256. For the approach based on word embedding, we adopt the Skip-Gram implementation provided by *gensim* [14], and keep the default parameters provided by the software. For the approach based on graph embedding, i.e., DeepWalk, we set the length of each walk to 80 and the number of walks per node to 10.

In the end, we need to feed these extracted features into a logistic classifier to make predictions. In this way, we evaluate the following two approaches to our prediction tasks: Word2vec+logistic and DeepWalk+logistic.

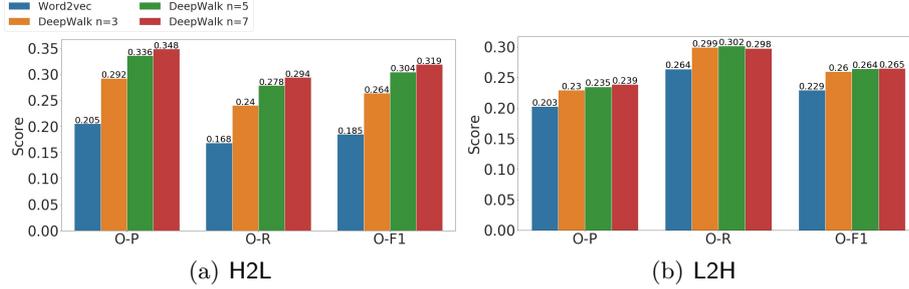


Fig. 3. (a): Experimental results of the task H2L with Word2vec embedding, and DeepWalk embeddings with the different minimum number of hashtags per image ($n = 3, 5, 7$). (b): Experimental results of the task L2H with Word2vec embedding, and DeepWalk embeddings with the different minimum number of labels per image ($n = 3, 5, 7$).

Results. The results for the task H2L are listed in Figure 3(a). We can see that all the O-P scores are no less than 20% for all four settings. Moreover, the results of different DeepWalk embeddings are better than the results of Word2vec embedding (for example, a 70.1% O-P gain and a 72.4% O-F1 gain for DeepWalk with $n = 7$). This indicates that DeepWalk could explore a more comprehensive relationship between hashtags and image contents than only considering the hashtag semantics. Moreover, the results of DeepWalk get better when increasing n . When compare the results of DeepWalk embedding with $n = 7$ and the DeepWalk embedding with $n = 3$, there is a 19.3% O-P gain and a 20.9% O-F1 gain. This indicates an image’s content has a more significant relationship with hashtags, when the image are tagged with more hashtags.

The results of the task L2H are listed in Figure 3(b). We can find that all the O-P scores are more than 20% and the O-F1 scores are more than 22% for these four settings. Similarly, DeepWalk embeddings achieve an improvement when compared with Word2vec embedding. For the O-P scores, the performance gain of DeepWalk embedding with $n = 7$ is 17.9% compared with the Word2vec embedding. But the improvement of DeepWalk embeddings in the L2H task is less significant than in the H2L task. This indicates that for the L2H task, the information provided by the graph relationship among images has a similar strength as only exploring the label’s semantic meaning. Besides, comparing the results of DeepWalk embedding with $n = 7$ with the DeepWalk embedding with $n = 3$, there is a 4.0% O-P gain and a 2.0% O-F1 gain. It indicates that more knowledge of image contents could not significantly help us to predict hashtags for the image.

4 Application

After verifying and quantifying the relationship between hashtags and images, we focus on whether this relationship can be used to improve a downstream task. In particular, we aim to use hashtags’ information summarised by the approach

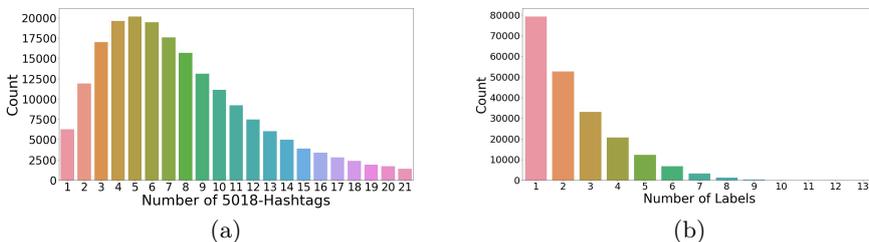


Fig. 4. (a): Distribution of the number of 5018-Hashtags associated with each image in the *NUS-WIDE* dataset. (b): Distribution of the number of labels for each image in the *NUS-WIDE* dataset.

based on DeepWalk to improve the performance of a baseline model on the multi-label classification task.

