



HAL
open science

Social media and European cultural routes : Instagram Networks on the Via Francigena

Aurore Bellens, Fabien Eloire, Natalia Grabar, Nathalie Valmond Le Blanc, Eric Kergosien, Marta Severo

► **To cite this version:**

Aurore Bellens, Fabien Eloire, Natalia Grabar, Nathalie Valmond Le Blanc, Eric Kergosien, et al.. Social media and European cultural routes : Instagram Networks on the Via Francigena. The International Conference on Management of computational and collective Intelligence in Digital EcoSystems (MEDES) 2016, Nov 2016, Hendaye, France. pp.122 - 128, <10.1145/3012071.3012082>. <hal-01426834>

HAL Id: hal-01426834

<https://hal.science/hal-01426834v1>

Submitted on 9 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Social Media and European Cultural Routes: Instagram Networks on the Via Francigena

Aurore Bellens
Master RSN
University of Lille
Bâtiment SH2
Cit  Scientifique
33 (0)320434081

aurore.bellens@orange.fr

Nathalie Valmond Le Banc
GERIICO
University of Lille
Domaine universitaire Pont de Bois
59653 Villeneuve d'Ascq (France)
33 (0)320416825

n.valmondleblanc@yahoo.fr

Fabien Eloire
Clerse
University of Lille
Bâtiment SH2
Cit  Scientifique
33 (0)320434081
fabien.eloire@univ-lille1.fr

Natalia Grabar
UMR8163 STL, CNRS
University of Lille
Domaine universitaire Pont de Bois
59653 Villeneuve d'Ascq (France)
33 (0)320416851

natalia.grabar@univ-lille3.fr

Eric Kergosien
GERIICO
University of Lille
Domaine universitaire Pont de Bois
59653 Villeneuve d'Ascq (France)
33 (0)320416825

eric.kergosien@univ-lille3.fr

Marta Severo
DICEN
University of Paris Ouest
200 avenue de la R publique
92001 Nanterre Cedex
33 (0)140977200
marta.severo@u-paris10.fr

ABSTRACT

In the last few years, European Cultural Routes have become important tourist destinations. What makes this kind of tourism interesting is both the complexity of values and significances that motivate it and the variety of social actors (walkers, inhabitants, institutions, commercial actors, etc.) that are involved. Empirical studies on this phenomenon are still limited because they are really time-consuming and resource-demanding. Yet, social media data seem to provide new perspectives on this field by generating digital traces of the tourists' behavior. This paper aims at exploring how social media data can be employed in order to study tourism on European Cultural Routes. We will consider the case study of the platform Instagram on the route "Via Francigena".

Categories and Subject Descriptors

J.4 [Computer Applications]: Social And Behavioral Sciences – *Sociology*.

General Terms

Algorithms, Management, Experimentation, Human Factors, Theory.

Keywords

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MEDES'16, 1-4 November 2016, Hendaye/France.

Copyright   2012 ACM 978-1-4503-4267-4/10/10...\$10.00.

Cultural Routes, Instagram, Social Networks, Cultural Heritage, Spatial Feature, Text Mining.

1. INTRODUCTION

European Cultural Routes are those itineraries based on a historic route, a cultural concept, figure or phenomenon, conveying a transnational importance and significance for the understanding and respect of common European values. Since 1987, they have been protected by an official Programme established by the Council of Europe (<http://culture-routes.net/>). Until now, 33 itineraries have been awarded with this European label and several others are currently applying for a similar certification. Among them, we can mention the Way of St. James to Santiago de Compostela, the Via Francigena between Rome and Canterbury, the Cluniac Sites in Europe, the Viking Routes, etc.

One of the main goals of Programme is favouring touristic attraction and facilitating management of these trails [1]. As a matter of fact, in the last few years, the number of visits of the cultural routes has strongly increased. What makes this kind of tourism interesting is both the complexity of values and significances that motivate the visits and the variety of social actors (walkers, inhabitants, institutions, commercial actors, etc.) that are involved [2]. Yet, empirical studies on this phenomenon are still limited because they are really time-consuming and resource-demanding. For the most famous itineraries, such as the Way of St. James and the Via Francigena, the institutions that manage the routes provide few statistical data mainly about frequentation: how many pilgrims, their age, gender, etc. [3]. Occasionally, more interesting information about the walkers and their motivation or social phenomena along the trail have been collected thanks to fieldworks on small portions of the trail [4].

