

# Summarizing social media content for multimedia stories creation (DISCUSSION PAPER)

Flora Amato<sup>1</sup>, Francesco Moscato<sup>3</sup>, Vincenzo Moscato<sup>1</sup>, Antonio Picariello<sup>1</sup>,  
and Giancarlo Sperli<sup>3</sup>

<sup>1</sup> University of Naples, Federico II (DIETI)

via Claudio 21, 80125, Naples, Italy {flora.amato,vmoscato,picus}@unina.it

<sup>2</sup> Università degli Studi della Campania

Viale Ellittico, 31, 81100, Caserta, Italy francesco.moscato@unicampania.it

<sup>3</sup> CINI (Consorzio Interuniversitario Nazionale per l'Informatica)

Via Cinthia, 80126, Naples, Italy giancarlo.sperli@consorzio-cini.it

**Abstract.** This article represents an extended abstract of our previous work on multimedia summarization. In particular, we propose a novel summarization technique of social media content for *multimedia stories* creation, using a graph-based modeling approach and influence analysis methodologies to detect the most important multimedia objects related to one or more topics of interest. Consecutively, from the list of candidates, we obtain a multimedia summary exploiting a summarization model that satisfies several properties such as *Priority* (w.r.t. user keywords), *Continuity*, *Variety* and *not Repetitiveness*. The summary objects are finally arranged in a multimedia story.

**Keywords:** Social Network Analysis · Summarization · Graph DB.

## 1 Introduction

*Online Social Networks* (OSNs) represent interactive platforms where users comment events and facts, express personal opinions on specific topics, report moments of everyday life and so on, by creating on-line profiles and continuously sharing large amount of information (especially multimedia data). Thus, *social media* content coming from OSNs can be considered, without any doubt, the essence of *Big Data*, providing at the same time new opportunities to investigate and analyze social dynamics within these environments. In the last decade, *Social Network Analysis* (SNA) has been introduced to understand OSNs' structure and properties aiming at supporting a wide range of applications: information retrieval, recommendation, viral marketing, event recognition, expert finding, community detection, user profiling, security, social data privacy, etc., and, in particular, summarization [1].

---

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors. SEBD 2019, June 16-19, 2019, Castiglione della Pescaia, Italy.

The *summarization* process from OSNs can be considered a “distilling” process of the most important information from a variety of logically related sources, in order to obtain a brief and significant version of the social media content. The heterogeneity of the user generated content leads to the creation of a *multimedia story*, i.e. a sort of summary integrating different kinds of multimedia data (e.g. images, videos, audios, texts, etc.).

Let us consider, for instance, the typical behavior of a user that desires to retrieve particular social media content (e.g., photos posted on Flickr or video on Youtube) related to a specific event (e.g., New year’s day in London) described by a set of keywords (e.g. ‘London’, ‘new year’s day’) and concerning a given topic (e.g., holidays). Once determined the most important objects composing the summary, they have to be properly organized in a multimedia story according to some preferences and needs and delivered to final users[5,7].

Concerning the Related Work on social media content summarization, the majority of approaches focuses on how different features of user generated multimedia content crawled by OSNs can support in several ways visual summaries building related to particular events [4,8,6].

Here, we propose a novel summarization technique of social media content for multimedia stories’ creation. In particular, for each *Multimedia Social Network* (MuSN) - i.e. a particular OSN focusing on the management and sharing of multimedia information - we use a graph- based modeling approach and exploit influence analysis methodologies to detect the most important multimedia objects related to one or more topics of interest. Consecutively, from the list of candidate objects we obtain a multimedia summary leveraging a summarization model that considers several properties such as *Priority* (w.r.t. user keywords), *Continuity*, *Variety* and *not Repetitiveness* of generated summaries. The summary objects are finally arranged in a multimedia story and presented/delivered to final users.