In order to make sure the reliability of images’ labels and to prove the universality of our method, we use the *NUS-WIDE* dataset, which contains human-generated labels and hashtags shared by real users, for this task. *NUS-WIDE* is a web image dataset [4], and it contains 269,648 images from Flickr. It has two types of hashtags: (i) 5018 unique hashtags (5018-hashtags); (ii) 1000 cleaner hashtags without noisy and rarely-appearing hashtags. Figure 4(a) presents the distribution of the number of 5018-hashtags associated with each image. We could see that the most frequent numbers of hashtags with images are 4, 5, and 6, and this dataset has quite some images with less than three hashtags.

The images in the dataset are also manually annotated using 81 labels by human annotators, which cover different aspects including object classes, scenes, and attributes. The labels on each image are considered as ground truth to represent the image contents.⁷ On average, each image contains 1.87 such labels. The Figure 4(b) presents the distribution of the numbers of labels. We can find that images with only one label have the largest count, and there are only a few images with more than 8 labels.

Preprocessing. To demonstrate the application of the relationship between hashtags and images to improve the performance of image multi-label classification, we use a pre-trained convolution neural network (CNN) as the baseline approach to extract the image features (or image embedding). This technique has been successfully used for many image-related tasks, i.e., image classification [1, 20], image recognition [16], etc. Then, we use the returned image embeddings to train a classifier to make predictions. Second, we use the 5018-hashtags, in this way we keep all the information provided by users. Third, while building the graph structure, we use the same settings as in Section 3.3, and we use 81 labels and set $n = 1$, i.e., we keep all available images from the *NUS-WIDE* dataset.

Implementation details. For the baseline CNN, we use 16 layers VGG network [18] pre-trained on ImageNet 2012 classification challenge dataset [5] using

⁷ This explains why we cannot directly use our Instagram dataset as we don’t have such ground truth.

Methods	O-P (%)	O-R (%)	O-F1 (%)
CNN	48.3	59.5	53.8
Word2vec	40.3	49.4	44.3
DeepWalk	51.3	62.9	56.5
CNN+DeepWalk	55.9	68.5	61.6
CNN+RNN	49.9	61.7	55.2

Table 3. Comparison of the experimental results of the top 3 image multi-label classification on the *NUS-WIDE* dataset with 5018-hashtags.

Pytorch deep learning framework. For our DeepWalk-based approach, we use the graph structure \mathcal{G} , the same as discussed in Section 3.3. The dimensions of the CNN embeddings, Word embeddings and DeepWalk embeddings are set as the same (256). To put different embeddings together, we simply concatenate them. In the end, we feed these extracted features into a logistic regression classifier to make predictions.

Results. We use the same evaluation metrics as discussed in Section 3.3. Table 3 presents the classification results of approaches using the CNN embeddings, the Word2vec embeddings, the DeepWalk embeddings and the CNN+DeepWalk embeddings, respectively. From the results in Table 3, we could first find that the CNN+DeepWalk embeddings can improve the classification performance when only using the CNN embeddings for multi-label classification (with 17.0% O-P gain and 16.2% O-F1 gain). Second, the performance of the DeepWalk embeddings is still better than using the CNN embeddings (with 6.2% O-P gain and 5.0% O-F1 gain). This indicates the relationship between hashtags and image contents is significantly useful for image multilabel classification.