Yet, results of these studies can hardly be generalized on all the trail.

Today, social media data seem to provide new perspectives on this field of study. By catching digital traces of the tourists' actions, they are expected to provide new information on social interactions and opinions related to places [5]. This paper aims at exploring how social media data can be used in order to study tourism related to European Cultural Routes. This paper presents a preliminary case study, based on the use of the platform Instagram on the route "Via Francigena". Then, methods developed for this case will be reproduced on other itineraries.

Via Francigena is a 1800 km itinerary stretching from Rome to Canterbury across Italy, Switzerland, France and UK. It has been taken by more than 50,000 pilgrims in 2014. While in the past centuries it was mainly a pilgrimage route, today people walking along it have most of the time different motivations: culture, tourism, wellness, environment or spiritual ones.

In Section 2, we introduce the context and the rationale of our work. Section 3 presents our methodology. Section 4 introduces the preliminary results of the project: it first describes social interactions along the Via Francigena and then presents an approach to automatically mine territories along the Via Francigena from Instagram data. Finally, section 5 gives some concluding remarks and perspectives for future research works.

2. STATE-OF-THE-ART

Several scholars have already validated the use of social media data for studying phenomena related to tourism [6]. Indeed, the interest of social media is that any interaction based on them leaves traces that can easily be saved, stored and massively retrieved. This intrinsic traceability of digital platforms promises, if it is controlled by an appropriate methodology, to provide a new source of data for the study of social facts [7]. In the field of tourism, geotagged photos published on social network sites are surely the more exploited object for studying people's activities and behaviour related to touristic places [8]. More particularly, considering the case of cultural routes, Chayeron et al. [9] showed that the analysis of digital traces retrieved on the photo-sharing platform Flickr can help to rebuild the spatial path of a given tourist and to identify the most popular tourist points.

Although there are numerous empirical studies of this type, we agree with [6] that the use of social media data raises several methodological and theoretical issues and, consequently, more research is still necessary. In particular, they risk to produce quite naïve findings that can be already obtained by mainstream tourist studies. This article will contribute to this field of research by investigating how social media data can be used for studying social phenomena related to complex geographical and social touristic objects, such as the cultural routes. Yet, this study doesn't require big data techniques, because the corpus is not so vast and various. As said, we will consider the case of the Via Francigena through the analysis of the traces left by users of Instagram.

Among the existing social network platforms, visual social media such as Flickr, Panoramio and Instagram have emerged as the most suitable sites for studying social phenomena related to touristic places. Lately, Instagram has become the social platform most used by tourists for sharing pictures of the top moments of their trips. From the social scientist's viewpoint, the interest of

Instagram, as all the other visual social media, is that each publication is linked to space and time. Indeed, it is connected to a specific couple of geographical coordinates and fixed in a precise time span. Thanks to this feature, Hochman and Manovich [10], several others after them [11], have elaborated advanced techniques to identify social and cultural patterns based on large visual datasets especially for studying urban tourism.

Even if Instagram is a visual social media based on photo sharing, it is important to highlight that photos are generally accompanied by a textual caption (Figure 1). Caption may include two types of items very useful for tourism research's sake: hashtags and mentions. Hashtags have been already used for studying exchanges content [15] while mentions allow performing social network analysis on Instagram [16].



Figure 1. Example of Instagram photo with caption including mentions.

In particular, several studies related to social media textual data have described the particular way of expressing named entities in natural language from social network data. Among these works, a first family has focused on the identification and the categorization of named entities (person, place et organization) using Wikipedia [12][13]. In these works, the authors seek to take into account only toponym names as places (for examples *Rome*, *Paris*, *London*, etc.). Despite significant works in spatial entities (SE) recognition, a second smaller family has focused on the identification of more complex SE. The TEXT2GEO method [14] presents a hybrid approach that combines natural language patterns and a machine learning approach. The combination of both methods significantly improves precision and recall results¹. However, The main disadvantage of machine learning approaches is that a part of data has to be indexed manually, and that is a complex task and time-consuming.

In this paper, we will focus on these two items: (i) on hashtags, especially to investigate more deeply the spatial entities related to the cultural route cited in the captions of photos related to the Via Francigena; (ii) on mentions, in order to investigate the dynamics of social connections related to the cultural route. In the next paragraphs we briefly present the methodology employed for studying these two aspects and the obtained results.

