

Chirag Shahani (SUNetID: 06237446)
Siddhesh Dindorkar (SUNetID: 06246946)

Flickr: Predicting Top Photo Categories For A Location

Abstract

The popularity of image sharing sites brings an interesting problem of analyzing and co-relating the images by leveraging the relationships between metadata information from these images. By grouping and analyzing image annotations we tried to categorize images into predefined set of categories and ranking the categories for a given location. To model a graph for a set of images based on a given location we considered the images' categories and the authors' connections with each other. We found that using authors' following information gives us a rich subset of images and users that are popular in a region and define the popularity for that region. This data of popular categories, images and users would help a new photographer just starting out in his quest to get some influence of what is a popular category in that region and get a headstart in defining his style by studying popular images and the users.

1. Introduction

Flickr is among one of the widely used image sharing website, contains rich metadata information, is a social network of photographers who like or comment on other images, submit their image to

groups/communities and hence an ideal contender for not only studying images and their associated tags but also modelling the social network formed by user interaction and interdependencies between common metadata annotations. We use Flickr to analyze how image sharing differ based on influential users/popular images in a particular region and study how a particular genre of photography or a group of users influence that region to create/share similar images.

Apart from doing content analysis of an images, using the image metadata that users create when uploading the image on flickr and further based on the interactions they have with other users/images by liking, commenting on images, following other fellow photographers, submitting the images to flickr groups and forums, proves much beneficial to characterise the image and judge the image popularity in relative terms based on other images with similar properties. For example, it is fair to compare images from same location/city and images belonging to same categories. This social analysis using the image annotations helps predict most popular image categories and within those categories the set of influential images and influential users. This leads to much better results than just merely relying on image view count and likes. Our goal was to find out, given a location, determine a list of popular image categories that the region is famous for or good at and provide a list of images and their users who form the influential set of this category which other users/new photographers can look up to.

This paper is divided into following sections: *Section II. describes the Related Research our paper is based on, Section III describes*

our Data Collection process, Section IV describes our Graph Modelling and application of Newman and Girvan's divisive method, Section V. describes our algorithm's Evaluation and lessons learnt and section VII describes the Future Scope for this project.

II. Related Research

We studied and got a head start by using the graph generation and evaluation concepts from *J. McAuley and J. Leskovec. Image Labeling on a Network: Using Social-Network Metadata for Image Classification [1]* paper. This paper talks in depth about how to not treat image labelling a problem of just individual image classification by content analysis but also suggests to use the image metadata relational properties by exploring the interdependencies between them. The paper talks about modelling the relational data found in images in the flickr dataset and how to use this model for different image labelling tasks. From the experiments, the authors prove that social network analysis of the the image metadata based on image submitted to common groups, common collections, taken by the same photographer leads to a substantially rich model for image labelling as compared to content based image classification techniques.

The authors found out that labelling an image is just not a function of neighboring nodes but it is far more worthwhile to consider the image metadata and the relative relationship between the metadata objects. Based on their study, the authors

note an interesting observation that images that share a group or tag are lot more likely to have common labels. Using the relationships between the image annotations for image classification, the authors show that their model outperforms the flat SVM model which analyses each image independently of the other.

For related research we also referenced *Aditya Khosla , Atish Das Sarma , Raffay Hamid, What makes an image popular? [2]*. This paper predicts how popular a photo will be on the social Web. The paper claims to examine 2.3 million images on Flickr and concluded they could predict an image's popularity based on the content and the social context. To predict how popular some image would be based on content, the authors looked at the image contents, the colors used in the images, the texture and gradient, as well as which objects were present. They also looked at how social context impacts popularity, studying follower counts, how many views a photoset received, and how long users had been on the social network.

