Hashtag of Instagram: From Folksonomy to Complex Network

Simona Ibba, Matteo Orrù, Filippo Eros Pani and Simone Porru
Department of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, Cagliari, Italy

Keywords: Instagram, Social Tagging, Knowledge Management, Complex Network, Folksonomy.

Abstract: The Instagram is a social network for smartphones created in 2010 and acquired by Facebook in 2012. It currently has more than 300 million registered users and allows for the immediate upload of images (square, inspired by Polaroid), to which users can associate hashtags and comments. Moreover, connections can be created between users that share the same interests. In our work, we intend to analyze the hashtags entered by users: the use of such hashtags, as it happens in other social networks like Twitter, generates a folksonomy, that is a user-driven classification of information. We intend to map that folksonomy as a complex network to which we can associate all the typical analysis and evaluations of such a mathematical model. Our purpose is to use the resulting complex network as a marketing tool, in order to improve brand or product awareness.

1 INTRODUCTION

Social Media Marketing has become essential for every company: it is almost free, and, if used correctly, it is incomparable in terms of audit and Customer Relationship Management.

The main purpose of social strategy is to increase brand reputation, brand loyalty and brand awareness. The main tools of Social Media Strategy have seen the recent entry in their ranks of Instagram, together with the more widespread Twitter, Facebook or Youtube. Instagram is a mobile application started in 2010, that in 2015 surpassed the mark of 300 million registered users. Instagram has different features from the other tools: it is an exclusively mobile application, it does not allow the posting of direct links to sites, and has a difficult integration with Social Media Management tools. Those characteristics, which could appear as limits, make Instagram a very fitting tool for the creation of brand awareness. The brand Nike, for example, (http://instagram.com/nike) owns an Instagram profile with 17,822,162 followers currently, with an Avg Daily Followers number of 40,752 (http://socialblade.com/instagram/user/nike). These figures would suggest that Instagram will soon integrate advertising systems that would allow brands to pay to reach users that are still not followers. In the meantime, the main tool in the hands of a manager of an Instagram profile is hashtags: a suitable hashtag makes the content visible to all users interested in that specific topic, and also gives visibility to the profile, generating additional followers. Good content with suitable hashtags generate likes and comments that make the account visible in the "Explore" tab, where new content of interest can be found.

With our work, we intend to research the hashtags entered by users, through the use of Instagram's Application Programming Interface (API). We plan to establish a relation between hashtags and metadata associated to uploaded images, and to analyze those relations. Social tagging on Instagram leads to the generation of a folksonomy, that is a collaborative, collective, and social organization, at the metadata level, of information entered by users, as suggested by Angius et al., (2014). Lastly, our aim is to map hashtags through a complex network and find a tool for the creation of new hashtags relevant to the images and that contribute to increase the visibility of pictures.

The general objectives of our research are summarized below:

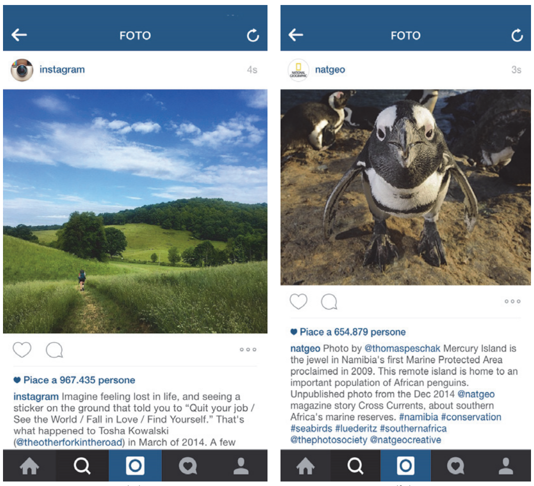
- Analyze the properties of complex networks originating from Instagram hashtags.
- Elaborate predictive models for the content posted by users
- Locate the content posted by users on Instagram to develop appropriate marketing strategies.