3. METHODOLOGY

In this paper, we study the content related to the Via Francigena published on Instagram between September 2011 and

¹ Performance measurement criteria in Information Research field

April 2016. Thank to the tool *Hashtag Instagram Explorer* [17] developed by the DMI of the University of Amsterdam, we could extract and analysed 7 034 photos (and related captions) including the following hashtags: #viafrancigena, #francigena, #camminafrancigena, #francigenatrip and #visitfrancigena. We carried out two techniques of analysis: (i) social network analysis and (ii) automated textual analysis.

As regards the first analysis, we looked at whether it was possible to reconstitute social networks between Instagram users based on mentions included in captions. Indeed, Instagram allows users to identify (and “mention”) other users related to the posted photo, and, consequently, create a link with them. Thus it is possible to retrace these links by building a “co-mention” network. In our case, all Instagram accounts can be either receiver or sender.

In order to extract mentions included in our corpus, we, first, developed a Python program capable to identify and extract the name(s) of other user(s) included in each ‘post’. Second, we build a directed graph where nodes are the users and the arcs correspond to mentions of the users. Finally, we used the software Gephi and Pajek to visualize the graph and analyse it, thanks to a social network analysis approach. We considered the following measures: density, centrality, reciprocity and cluster measures.

As regards the second analysis, we seek to identify complex spatial entities (SE) from Instagram textual data in order to apprehend territories that Via Francigena crosses. To achieve this, we propose a semi-automatic method by combining a morpho-syntactic approach with a lexicon approach based on the list the Via Francigena’s stops and on Wikipedia data.

To create the list of SE from Wikipedia, we have selected Wikipedia articles related to places located in regions that Via Francigena crosses. Each SE is enriched with several metadata extracted from the article (city, region, description, geo-coordinates, hyperlink to the article). Then, the indexation of Instagram data using the lexicon allowed us to generate a list of SE useful to describe the Via Francigena. For each SE, one (or two) label is attached to describe its type (stop or other place from Wikipedia).

Finally, a disambiguation step is applied on spatial entities, identified using Wikipedia data. From the list of SE, a SE is validated if its region or city name is identified in one or more Instagram message. Geo-coordinates extracted from Wikipedia allow calculating the distance between each SE and the stops. We have carried out experiments on Instagram data in order to identify the SEs that are related to the Via Francigena stops (5, 10, 15, 20, 30, 50, 100km). A correct distance is validated if the SE candidate is related to the cultural route into Instagram textual data. Then, a SE is validated if it is located less than 15 kilometers from the nearest stop. For example, the SE named « passo della cisa » is identified into 3 messages and characterized by Wikipedia as a place from a region that Via Francigena crosses. This SE is validated because it is located at 6 kilometers from the stop called “Berceto”. The results were evaluated in terms of precision and recall.

4. DISCUSSION

4.1 Studying social interactions along the Via Francigena

In this section, we are interested in the networks that users of Instagram, that have used different hashtags linked to Via Francigena (between August 2014 and April 2016, a total of 4 619 photos), have created when they ‘mention’ each other within their ‘caption’. As said, our data were composed of two kinds of data: first, a list of users, which will be represented by nodes in our graph; and second, a list of directed arcs (based on co-mentions within ‘posts’), which will be represented by links in our graph. The network aggregates 304 nodes and is composed of 442 arcs (see Figure 2).

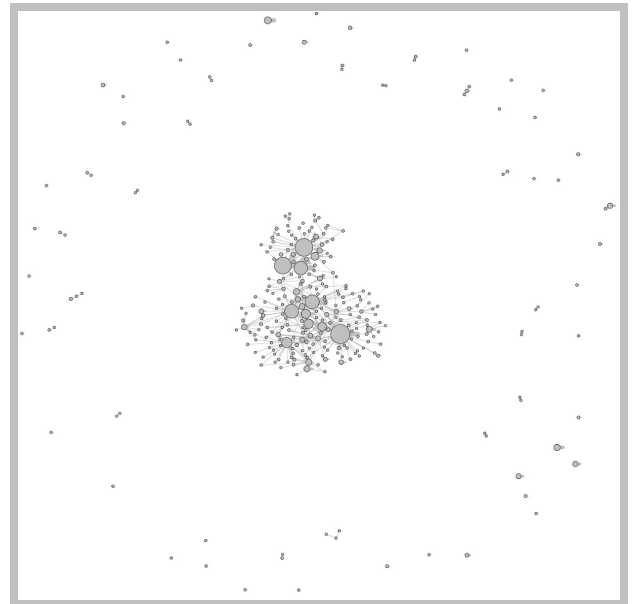


Figure 2. The co-mention network of users using hashtags related to the Via Francigena