The authors explore socio aspect contributing to image popularity by using social cues contained in image metadata. The authors conclude that social cues influence the popularity of the image along with image content and say that social cues perform better in case of user-mix dataset and image content performs better in case of user-specific dataset. Based on the intuition of the authors, they observe that social cues become less relevant when users upload many images. Lastly, the authors conclude that if both image content and social cues are considered together to relate to popularity, the results are much

better as compared to when the either one is considered to relate to popularity. We found this paper to not utilize much of the social aspects of an image. They considered a image popularity on more of absolute terms per image rather than taking several related images into consideration. They tried to predict if an image will be popular based on image content and image annotation, however, we found this image popularity data of individual images hard to correlate amongst each.

We also referenced *Lindstaedt, S., Pammer, V., Mörzinger, R., Kern, R., Mülner, H., Wagner, C. Recommending tags for pictures based on text, visual content and user context [3]*. The authors developed a model named *tagr* to recommend tags for images using their visual, textual content and user info. The goal of the authors was to suggest tags for images, predict similar images to the reference image and suggest similar users as the user to whom the referenced image belonged.

The authors use an interesting technique for tag prediction using *WordNet*, a lexical database that collects terms around senses and uses it to predict more tags / senses by taking source tags as input senses. For each of the sense, it finds the synonym senses and hyper/hyponym senses. The use of WordNet helps the authors to back their tag prediction algorithm with second and third level source of data. This makes tag prediction much more accurate, as the tags not only predict the main focus of the image but also tries to identify the rest of the things in the image. We used wordNet in a similar way to help us categorize images into predefined set of categories.

III. Data Collection

Similar to [1], to form the flickr dataset, we used four image collections already available and then used Flickr's APIs to collect the metadata info for these images. FlickrXml contained image metadata info obtained from four data sources: CLEF, MIR, PASCAL and NUS. As noted in [1], we found the metadata information very limiting in PASCAL dataset.

We divided the data collection process into three parts:

- *Location shortlisting:*
Based on the image metadata available in FlickrXml we extracted location information from the owner tag. Images with blank location were ignored. We could generalize the location information to a set of repetitive patterns, usually in the form of <city, state, country> or <city, country> or <country>. For simplicity and to avoid drastic variations in our location comparison, we considered images belonging to same country rather than different cities, as not all images belonging to a city were found to be tagged correctly and each city was found to have very significant difference in the number of images as well. Once the algorithm matures, it would be easy to find and apply the logic on images from a particular city and try to be more granular location wise.

- *Category generation:*

We defined a set of predefined categories that an image can belong to. The image can belong to more than one category. But for sake of implementation, we just used the top most category with most ranks as the category the image belonged to.

We considered words from <title>, <description>, <tags>, <labels> and <groups the image was submitted to>. We created a learning category dictionary which tries to match the words with predefined categories. Each word in the dictionary contains a list of categories it can belong to. To start with we input the categories themselves and special signifying words per category into the dictionary. Per image we keep another image dictionary with an entry per category denoting the number of hits for that category for this image. As we scan through images and encounter new words, we use wordNet to find synonyms of the words. At each step of evaluating words and their synonyms, we check if the word is present in the category dictionary. If yes, we increment the categories of the word for the image and proceed to pick the next word from the image annotations. If we don't find a word in category dictionary, we search for the word's synonyms using wordNet and evaluate the if the synonym is present in category dictionary. If yes, add this synonym to the category dictionary. If not find next set of synonyms for this synonym and continue the lookup procedure. We do this recursive lookup for 3 levels of depth with the hope that the

synonyms will eventually be found in the category dictionary. Since we take labels into consideration, the prediction of results yeild better results if we compare the predicted category against the labels themselves.

- *User follower information:*
Based on the images in the flickXml dataset, we extracted the owner id of the image and using flick API got the list of users followed by the user of the image. For simplicity of calculations and to have an undirected graph, we assumed that if user A follows user B, user B follows A as well.

IV. Graph Modelling

In our graph, based on the country the user inputs, we select all the images belonging to that country.