Ibba, S., Orrù, M., Pani, F. and Porru, S..
Hashtag of Instagram: From Folksonomy to Complex Network.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 279-284
ISBN: 978-989-758-158-8

This paper is structured as follows: Section 2 describes the context in which the proposed project lies, in Section 3 we present an overview about the state of the art; in Section 4 we describe our approach, while in Section 5 we explain our project for the use of Instagram hashtags as a complex network. The last section hosts our final observations about the project.

2 BACKGROUND

Instagram allows the upload and sharing of photos and videos through the use of a mobile device.



(a) (b)

Figure 1: Interfaces of Instagram. (a) Instagram photo with comments (b) Hashtags entered by users.

It offers its users a unique way to post their pictures, and allows immediate editing through 22 filters. It also allows users to add captions, hashtags using the # symbol to describe photos or videos, or to mention other users using the @ symbol (the @ symbol creates an actual link among the accounts). Instagram also allows users to follow posts from all selected profiles, and to have their own followers. Each follower must be specifically approved by the profile owner. Every user can set their own privacy preferences, and can make the pictures visible to everyone or only to followers. Pictures in profiles are shown in chronological order, starting from the latest. For each picture, it is possible to enter likes or comments. Hashtags and user mentions can be entered inside comments.

Moreover, according to a research by (Hu et al., 2014), Instagram photos can be roughly categorized into eight types based on their content: self-portraits, friends, activities, captioned photos (pictures with embedded text), food, gadgets, fashion, and pets, where the first six types are much more popular.

3 RELATED WORK

In the literature, studies on Instagram are in markedly smaller numbers than studies on Twitter. The latest developments and the big figures in Instagram usage as of late, lead us to believe that this mobile app is deserving of more attention from the scientific community. Understanding Instagram and its mechanics means working on human, cultural, social, and environmental dynamics that open several scenarios, also from a business, communication, and marketing perspective.

The main key to Instagram is the hashtag: hashtags can function as descriptive elements in the image, or can be related to its caption, explaining its content better. Hashtags are thus a way to manage and organize information, as stated by Mathes (Mathes et al., 2014) as regards the social bookmarking system in Delicious, in that it leads to the generation of a folksonomy. Bruns and Burgess (2014) maintain that the use of certain hashtags can allow certain types of communities to emerge and form, including ad hoc publics, forming and responding very quickly in relation to a particular event or topical issue.

The main purpose of social tagging is thus to facilitate visibility of information (visibility of images in Instagram's case) for the creation of recommendation systems. Our work is based on the intention of finding, in an automated way, using complex networks and missing links, the most suitable hashtags to the retrieval of images on Instagram, making an account more visible as a result. Similar works have been carried out by Marlow et al., (2006), and Ames and Naaman (2007), who view content management and retrieval as two of the most important incentives to tag resources. Hashtags in Instagram can describe the content of the image, but can likewise represent subjective opinions, feelings, places, or a variety of expressions pertaining to colloquial language. In our work, we took our cue from Bischoff et al., (2007), who analyzed the data sets of tags extracted from several tagging systems like Flickr and Delicious, and studied the distribution of tags associated to the different resources to find the implications derived from the usage of different tags to improve research and visibility of content. One of the first authors that thought of building a network of tags and of using a graph-based approach is Mika (2005). In their work, they apply clustering techniques to tags, uses their co-occurrence statistics, and produce conceptual hierarchies. A useful approach to define sets of related tags is the one devised by Gemmell et al., (2008): it is based on the hypothesis that clustering tech1niques are better if the context in which the tags appear can be determined. In De Gemmis et al., (2008) suggest an interpretation of tags through the Wordnet thesaurus: related tags can be defined in this way, reducing the issue of ambiguity.