The network average degree for a node, that is the mean number of links received and sent by one node, is equal to 18, but this number masks large disparities, because, for instance, the more central node (that is francigenatoscana) has a degree centrality of 122 links, whereas 30% of nodes just have one link. The density of this network² is very weak: 0.5%. What is interesting to note is that we clearly have a ‘core-periphery’ network structure: 84 nodes (28%) belongs to what we call the ‘periphery’, because they own only one tie within the network; whereas the ‘core’ is composed of 220 nodes (72%). The fact that users practice ‘auto-mention’ in the network, lead to the existence of some isolated nodes, which lower the density measure.

In order to analyze the core of the network, we decided to delete the periphery, that is to say 84 nodes (see Figure 3). The density remains weak, 0.7%, but quite higher than that of the global network. Regarding the reciprocity phenomenon³, it appears that less than 5% of the links are reciprocal, which is very

² I.e. The difference between the number of observed links and the number of links if all the nodes were connected.

³ I.e. When A sends a link to B and B sends also a link to A.

low compared to the networks of the real life where reciprocity is a quite common feature of social relations. It indicates that these kinds of digital networks allow people to express their emotions about their experience on the cultural route, but that, objectively, they don't foster contacts and exchanges with others. Similarly, more than 20% of links just arise from people who auto-mention themselves outside any other kind of contact. So mentions are not really meant to create connections with other walkers but mostly to build their own visibility and digital reputation.

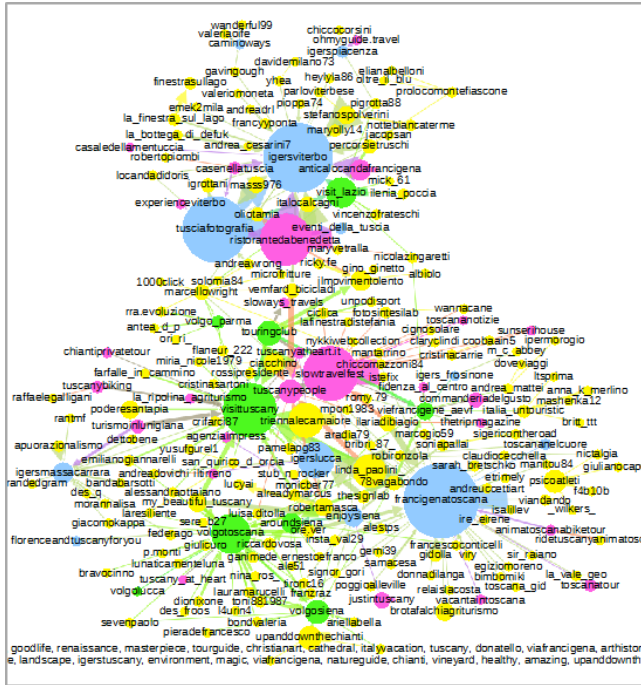


Figure 3. Core of the co-mention network related to the Via Francigena.

This structural description of the network can be fruitfully completed by considering nodes' features: who are the people or entities that have participated to generate this kind of network related to "Via Francigena" cultural route? We have built a typology of users dividing them in four classes according to their nature: class 1 (yellow nodes) is composed of individuals (physical persons, including some politicians); class 2 (green nodes), of tourism associations or institutions; class 3 (pink nodes), of firms involved in the tourism sector (for instance: hotels and restaurants); and class 4 (blue nodes), of Italian tourism websites.