## 2 Multimedia Social Network modeling

The proposed model (see [2,3] for more details) permits to represent in an effective way any kind of entities (i.e., *users* and multimedia *objects*) and relationships (e.g., publishing, sharing, commenting, similarity, etc.) in any type of MuSNs (e.g, YouTube, Flickr, Instagram, Last.fm, etc.). In particular, our idea consists of modeling any MuSN as a particular *database graph*.

**Definition 1 (MuSN).** A MuSN (*Multimedia Social Network*) is an undirected edge-labeled graph  $G = (V, L, E)$ ,  $V$  being the set of graph vertices, representing main entities of a social network,  $L$  being a set of labels (belonging to a given vocabulary), describing the different kinds of relationships that can occur among the social network entities; and  $E \subseteq V \times L \times V$  being the set of edges;  $V$  and  $E$  being abstract data types with a set of properties (expressed using several attributes that can be different depending on the type of nodes and edges).

*Example 1 (Example of MuSN).* In the case of *Flickr*, entities of the social network are *Users*, *Groups* and *Photos* ( $V = U \cup Gr \cup P$ ). Users, Groups and Image properties can be described leveraging proper attributes (e.g., username, name, surname, number of followers, etc. for Users; title, description, number of photos, number of users, etc for Groups; and title, description, number of favorites, tags, etc. for Images). Labels correspond to the several activities ( $L = \{\text{'publishing'}$ ,  $\text{'following'}$ ,  $\text{'mark\_as\_favorite'}$ ,  $\text{'comment'}$ ,  $\text{'visualization'}$ ,  $\text{'add\_to\_group'}$ ,  $\text{'discussion'}$  $\}$ ) on the social network (i.e., a user can publish a photo, a user can follow another one, a user can mark as favorite a photo shared by other users, a user can perform a comment on a given photo, a user can visualize a photo, a user or photo can be added to a group and a user can add a discussion to a group). Edges properties are described by proper attributes (e.g., publishing relationships by timestamp and topic, discussion relationships by the timestamp and text of a discussion together with the related answers, etc.).

In addition, particular edge-labeled paths, named *social paths* can be instantiated between two nodes leveraging the different kinds of relationships in MuSN: a given path can “directly” connect two users because they are “friends” , or “indirectly”, as they have commented the same photo, or even, two distinct but similar pictures. Among the different types of social paths, the *relevant social paths* ( $p = (v_i, e_i, \dots, e_k, v_j)$ ) – i.e. particular paths that present certain properties  $\Theta$  – assume a particular importance for the social network analysis purposes. Eventually, we can easily observe that relevant social paths can be obtained as results of a *Regular Path Query* (RPQ) on the graph database representing the given MuSN. To extract relevant social paths, we can first exploit regular path queries and after filtering the obtained results on the base of  $\Theta$ .

*Example 2 (Influential Paths).* A particular kind of relevant social path is constituted by *influential paths* connecting two users; in particular, a user can “influence” in some way other users. As an example, in Flickr a given user  $u_i$  influences another user  $u_j$ , if  $u_j$  adds to her/his favorites any photo of  $u_i$ , or if  $u_j$  positively comments a photo (or one similar to) that  $u_i$  has just published. In Twitter, the influence is mainly related to the re-tweet actions, thus the user  $u_i$  influences the user  $u_j$ , if  $u_j$  has re-tweeted any tweet of  $u_i$ . Similarly in Yelp, the user  $u_i$  influences  $u_j$ , if the user  $u_j$  posts a review of the same sentiment of the review previously posted by  $u_i$  on the same business object. Indeed, the type of influential paths that can be considered depends on the Social Network and on the analytical goals. Concerning the first case of Flickr, all the influential paths can be extracted using the following RPQ:

$$(u_1, e_1 e_2, u_2) \tag{1}$$

where:

$$e_1.\text{type} = \text{"publishing"} \wedge e_2.\text{type} = \text{"mark\_as\_favorite"} \wedge (e_2.\text{time} - e_1.\text{time}) \leq \Delta t$$

constitutes the set of conditions  $\Theta$ , being  $\Delta t$  a given time.