- Nodes are the images themselves
- Edges are formed if
 - two images share the same category and the users follow each other
 - or the two images belong to the same user

Adding the condition of forming an edge if users follow each other may cause fragmentation of the graph, since even if the image belongs to same category, it may not be connected to other images with same category, especially if the user is new or if that user does not follow or is followed by anyone. However, we find this to be a transient state and overtime, based on clustering coefficient principles, as users

become friends and they follow each other they make friends of friends. To further aid this to happen, based on the generated graph, we could query the graph to find the most popular set of images and their users in the required category as well as top category in her/his region, recommending her/him to follow the suggested users. Thus making the graph better connected.

The main goal of our paper was to be able to ask:

- What are the most popular categories in a region?
- Given a category,
 - which images are popular?
 - which users are popular?

Given a region, finding popular category will help a newcomer to gauge the region for well received image categories and finding popular images and users in these categories will help the newcomer/struggling photographer to view which images work. Another application of our graph model can be, that given a category as per the newcomer/struggling photographer's liking, finding the influencing images and users within this category will help him take inspiration from those users and those images. This was the social network grows with better content and better connectivity between the nodes as more and more users follow each other and new photographers.

To our modelled graph, we apply the *divisive method algorithm by Newman and Girvan[4]* to divide our graph into distinct categorial components. As a result, we find our graph divided into multiple components representing a single category of the image. We then analyze the individual clusters to determine properties of each cluster, such

as, number of nodes, top images in this SCC based on view counts, likes and favs, top users of those images. Once we do that, we can also calculate a metric per cluster or per category by calculating:

- To get average ratio of views of images per category to determine which category sells in a region:

$$\frac{\text{sum of number of views of all the images}}{\text{num of images}}$$

- To get average ratio of likes to the views to determine likeability of images within an category:

$$\frac{\text{sum of likes of all the images}}{(\text{sum of number of views of all the images} * \text{num of images})}$$

Newman and Girvan prove that intra community links are short and strong rather than inter community links. They show that the intra community links join set of similar and cohesive nodes. Applying similar principles, to achieve the desired clustering of graph into clusters of distinct categories, they find and pick intra community edges, that is, edges with high betweenness quotient. The goal is then to find edges connecting two or more communities. If we find these edges and remove them, what is left are the individual category clusters.

Our goal here was to find the categories in the graph model we created without explicitly specifying the number of communities. The motivation is that the few edges that lie between communities can be thought of as forming bottlenecks between the communities of one kind or another that flows through the network will have to travel along at least one of these bottleneck edges if it wishes to pass from one community to another. Thus if we consider some model of traffic on the network and look for the edges with highest traffic, we should find the edges

between the communities. Removing these should then split the network into its natural communities. The betweenness of an edge is defined to be the number of shortest paths between vertex pairs that run along the edge in question, summed over all vertex pairs. This quantity can be calculated for all edges in time that goes as $O(mn)$ on a graph with m edges and n vertices. This method provided no guide to how many communities a network should be split into.

To overcome this, we started to find modularity. We started with a random starting node in the graph, say S . W_i is the weight of node i which is calculated based on number of distinct paths from S to i . So considering two nodes i and j connected to S , where j is farther from S , [4] say that the fraction of paths from j to S that go via i can be given by W_i / W_j . So to calculate the edge betweenness from all shortest paths source at node S :

- Find all the leaf nodes t
- For each node i connected to t , assign a weight of W_i / W_t to the edge E_{ti} .
- Similarly using BFS, try to move up one step at a time from node i to eventually node S , and assign weight $(1 + W_i / W_j) * W_j / W_k$ Where t is farther than i , i is father than k and k is farther than j .

The above BFS takes $O(m)$ time complexity and doing this for n set of nodes after removing an edge with largest betweenness quotient each time the above steps are executed, gives the overall time complexity of $O(nm)$ where n are the number of nodes and m are the number of edges.