	Core	Periphery	Core + Periphery
Class 1 (yellow) Individuals	173	61	234 (77%)
Class 2 (green) Asso/institution	9	5	14 (5%)
Class 3 (pink) Firms	26	12	38 (12%)
Class 4 (blue) Websites	12	6	18 (6%)
Total	220	84	304 (100%)

Table 1. The 4 classes of nodes within Instagram's co-mention network

Individuals (class 1) represent more than 75% of the network users. In most cases, they are tourists or pilgrims who have walked on Via Francigena: they are not Instagram's specialists but they just share pictures from their different trips. One also finds globe-trotters, such as for instance *italian_backpacker*. Finally, in this class 1, we also find two local politicians: the first one is Nicola Zingaretti (*nicolazingaretti* on Instagram), president of the Lazio region in Italy ; and the second one is Enrico Rossi (*rossipresidente* on Instagram), president of the Tuscany region. It is interesting to note that our methodology based on co-mention networks allows to easily identify 'false' individuals, i.e. computer robots programmed to be "social capitalists" [19] searching for e-reputation⁴, which appear on Instagram⁵ but disappear from our network because they use only popular hashtags to create links and increase their visibility but never use the co-mention's strategy.

Among associations and institutions (class 2), one finds, for instance, the official tourist offices' pages of Lazio, Tuscany and Pavia regions; but also the *visittuscany* page, which is subsidized by the Tuscany region and European Union, whose aim is to promote events and good addresses (i.e. hotels, restaurants) and whose strategy is to re-post, through its profile, pictures from blogs or individuals. Among firms (class 3) one finds lodges or restaurants, which are mentioned by walkers, staying or eating in these places; one also finds firms proposing visits of the region, as *chiantiprivatetour*, which mentions the most popular associations or institutions as to benefit from their visibility. Finally, among tourism websites (class 4), one finds groups of amateurs, whose aim is to share their passion for the region or Via Francigena. For instance, *francigenatoscana* which re-posts the pictures of walkers thanks to the hashtag #francigenatoscana.

In terms of degree centrality, seven accounts are really central: *francigenatoscana* (122 received or sent links), an account which invites people to post photos of Tuscany with the hashtag #francigenatoscana to allow them to repost these photos, *igersviterbo* (111), an account which aims to discover the city of Viterbo, *tusciafotografia* (105), corresponding to a tourism website to discover Tuscany, *slowtravelfest* (85), a music festival, *visittuscany* (84), the official account of the Tuscany Region, *restaurantedabenedetta* (80) a restaurant in the city of Vetralla, and *volgotoscana* (57), an association which aims at discovering the Toscana Region. Interestingly, none of these accounts belongs to individuals' class 1: the three most mentioned users are tourism websites, others two are firms and the last two are in class 2 'associations or institutions'.

All these seven main accounts belong the core-network and play an intermediary role, in the sense of the betweenness centrality⁶. But it is interesting to note that two nodes, that are not central in terms of degree, are, however, central in terms of betweenness. The first one is *Viefrancigena_aevf*, which is the

⁴ These robots are usually hidden behind 'false' young American profiles, who post at least one time per hour.

⁵ We tried to contact them several times but they never answered us, except the fact that they sent us a subscription proposal to our profile, after we had subscribed ourselves to their profile.

⁶ Betweenness centrality is the number of times a user is a path between others actors within the network.

official account of EAVF⁷. The posts in which it is mentioned are those concerning events on the Via Francigena, or those of walkers posting photos of landscapes or monuments of the Via Francigena. The second account is *Ilariadibagio*, which is an Italian photographer, whose account contains essentially pictures of landscapes. She mentions only three other users: a magazine, a tourism website and a festival.

Coming back to the four classes we have qualitatively identified, they are different, not only because of their intrinsic characteristics, but also because of their structural position within the network. The structural analysis of the network (see Figure 4⁸) offers a better visibility concerning the structure of the flows of mentions within and between each class. Some classes are not central in the network, such as class 1 and 3, whereas others are prominent because they send and/or, especially, receive many links, such as class 4 and 2. This structural result is not accidental, but can be explained in terms of the roles played by the different kinds of actors within the network.

Individuals (from class 1) are not central because their posts are seldom, but as they are numerous, they actively participate to the centrality of associations-institutions (from class 2) and tourism websites (from class 4) by diffusing information of these two kinds of actors. Thus, the structural role of individuals is to improve the visibility of others. Yet, it is not their main goal, but an effect of their willingness to share their experience of the cultural route, by posting links related to it. Conversely, the centrality of classes 2 and 4 is generated by the fact that these actors receive links three times more than they send. Finally, the firms (from class 3) send links more than they receive, which indicates that they try to generate activity and obtain visibility around their profile, in a commercial perspective. This strategy seems not entirely efficient at that time because they remain, in average, little central in the network.