*Example 3 (Recommending Paths).* Another kind of relevant path is the *recommending path* that represents a specific path between two objects by which a given object can “recommend” other objects. As an example, in Flickr it is proper to assume that a given object  $o_i$  recommends another  $o_j$ , if a user visualized/published  $o_i$  and  $o_j$  in consecutive temporal instants of the same browsing session, and the objects are similar or if a user provided two positive reactions or comments to  $o_i$  and  $o_j$  in successive times or if a user marked  $o_i$  and  $o_j$  as favorite in consecutive temporal instants. In the last case, all the recommending paths have the form:

$$o_1, e_1 e_2, o_2 \quad (2)$$

where:

$$e_1.\text{type} = \text{“mark\_as\_favorite”} \wedge e_2.\text{type} = \text{“mark\_as\_favorite”} \wedge \\ \text{similar}(o_1, o_2) \wedge (e_2.\text{time} - e_1.\text{time}) \leq \Delta t$$

constitutes the set of conditions  $\Theta$ , being  $\Delta t$  a given time and  $\text{similar}(o_i, o_j)$  a predicate that is true in the case the two objects are similar in terms of multimedia content. Recommending paths are surely useful for different applications such as recommendation and summarization, being the goal to determine the subset of most relevant objects that could be of interest for a large community of users on the base of their multimedia content. Thus, we can consider recommending paths as a sort of influential paths between objects. The most influential objects are surely good candidates to compose a multimedia summary.

### 3 Story creation

Our goal is to determine the most *important* objects of a MuSN for summarization purposes exploiting an *Influence Maximization* (IM) strategy that allows to obtain a set of suitable candidates (“influentials”), together with the related overall social *importance* w.r.t. a given topic. We successively apply a summarization algorithm on the influentials in order to generate a summary following a set of optimization criteria.

For the summarization goals, we deal with a particular homogeneous graph – *Summarization Graph* – that is derived from a MuSN topic-based view using relevant paths.

**Definition 2 (Summarization Graph).** A Summarization Graph is the triple  $SG = (V; Es; \omega)$ ,  $V$  being a set of vertices related to specific objects of a MuSN,  $Es$  a set of edges and  $\omega$  a weight function. In particular, there exists a unique edge  $e$  between two vertices  $v_i$  and  $v_j$  for all recommending paths connecting  $v_i$  and  $v_j$ . For each edge the related weight will be determined as in the following:

$$\omega(e_{i,j}) = \frac{\sum_{k=1}^M \gamma(p_k(v_i, v_j))}{N_j} \quad (3)$$

$M$  being the number of distinct relevant paths between  $v_i$  and  $v_j$  and  $N_j$  the number of relevant paths of having as destination vertex  $v_j$ .

It is then possible to apply on the SG all the most diffused models and techniques for influence maximization and diffusion to determine the most important objects (*influentials*) of a MuSN. In particular, we have chosen to model how the influence spreads over a network using an *Independent Cascade* (IC) model, where the “activation” of each node is based on the behavior of its active neighbors, and can occur by a single chance. Among all possible approaches defined in the literature, we have used a biologically inspired technique for influence maximization, in particular the *ABC* algorithm based on the bees’ behaviors within a hive (see [2,3] for more details).

In our vision, a *multimedia summary* is a sequence of *summarizable objects* (i.e., influentials represented by multimedia data with topic labels derived from user annotations) that can be semantically correlated (also w.r.t user keywords). On the top of summary definition, we have then introduced four different properties for evaluating the generated summaries (see [2,3] for more details): *Priority*, *Consistence*, *Variety* and *Repetitiveness*.