The use of a folksonomy as a Complex Network was introduced also by Shen et al., (2005), who highlighted that, since folksonomy is a classification system of web contents, its properties both static and dynamic can also serve to search and retrieve the related information.

Complex networks have recently received a lot of attention by researchers of different disciplines (Newman, 2003). These systems are found in many area of the natural (metabolic network, etc) and human (Internet, the web, power networks, etc). Althogh their different nature, several different networks share some properties: scale-free network, small-world, (Valverde et al., 2002; 2003), and community structure (Girva et al., 2001) (Fortunato, 2010), just to name a few of the most studied.

Many complex networks of different nature present the small-world property (Milgram, 1967), showing small values for mean shortest paths (if compared to a random network of the same size) and high value of the clustering coefficient. In this kind of networks the paths that separate each pair of nodes are relatively short.

Another interesting properties regards the distribution of the degree (in and out) of complex networks, that usually follows an exponential distribution. This led to the definition of scale-free networks. This kind of network present few nodes having high degree, whereas the vast majority of nodes have a low degree.

One of the most important property of complex network is the community structure (Girvan et al., 2001), (Fortunato, 2010). A community is a subnetwork of densely connected nodes. The nodes of a community are more connected to each other if compared to nodes that are outside the community. Determining a significant community structure is not straight forward and many authors proposed their algorithm to determine the community structure of a network. One of the most used is the FastGreedy algorithm (Clauset et al., 2004).

Complex networks show also a hierarchical structure, namely the network's communities are organized at different levels, being some of the included into others. A method to retrieve the hierarchical structure of a network has been proposed by Clauset et al., (2007; 2008) and consist in using a Monte-Carlo Markov Chain approach to compute the dendrogram of an associated hierarchical random graph. This method can be used to for the detection of missing links between nodes.

Figure 2: Collaboration network of jazz musicians. Each node is a Jazz musician and an edge denotes that two musicians have played together in a band (Gleiser and Danon, 2003). Different colors represent the community retrived with the FastGreedy Algorithm (Clauset et al. 2004).

4 THE PROPOSED APPROACH

The proposed approach is based on the concept of complex network (or graph), that is a structure with two main separate elements: nodes, which represent the basic elements of the graph, and node connections, called branches or arches. An example of a network we face every day is the road network. We could associate cities to nodes, and roads connecting cities to branches. The complexity is intrinsic to the networks as their size grows, that is to the number of involved elements (nodes and branches), to the point that it becomes extremely complex to understand its structure, its behavior, and its evolution. For this reason, in order to study complex networks, tools, methods, and algorithms coming from a multi-field knowledge corpus (statistical physics, sociology, etc.) are used, a corpus that has been gradually built over the years.

What makes complex networks interesting is the fact that they constitute a mathematical model that can represent facets and artifacts of human life, and several natural phenomena. Networks of varied origin and nature (from electric networks to metabolic ones) possess the same properties.

It is possible to apply the same concept of network to human knowledge. In this case, the term knowledge network is used. An example of knowledge network is represented by bibliographic networks created from author collaboration (co-authorship) information or quotations appeared in reference sections.

In our work, we aim thus to map the hashtags and metadata retrieved from the Instagram app as a complex network, and we intend to study some applications that, through the study of the network, could be developed to make Instagram a more rounded marketing tool.

In our study, we want to evaluate the effect of hashtag categorization in the context of folksonomy-based item recommendation using data crawled from the multi-topic social networking system of Instagram.

Shen k. et al. (Shen et al., 2005) demonstrated that the folksonomy at Del.icio.us it can be studied as a complex network formed by tags because displays both nature of small world, scale free and highly clustered.

We want to apply Instagram hashtags metrics typical of complex networks. In particular we want to study the incidence matrix of a subgraph to determine the properties of the network. We can thus identify the most important nodes, the ease of transition from one node to another, the indices of centrality, the hub nodes and those autority and indexes of the graph. This information will allow us to interpret the content posted by users and see which are the most important information for each geographical area and how this information can be linked to content of other subgraphs network.