Although network density is weak, class 1 is the denser one (0.6% within the core network): indeed, individuals often practice the ‘co-mention’ (which relates to reciprocity phenomenon) between them, compared to the actors of other classes (52 co-mentions on 90 within the whole network). It is due to the fact that some people mention others with whom they walk and have an Instagram account, such as *soniapallai* who belongs to a group of hikers and who mentions her companions in several of her posts. Her pictures can be group photos, but also landscape photos. This confirms that Via Francigena’s walkers are not necessarily pilgrims or foreigners, but also local inhabitants which walk, for pleasure, with their friends.

Moreover, individuals are those who have the most numerous links with the actors of other classes. If they mention a lot tourism websites (from class 4) and associations-institutions (from class 2), this is due to the fact that some accounts of these classes re-post some individuals’ pictures. Thus, walkers employ hashtags from tourism websites and associations-institutions: it is

⁷ European Association of Via Francigena, i.e. the association in charge of the Via Francigena.

⁸ In this graph, we have grouped all users of the same class in a single node (procedure named “shrink” in Pajek software). The size of the nodes is proportional to the number of users of each class (see Table 1), and the weight of links corresponds to the total number of links between each class of users.

a very simple way for them to share their experience of Via Francigena.

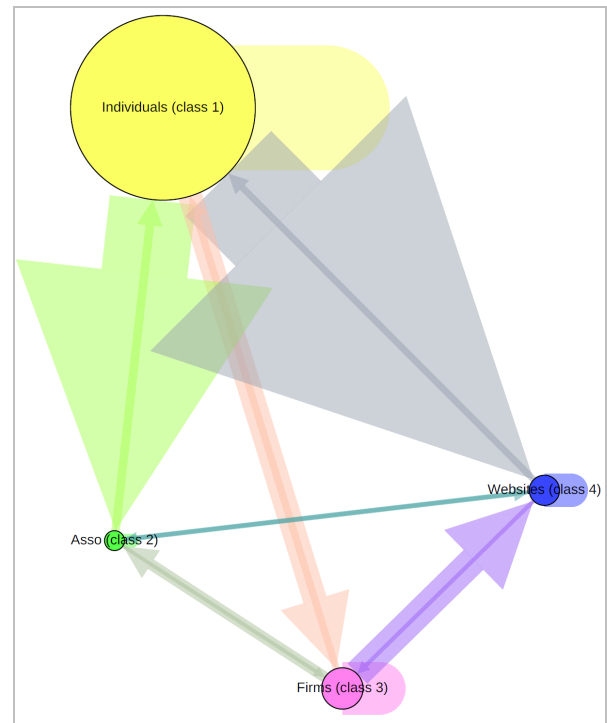


Figure 4. Intra- and inter-class links: a structural approach of the co-mention network.

Conversely, tourism websites and associations-institutions don’t make links between each other (see Figure 4): this may be explained by the fact that these accounts are competitors in the war for the e-reputation. This can be the case, for instance, of accounts that have a similar goal such as two cyclists’ clubs. Or, this can be the case of websites financed by advertising, such as *eventi della tuscia*⁹ which revenues is subject to its audience (in terms number of visitors, numbers of clicks on advertising).

To conclude, social network analysis offers an original point of view to understand how different kinds of actors mobilize social media to achieve their goals. In this case, Instagram’s data allow us to build a network of co-mentions, which reveals some density, centrality, reciprocity and clustering phenomena. In further researches, we plan to build the same kind of networks from Twitter and Facebook, to see if there are common users, roles and practices on these social media.

4.2 Mining spatial entities along the Via Francigena

The corpus employed in the second analysis is constituted by 2598 Instagram posts (in Italian, English and French) published between September 2011 and January 2016. As previously stated, our methodology is organized in two phases: (i) extraction of stops and SE related to the Via Francigena from Wikipedia and (ii) disambiguation of SE.

⁹ <http://www.eventidellatuscia.it>

Based on these first exploratory treatments, we can provide some statistical figures. Firstly, 2099 of 2598 posts contain one or more relevant SE. In them, 12917 occurrences of SE have been identified: 4359 related to Via Francigena stages and 8558 related to other ES identified on Wikipedia. Figures 5 and 6 highlight the most popular ES respectively stages and SE from Wikipedia.