Based on the above results, the edges connecting two category of images had high

betweenness quotient and thus removed before the edges inside a category gets removed. We stopped the computation when the variance of category of the nodes in the BFS is less than a certain threshold, say 1.

As we apply clustering algorithm to find the clusters, to measure the quality of the community structure, we can make use of modularity principles. Modularity will quantify the extent to which we will have more intra community links than the same community structure if the links were rewired using Erdos Reyni graph model.

Modularity was computed as

$$Q = 1/2m \sum (A_{i,j} - K_i k_j / 2m) S(c_i, c_j)$$

$A_{i,j}$ -> adjacency matrix (observed number of intra-community edges)

K_i -> degree of node i

C_i -> community of node i

$S(c_i, c_j) = 1$ if i, j belong to same community

M = number of edges in a graph.

We compare the modularity of a random graph model (Erdos Reyni) and the graph created by the flickr dataset. The modularity of a Erdos Reyni graph model will be 0, it being a randomly generated graph.

V. Evaluation

Since we used wordNet to find synonyms and categorize each word into a list of categories, multiple categories were included for the image, which did not even have any relation to the image. We compared the results of the image categorization with the image labels themselves and found that to be able to

determine correct category for an image, we need heuristics per word. It is not enough to treat all words with equal priority, some common words describe most of the categories or may not be applicable to any categories at all. We needed to eliminate these words, but still need to find a better automated logic to do it for us instead of manually creating a blacklist of words to not use for categorization.

The main disadvantage of the algorithm of Girvan and Newman divisive method, we found is that it is slow. Since there are m edges to be removed in total and each iteration of the algorithm takes $O(mn)$ time, the worst-case running time of the algorithm is $O(m^2n)$, or $O(n^3)$ on a sparse graph. However, it did give us results where images of different categories were deduced. We were able to make the partitions in the graph data. Applying the algorithm to our graph, we found that communities there were multiple overlaps. We saw that amongst many overlapping subsets, a notable one was communities of *Landscape* and *Portrait* had overlaps from community of *Cities & Architecture*.

VII. Future Scope

To be better able to categorize our images, along with the image annotations, it would have been beneficial to take image content analysis algorithms into consideration. Using the content analysis tools we could have outright rejected certain categories which are absolutely not relatable to the image.

We could also assign weights to the edges and nodes so as to better model the real

world scenario of flickr social network. For example, we can increase the weight of an edge between two images, if the two images share more than one common category by a factor of how many common categories do the images share. We could also assign weight to the nodes based on the likeability of the image, that is a factor of (likes, comments and view count). This weight assignment to the nodes would have helped us to

- a. Take image popularity into account, more the views, likes, comments, better is the image. It may happen that in a particular city, pet photography is really common and hence without any node weight pet photography would be counted as the top category. However, if landscape photography is a niche segment that not many photographers do, but those who do are really well received, the nodal weight model will help to accurately predict the better category rankings.
- b. As with any typical social network, more an user is connected more will his images be well received. However, this may not signify that his images are absolute best. A newcomer to the social network may have a some of the best images in the well connected user's category, however, due to not many connections, the newcomer's images are not discovered. With assigning nodal weight as a factor of view count and the corresponding likes, we can fine tune the algorithm to give higher weightage to the newcomer's images.

References

- [1] J. McAuley and J. Leskovec. *Image Labeling on a Network: Using Social-Network Metadata for Image Classification*
- [2] Aditya Khosla , Atish Das Sarma , Raffay Hamid, What makes an image popular?
- [3] Lindstaedt, S., Pammer, V., Mörzinger, R., Kern, R., Mülner, H., Wagner, C. *Recommending tags for pictures based on text, visual content and user context*
- [4] M. E. J. Newman, M. Girvan, *Finding and evaluating community structure in networks*
- [5] S. Fortunato *Community detection in graphs, Arxiv 2009.*
- [6] Vincent D. Blondel, *Fast unfolding of communities in large networks*