5 HASHTAG AS NODES OF A COMPLEX NETWORK

Our analysis is based on the Instagram data collected using the Instagram API. Instagram Application Programming Interface (API) allows to research tags placed by users, and provides complete information on images and videos. Through the API, for each photograph it is possible to find an univocal identifier (id) and the link to two versions of the same image (low resolution and standard resolution), the metadata that describe user name, date and time of image creation, the location where the picture was taken, the caption entered by the author, comments, tags associated to the image, number of likes and names of the users that gave their like. Other than the metadata associated to images, metadata associated to each user that has posted a picture can be extracted. These metadata allow to find number of followers and following, email address, number of posts and a brief biography.

The high number of available data allows to perform quantitative and qualitative analysis, to verify the stream of content over time and find the most interesting topic to users. The information gathered from it makes it also possible to map data according to their geographic location, and to analyze the geolocalization of users in relation to specific tags of interest. The analysis on tags allows to find users' consumption models, the type of posted content, and the specific locations where the same content is posted, together with time data.

As suggested in Laniado et al., (2010), who analyzed Twitter's hashtags network, we are going to define a set of metrics for describing hashtags usage from different perspectives: frequency, specificity, consistency and stability over time.

The main purpose of Instagram users, especially of those users that manage communities (http://instagramersitalia.it/iger/igers_sardegna/) is to make their posted photos as visible as possible: with a higher visibility, the number of followers increases, and consequently so does the size of the community and the number of users who can automatically see the pictures posted by a specific profile (and can thus comment or enter a like). The pictures can represent a communication strategy, and can be a part of a marketing plan. For example, they could lead to the download of an app, or to the promotion of a restaurant, etc. In order to make a photo visible, and iteratively increase the number of followers of the profile, it is necessary to use suitable tags. Instagram users seldom find the most suitable tags for a specific photo.

The aim of our work is, therefore, to map Instagram tags through a complex network (or complex graph). As proposed by (Cantador et al., 2011) for the Flickr social network, we intend to use an approach based on Random Walks with Restarts theory (Lovasz, 1996), which allows us to directly predict the preference of users to particular photos from the data collection acquired, by taking into account not only their personal profiles in terms of item preferences but also their tagging behavior, social network as well as similarly tagged items.

Specifically, we want to create our social graph by representing users, photos and hashtags as nodes. User relationships are encoded using either unidirectional or bidirectional edges between the corresponding nodes. Similarly, we add edges between items and tags as well as users and hashtags.

Through the analysis of the complex network, we want to find the missing links, that is we want to find and suggest new hashtags based on the first tags entered by users that are found to be especially suitable to the posted photo, consequently increasing the visibility of the image: namely, socially relevant tags. This analysis can be applied to the definition of prediction algorithms that monitor sudden changes in a network. This study is extremely interesting for the search of trending topics associated to a specific location or a specific user type.

6 CONCLUSION

Our proposal stems from the aim to analyze and use the hashtags on Instagram. We hypothesized the creation of a social graph, interpreting users, photos, and hashtags as graph nodes. The relations between these elements constitute the graph links: unidirectional or bidirectional. We propose to associate typical analysis of those models to the complex network obtained with the above process. We intend to interpret data gathered from it as useful tools for marketing operations, so as to improve brand awareness. In particular, we want to find missing links to define new hashtags relevant to the pictures uploaded on Instagram and to the profiles of specific communities, so as to give a higher visibility to profiles. In the future, we believe that this analysis might be used for the interpretation of the most relevant informative content for a specific user type and in a specific location. In the tourism industry, for example, the complex network and the study of the missing links could provide, in a semi-automated way, tour routes associated to different user types that could be found though pictures posted on Instagram. A user that visits a certain region and posts pictures of its monuments could automatically receive new suggestions of interesting spots in their itinerary from the application.