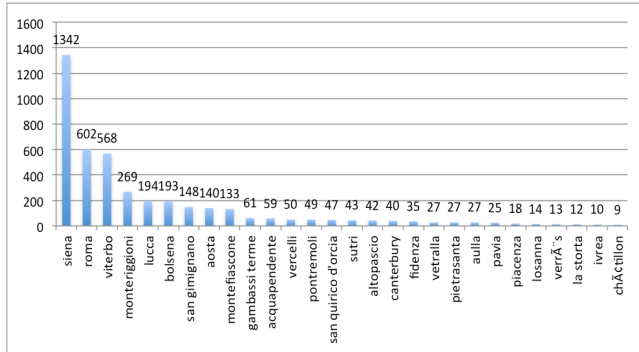


Figure 5. Distribution of the Via Francigena stops by counting the number of occurrences in Instagram data

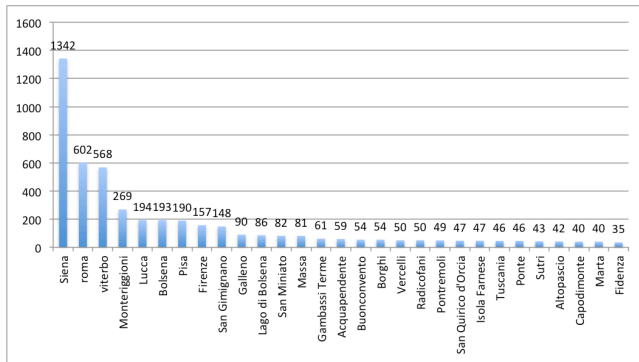


Figure 6. Distribution of the Wikipedia SE by counting the number of occurrences in Instagram data

On 5755 occurrences of SE from Wikipedia, there are 89 different SE including 28 stages. Besides the most popular stops are located in Italy, we identified 40 occurrences of “Canterbury” and 9 occurrences of “Châtillon”.

The Wikipedia data allowed finding other types of SE than cities. For instance, in Figure 6 Bolsena Lake appears 86 times in the corpus. We also can identify churches and cathedrals that are situated near a stop. Moreover, as regards the mentions of countries, not surprisingly we noticed that the country appearing most often is Italy which represent 95% of the countries mentions. The other countries are very poorly represented with 3% for Switzerland and 1% for France and England. For the regions of Italy, the region of Tuscany is the most popular with 85% of the SE, 6% for Lazio, 5% for the Aosta Valley and 1% for the Piemonte, Lombardia, Emilia-Romagna and Liguria.

To evaluate our processes, we calculated the precision and recall on a sample of 50 posts. A group of three annotators manually annotated all SE from the 50 posts. This allowed us to evaluate the quality of our variant SE extraction method. On the sample of 50 posts, the recall is 1 and the precision is 0,97. Indeed, irrelevant SE are identified. For example, “Fondo”, which

is a part of the city “Traversella” is also a common noun, and the meaning is “the lake bottom”.

As a final step, a demonstrator has been developed as a web application based on Java/Javascript and Google Maps API technologies. An extract is shown in figure 7. This demonstrator allows a user to visualize the Via Francigena cultural route (green color); and to show on the map relevant information related to SE identified using Wikipedia (red color). Links to complete description of SE available on Wikipedia are being integrated.

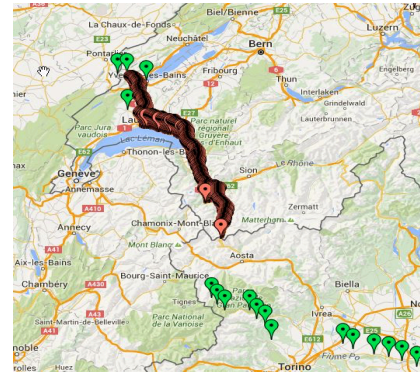


Figure 7. Mapping of SE related to Via Francigena from Instagram data

5. CONCLUSIONS

In the last few years, social media data have emerged as an original and rich source of information for investigating social phenomena related to tourism. Even if numerous empirical studies have been recently developed in this field, several methodological and theoretical aspects need further exploration. In this paper, we present the preliminary results of the PEPS Study on “Digital representations of European Cultural Routes” that will contribute to this field of research. Particularly, it focuses on the use of social media data, notably Instagram data, for studying European Cultural Routes. The case of the Via Francigena has been considered.