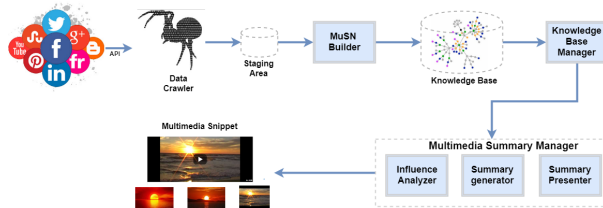
In more details, the *Priority* criterion measures the relevance of objects in the summary with respect to some user keywords; *Continuity* and *Variety* criteria give more importance to multimedia objects published in the same temporal intervals and by different users, respectively; *Repetitiveness* criterion, eventually, measures how semantically similar are the selected objects. Clearly, it is desirable to have a summary with priority and not repetitive contents that presents a temporal continuity and a certain variety in terms of multimedia sources (i.e. users and social networks).

In the previous work, authors have demonstrated that the optimal summary evaluation is a NP-hard problem. To this aim, we provide a greedy strategy that find a sub-optimal solution for the summary evaluation problem in a more efficient way. Our summarization algorithm is based on genetic programming with the following characteristics (see [2,3] for more details):

- it starts using as input the most important  $k$  objects (influentials) computed by the *ABC* influence maximization algorithm applied on the summarization graphs related to all the considered OSNs;
- it works on an initial solution that considers only the priority criterion;
- it uses a *mutation* operator to generate more suitable solutions with respect to all the optimization criteria.

## 4 System Architecture and Implementation

Figure 1 reports an overview of the proposed summarizer system in terms of main components: the *Data Crawler* – that collects information about users, the related generated content and interactions among users and between users and content) and then store such information into a *Staging Area*: the *MuSN Builder* – that builds the social network using a graph database for each considered MuSN; the *Knowledge Base Manager* – that allows to manage the knowledge related to the different MuSNs; and the *Multimedia Summary Manager*, formed by *Influence Analyzer*, *Summary generator* and the *Summary Presenter*.



**Fig. 1.** Overview of the proposed summarizer system

We retrieve data from Flickr and YouTube. The Staging area has been realized using a document oriented database MongoDB that ensures a high horizontal scalability. For the Knowledge Base, we decided to adopt a graph-based approach and to exploit Neo4J functionalities. All the remaining components have been implemented in Scala on the top of Apache Spark, the related data processing libraries and HDFS. Multimedia stories have been realized as HTML pages using javascripts combining AJAX and JQuery technologies.

## 5 Experiments and Results

We used as dataset the YFCC100M<sup>4</sup> multimedia collection. We have thus instantiated the related MuSN using Flickr images and basic relationships (i.e., publishing, following, visualizations, comments, favorites, etc.). We considered images about *building* and *sport*, obtaining as topic-based view a graph with the characteristics depicted in Table 1.

Dataset	Vertices			Edges
	Users	Topic labels	Images	
YFCC100M	1K	40	1.3K	3.8K

**Table 1.** Topic-based view characterization

We performed a human-based evaluation for generated summaries using the *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE<sup>5</sup>) package.

We asked a group of 25 people<sup>6</sup> to generate, for two distinct triples of keywords (“commercial/residential/government”, “soccer/football/rugby”), two different sets of summaries, each one respectively containing 15, 25 and 50 images from the list of candidates computed by ABC algorithm on the Flickr MuSN.

<sup>4</sup> <https://webscope.sandbox.yahoo.com>.

<sup>5</sup> <http://haydn.isi.edu/ROUGE/>

<sup>6</sup> The people involved in the experiments were mainly students from the University of Naples related to the database and multimedia system courses having an account on Flickr.

The first group contains images that according to human judgment maximizing the not repetitiveness criterion and the second one endorses the variety criterion.

After this preliminary step and starting from the 300 obtained summaries, we have built, for each topic, 6 “optimal” summaries (composed by 15, 25 and 50 sentences and respectively giving more importance to variety and not repetitiveness) by considering those objects that have been more frequently chosen by humans. Then, such optimal summaries have been compared with those generated by our summarizer using the same variety and not repetitiveness criteria.

Such combinations have been then considered to obtain all the 12 possible system configurations. We computed system performances in terms of *average recall*, *average precision* and *F-measure* with respect to the human ground truth according to the ROUGE-2 and ROUGE-SU4 methods (see Table 2).