Moreover, we list the results of one state-of-the-art approach CNN+RNN [20], which combines image features and the corresponding hashtags for image multilabel classification. We can find that the results of DeepWalk and CNN+DeepWalk embeddings are better than the CNN+RNN embeddings (a 2.8% O-P gain and a 2.4% O-F1 gain for DeepWalk, and a 12.0% O-P gain and a 11.6% O-F1 gain for CNN+DeepWalk). It further indicates that our approach for relationship exploration is more comprehensive.

Observations. In this section, we present detailed examples to understand the different predictions given by the CNN embeddings and the CNN+DeepWalk embeddings.

In Table 4, there are two images with their associated labels and hashtags from the *NUS-WIDE* dataset, as well as the predictions made by the two approaches based on the CNN embeddings and the CNN+DeepWalk embeddings, respectively. For the image on the left, the CNN embeddings give one correct prediction (“person”) and two incorrect predictions (“sky”, “water”). We notice that this image is somehow unclear and over light. Since the CNN embeddings come from the image itself, it somehow mistakes this strong light in the background as “sky” or “water”. Besides, the correctly predicted label “person” is one of the most popular labels in the dataset (24.6% images contain this label), so this label could not provide precise information to identify the image contents.

		
5018-Hashtags	#film, #army #war, #historic	#fish, #photography #underwater
Labels (ground truth)	military, person	animal, coral, fish
Prediction (CNN)	person, sky, water	animal, coral, water
Prediction (CNN+DeepWalk)	military, person, sky	animal, coral, fish

Table 4. Two example predictions by the CNN approach and the CNN+DeepWalk approach on the *NUS-WIDE* dataset.

On the other hand, the CNN+DeepWalk embeddings correctly predict the two labels “military” and “person”. This indicates that this approach can capture more comprehensive information about this image itself.

For the image on the right, the CNN embeddings give two correct labels (“animal”, “coral”) and one incorrect label (“water”). However, this incorrect label is different from those two incorrect labels for the left image, as it is still relevant to the contents of the image. We could recognise that the image presents an underwater environment, so “water” is not wrong even it does not appear as one of the truth labels. The fish in the right image disguises itself in the environment. In this case, the visual CNN embeddings are not sufficient in capturing small objects (i.e., “fish”) in the image. On the other hand, CNN+DeepWalk embeddings succeed in predicting all three labels.

From these two example predictions on the *NUS-WIDE* dataset, we can confirm that our hashtag features through the DeepWalk embeddings can provide useful information to improve the image multi-label classification even when the image quality is not good enough, or the objectives in the image are not easy to be found by visual features.

We are also interested in knowing whether the DeepWalk embeddings could embed images into the correct position in the embedding space. More specifically, whether they can keep images with similar contents to be close in the space. For this aim, we select 9 groups of labels and each group to have 2 different labels and collect sample images only containing one of the groups of labels. We then transform these image embeddings obtained with DeepWalk into a 2D space using the dimensionality reduction algorithm t-SNE [9].

We visualise the result in Figure 5, and observe the existence of clustering structure in images’ embeddings. These images with different groups of labels are separated into different clusters, and related clusters are close in the space. For instance, in the figure, we can first find that images with the labels related to the animal (“cat”, “birds”, “fish”) are in the left side while the images with labels related to plants (“flowers”, “plants”) are in the right side and the images about the natural scene (“water”, “ocean”, “mountain”) are in the middle. Second, we can also find that images labelled by [“lake”, “mountain”] and images labelled

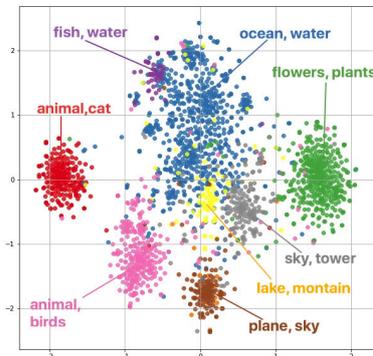


Fig. 5. Visualisation of our DeepWalk embeddings. The images information are mapped to the 2D space using the t-SNE package with learned DeepWalk image embeddings as input. We select some labels: [“animal”, “cat”], [“ocean”, “water”], [“flowers”, “plants”], [“fish”, “water”], [“airport”, “clouds”], [“lake”, “mountain”], [“plane”, “sky”], [“animal”, “birds”] and [“sky”, “tower”] and collect images have these labels.

by [“fish”, “water”] are mixed with the images labelled by [“ocean”, “water”]. This is due to the contents of the two images have similar semantics.