REFERENCES

Ames, M., Naaman, M., 2007. Why we tag: motivations for annotation in mobile and online media, in: *Proceedings of the 25th ACM Conference on Human Factors in Computing Systems* (CHI'07), 2007, pp. 971–980.

Angius A., Concas G., Manca D., Pani F. E., Sanna G., 2014. Classification and indexing of web content based on a model of semantic social bookmarking. In: *Proceedings of the 6th International Conference on Knowledge Management and Information Sharing*, KMIS 2014, Rome, Italy, 21-24 October 2014. ISBN: 978-989-758-050-5

Bischoff, K., Firan, C.S., Nejdl, W. R., 2008. Can all tags be used for search? In: *Proceeding of the 17th ACM Conference on Information and Knowledge Management* (CIKM'08), pp. 203–212.

Bruns, A., Burgess, J., 2011. "The use of Twitter hashtags in the formation of ad hoc publics," *paper presented at the European Consortium for Political Research conference*, Reykjavik (25–27 August), at http://eprints.qut.edu.au/46515/, accessed 14 October 2014.

Cantador, I., Konstas, I., & Jose, J. M., 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1), 1-15.

Clauset, A., Moore, C., Newman, M.E.J., 2008. Hierarchical structure and the prediction of missing links in networks. Nature 453, 98 - 101.

Clauset, A., Moore, C., Newman, M.E.J., 2007. Structural Inference of Hierarchies in Networks. In *E. M. Airoldi et al. (Eds.): ICML 2006 Ws, Lecture Notes in Computer Science 4503, 1 - 13.* Springer-Verlag, Berlin Heidelberg.

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E, 70(6):066111.

De Gemmis, M., Lops, P., Semeraro, G., Basile. P., 2008. Integrating tags in a semantic content-based recommender, in: *Proceedings of the 2nd ACM Conference on Recommender Systems* (RecSys'08), pp. 163–170.

Fortunato, S., 2010. Community detection in graphs. Physics Report, 486:75–174.

Gemmell, J., Shepitsen, A, Mobasher, M., Burke, R., 2008. Personalization in folksonomies based on tag clustering, in: *Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*.

Girvan, M., Newman, M. E. J., 2001. Community structure in social and biological networks. Proc. Natl. Acad. Sci. U. S. A., 99 (cond-mat/0112110):8271–8276.

Gleiser P. M., Danon, L., 2003. Community structure in jazz. Advances in Complex Systems, 6(4):565-573.

KONECT, 2015. Jazz musicians network dataset - KONECT.

Laniado, D., & Mika, P. (2010). Making sense of twitter. In The Semantic Web–ISWC 2010 (pp. 470-485). Springer Berlin Heidelberg.

Lovasz, l., 1996. Random walks on graphs: a survey, Combinatronics 2, 1–46.

Marlow, C., Naaman, M., Boyd, D., Davis, M.,. 2006. HT06, tagging paper, taxonomy, flickr, academic article, toread, in: *Proceedings of the 17th ACM Conference on Hypertext and Hypermedia* (Hypertext'06), pp. 31–40.

Mathes, Adam, 2004. Folksonomies-cooperative classification and communication through shared metadata.

Mika, P., 2005. Flink: semantic web technology for the extraction and analysis of social networks, Journal of Web Semantics 3 (2–3) - 211–223.

Milgram, S.,1967. The small world problem, Psychology Today 2, 60–67.

Newman, M. E. J., 2003. The structure and function of complex networks. SIAM REVIEW, 45:167–256.

Shen, K., & Wu, L., 2005. Folksonomy as a complex network. arXiv preprint cs/0509072.

Valverde S. Cancho, Sole R.V., Scale free networks from optimal design. Europhysics Letters, 60, 2002.

Valverde S., Sole R.V., 2003. Hierarchical small worlds in software architecture. arXiv:cond-mat/0307278v2.