**Table 2.** Comparison of ROUGE values between system generated summaries and human ground truth.

ROUGE-2			
Configuration	AverageR	AverageP	AverageF
high not repetitiveness, n=25, building	0.40014	0.42331	0.41032
high not repetitiveness, n=25, sport	0.38159	0.41372	0.39409
high variety, n=25, sport	0.38443	0.40959	0.39408
high variety, n=25, building	0.37694	0.40598	0.39104
high not repetitiveness, n=50, sport	0.35801	0.39220	0.37296
high not repetitiveness, n=50, building	0.34689	0.38160	0.36218
high variety, n=50, sport	0.33693	0.35843	0.34691
high variety, n=50, building	0.34627	0.34657	0.34513
high not repetitiveness, n=15, building	0.27698	0.30680	0.28899
high not repetitiveness, n=15, sport	0.29604	0.30803	0.30004
high variety, n=15, building	0.24785	0.26503	0.25412
high variety, n=15, sport	0.23099	0.25431	0.24035

ROUGE-SU4			
Configuration	AverageR	AverageP	AverageF
high not repetitiveness, n=25, building	0.43015	0.45132	0.43802
high not repetitiveness, n=25, sport	0.41684	0.43903	0.42302
high variety, n=25, sport	0.40913	0.44888	0.42691
high variety, n=25, building	0.40713	0.43912	0.42069
high not repetitiveness, n=50, sport	0.37835	0.41433	0.39801
high not repetitiveness, n=50, building	0.38752	0.42374	0.40270
high variety, n=50, sport	0.37023	0.39286	0.37943
high variety, n=50, building	0.37796	0.37801	0.37564
high not repetitiveness, n=15, building	0.32701	0.34106	0.33011
high not repetitiveness, n=15, sport	0.30891	0.34301	0.32201
high variety, n=15, building	0.28032	0.30203	0.28998
high variety, n=15, sport	0.26531	0.29301	0.27632

## 6 Acknowledgments

This work was co-funded by the European Union's Justice Programme (2014-2020), CREA Project, under grant agreement No. 766463.



## References

1. Aggarwal, C., Subbian, K.: Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)* **47**(1), 10 (2014)
2. Amato, F., Castiglione, A., Mercurio, F., Mezzanzanica, M., Moscato, V., Picariello, A., Sperli, G.: Multimedia story creation on social networks. *Future Generation Computer Systems* **86**, 412–420 (2018)
3. Amato, F., Castiglione, A., Moscato, V., Picariello, A., Sperli, G.: Multimedia summarization using social media content. *Multimedia Tools and Applications* **77**(14), 17803–17827 (2018)
4. Bian, J., Yang, Y., Zhang, H., Chua, T.S.: Multimedia summarization for social events in microblog stream. *IEEE Transactions on Multimedia* **17**(2), 216–228 (Feb 2015). <https://doi.org/10.1109/TMM.2014.2384912>
5. d’Acierno, A., Gargiulo, F., Moscato, V., Penta, A., Persia, F., Picariello, A., Sansone, C., Sperli, G.: A multimedia summarizer integrating text and images. In: *Intelligent Interactive Multimedia Systems and Services*, pp. 21–33. Springer (2015)
6. Lu, Z., Lin, Y.R., Huang, X., Xiong, N., Fang, Z.: Visual topic discovering, tracking and summarization from social media streams. *Multimedia Tools and Applications* **76**(8), 10855–10879 (2017)
7. Modani, N., Maneriker, P., Hiranandani, G., Sinha, A.R., Subramanian, V., Gupta, S., et al.: Summarizing multimedia content. In: *International Conference on Web Information Systems Engineering*. pp. 340–348. Springer (2016)
8. Qian, S., Zhang, T., Xu, C., Shao, J.: Multi-modal event topic model for social event analysis. *IEEE Transactions on Multimedia* **18**(2), 233–246 (2016)