5 Related Work

There has been a diverse array of academic works on exploring the information contained in hashtags. Tsur et al. try to explore what information are contained in hashtags based on a massive dataset from Twitter [19], and they view hashtags as ideas that could express users. As a result, they present the richness of information in hashtags. Furthermore, some work use hashtags to detect the topic of tweets on Twitter [21] and predict hashtags based on tweet contents [7, 17]. These work indicate there is a strong relationship between hashtags and text contents, and it is possible to make two-way predictions between them.

Focusing on the relationship between hashtags and images, Niu et al. propose a semi-supervised Relational Topic Model (ss-RTM) to use hashtags information to recognise social media images [12]. They first organise images into a network if they share some hashtags. Then, they treat this network as a document and use a statistical model RTM, which is widely used in natural language processing tasks to extract the topic relationship among documents, to extract images’ relationships into representative vectors. Compared with our work, they only use hashtags’ information to build up the network but ignore their semantic meaning in the final features. Besides, due to the computational cost of RTM, they cannot involve a large number of images in one network, and there might be a strong influence from noisy hashtags. Wang et al. propose a framework (CNN-RNN) which combines hashtags and image features to perform classification [20]. CNN-RNN mainly contains two parts – a CNN model for extracting

semantic representations from the images, and an RNN (recurrent neural network) to model image/labels relationship and hashtags dependency. Due to the advantages of RNN, this framework can utilise the order information among hashtags, and it can predict a long sequence of labels. It achieves better performance compared with ss-RTM, but it neglects the connections among images. Recently, Wang et al. utilise a hashtag-related knowledge graph to improve image multi-label classification [22]. They first build a large knowledge graph, which contains millions of hashtags and their semantic relationships. Then they apply the deep graph embedding methods to extract hashtags' relationship to representative vectors and use the representative vectors to assist the classification task. But it is a high-cost work to build up a knowledge graph with millions of hashtags, and they only consider the hashtags semantic information but neglect the graph structure among the images.

6 Conclusion and Future Work

In this paper, we have performed an empirical study on verifying and quantifying the relationship between hashtags and images based on real-world datasets collected from Instagram and Flickr, and we successfully applied the verified relationship to improve a downstream task.

We have implemented a statistical test to verify the existence of the relationship between hashtags and images on the Instagram dataset. Then, we designed bi-directional prediction tasks (H2L and L2H) and used the prediction performance to quantify the relationship. In particular, we proposed a new graph-based approach to integrate both the semantic meaning of hashtags (and labels) and the graph structure of the images, which indeed help to extract more comprehensive information for hashtags (and labels). In the end, we adopted a widely used dataset *NUS-WIDE* which has tags given by users and manual labels, and successfully applied the extracted features of hashtags from the H2L task to improve the performance of image multi-label classification and achieved a 12.0% overall precision gain compared to a state-of-the-art method.

Hashtags can be naturally organised into different categories, according to their semantics. In the future, we will first focus on the influence of hashtag categories, i.e., investigating the different relationship strength between each category of hashtags and images. Second, on OSNs different users have different habits of using hashtags, and we hypothesise that the richness of the semantic meaning contained in their hashtags could be different. How to explore this, e.g., to perform link prediction as in [3, 2, 24], is part of our future work. Third, so far we have only applied the extracted hashtag features for an image multi-label classification task in this work. In the future, we want to utilise the extracted label features (from the L2H task) to perform hashtag recommendation in OSNs.

Acknowledgements. This work is partially supported by the Luxembourg National Research Fund through grant PRIDE15/10621687/SPsquared.

References

1. Akata, Z., Reed, S., Walter, D., Lee, H., Schiele, B.: Evaluation of output embeddings for fine-grained image classification. In: Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2927–2936. IEEE (2015)
2. Backes, M., Humbert, M., Pang, J., Zhang, Y.: walk2friends: Inferring social links from mobility profiles. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 1943–1957. ACM (2017)
3. Cheng, R., Pang, J., Zhang, Y.: Inferring friendship from check-in data of location-based social networks. In: Proceedings of the 2015 Workshop on Social Network Analysis in Applications (SNAA). pp. 1284–1291. ACM (2015)
4. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: NUS-WIDE: a real-world web image database from National University of Singapore. In: Proceedings of the 2009 International Conference on Image and Video Retrieval (CIVR). ACM (2009)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the 2009 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 248–255. IEEE (2009)
6. Denton, E., Weston, J., Paluri, M., Bourdev, L.D., Fergus, R.: User conditional hashtag prediction for images. In: Proceedings of the 2015 ACM Conference on Knowledge Discovery and Data Mining (KDD). pp. 1731–1740. ACM (2015)
7. Godin, F., Slavković, V., Neve, W.D., Schrauwen, B., de Walle, R.V.: Using topic models for Twitter hashtag recommendation. In: Proceedings of the 2013 International Conference on World Wide Web (WWW). pp. 593–596. ACM (2013)
8. Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. *IEEE Data Engineering Bulletin* **40**, 52–74 (2017)
9. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. In: Eurographics Conference on Visualization (EuroVis). pp. 2579–2605. Eurographics Association (2008)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: Proceedings of the 2013 International Conference on Learning Representations (ICLR) (2013)
11. Miller, G.A.: WordNet: A lexical database for english. *Communications of the ACM* **38**(11), 39–41 (1995)
12. Niu, Z., Hua, G., Gao, X., Tian, Q.: Semi-supervised relational topic model for weakly annotated image recognition in social media. In: Proceedings of the 2014 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4233–4240. IEEE (2014)
13. Perozzi, B., Al-Rfou, R., Skiena, S.: DeepWalk: Online learning of social representations. In: Proceedings of the 2014 ACM Conference on Knowledge Discovery and Data Mining (KDD). pp. 701–710. ACM (2014)
14. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. pp. 45–50. ELRA (2010)
15. Richards, D.R., Tunçer, B.: Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services* **31**, 318–325 (2018)
16. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 815–823. IEEE (2015)

17. She, J., Chen, L.: Tomoha: Topic model-based hashtag recommendation on twitter. In: Proceedings of the 2014 International Conference on World Wide Web (WWW). pp. 371–372. ACM (2014)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 2018 International Conference on Learning Representations (ICLR) (2015)
19. Tsur, O., Rappoport, A.: What’s in a hashtag? Content based prediction of the spread of ideas in microblogging communities. In: Proceedings of the 2012 ACM International Conference on Web Search and Data Mining (WSDM). pp. 643–652. ACM (2012)
20. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., Xu, W.: CNN-RNN: A unified framework for multi-label image classification. In: Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2285–2294. IEEE (2016)
21. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In: Proceedings of the 2011 ACM International Conference on Information and Knowledge Management (CIKM). pp. 1031–1040. ACM (2011)
22. Wang, X., Ye, Y., Gupta, A.: Zero-shot recognition via semantic embeddings and knowledge graphs. In: Proceedings of the 2018 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6857–6866. IEEE (2018)
23. Wu, J., Yu, Y., Huang, C., Yu, K.: Deep multiple instance learning for image classification and auto-annotation. In: Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3460–3469. IEEE (2015)
24. Zhang, Y.: Language in our time: An empirical analysis of hashtags. In: Proceedings of the 2019 International Conference on World Wide Web (WWW). pp. 2378–2389. ACM (2019)
25. Zhang, Y., Humbert, M., Rahman, T., Li, C.T., Pang, J., Backes, M.: Tagvisor: A privacy advisor for sharing hashtags. In: Proceedings of the 2018 Web Conference (WWW). pp. 287–296. ACM (2018)
26. Zhou, B., Lapedriza, À., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**(6), 1452–1464 (2018)