Based on the preliminary results presented here, we may conclude that the interest of this case is double. Firstly, while previous studies have mainly focused on specific tourist points of interest (notably cities), it shows the potential of social media data for investigating a complex touristic object such as a cultural route. Text related to photos on Instagram has allowed us to identify most popular stops and localities related to the cultural route. In particular, we could develop based on Wikipedia in order to identify geographical places mentioned on Instagram posts. Secondly, this case study shows the interest to analyze textual data in visual social media. In particular, the analysis of mentions included in Instagram posts may allow investigating social interaction related to the cultural route. Yet, as a conclusion, it is also important to remind the limits and precautions necessary for using social media database [18]. Consequently, even if we already obtained interesting results, they need to be validated through further research. Firstly, findings of this case study will be tested through in-deep interviews to some of the identified Instagram users. Secondly, we will incorporate the usage of

wikidata and commons to improve the SE extraction process¹⁰. Thirdly, the methods developed here will be tested on other cultural routes.

6. ACKNOWLEDGMENTS

This research has been funded by the University of Lille and CNRS through a grant PEPS/University of Lille. More information about the global project “Digital representations of European Cultural Routes” at the url www.itineraireculturels.fr.

7. REFERENCES

- [1] Denu, P. 2015. The evolution of the Council of Europe Cultural Routes programme. In Berti, E. et al., *Cultural routes management: from theory to practice. Step-by-step guide to the Council of Europe Cultural Routes*. Council Europe. Strasbourg.
- [2] Mariotti, A. 2015. Tourism and Cultural Routes: clusters, cultural districts and tourism systems. In Berti, E. et al., *Cultural routes management: from theory to practice. Step-by-step guide to the Council of Europe Cultural Routes*. Council Europe. Strasbourg. 115-124
- [3] <https://oficinadelperegrino.com/estadisticas/>
- [4] Conti, E., Iommi, S., Rosignoli, S., & Piccini, L. 2015. *The European Cultural Routes as engine for sustainable development. The case of Via Francigena in Tuscany (No. 581)*. Istituto Regionale per la Programmazione Economica della Toscana.
- [5] Edwards A, Housley W, Williams M, Sloan L, and Williams M, 2014. Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology* 16, 3, 245–260.
- [6] Chareyron G., Da-Rugna, J., and Raimbault, T., 2014. Big data: A new challenge for tourism. *2014 IEEE International Conference on Big Data*, 5-7 DOI=10.1109/BigData.2014.7004475
- [7] Lazer, D., Pentland, A.S., Adamic, L., et al. 2009. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323, 5915, 721.
- [8] Kisilevich, S., Krstajic, M., Keim, D., et al. 2010. Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. *2010 14th International Conference Information Visualisation*. IEEE, 289-296.
- [9] Chareyron, G., Da-Rugna, J., and Branchet B. 2013. Mining tourist routes using flickr traces. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, New York, NY, USA, ACM, 1488-1489.
- [10] Hochman N. and Manovich L. 2013. Zooming into an Instagram City: Reading the local through social media. *First Monday*, 18, 7(1 July 2013) <http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698> doi:10.5210/fm.v18i7.4711
- [11] García-Palomares J.C., Gutiérrez J., Mínguez C. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63 (September 2015), 408-417, DOI=<http://dx.doi.org/10.1016/j.apgeog.2015.08.002>.
- [12] Pospecu M., Grefenstette G., Moëllic P-A. 2009. Mining Tourist Information from User-Supplied Collections. *Proceedings of the 18th ACM conference on Information and knowledge management*, 1713-1716.
- [13] Mohamed A. and Oussalah M., 2014. Identifying and Extract Named Entities from Wikipedia Database Using Entity Infoboxes. *International Journal of Advanced Computer Science and Applications*, 5, 7(September 2014), 164-169 · DOI= 10.14569/IJACSA.2014.050725.
- [14] Tahrat S., Kergosien E., Bringay S., Roche M., Teiseire M., 2013. Text2Geo : from textual data to geospatial information. *WIMS: International Conference on Web Intelligence, Mining and Semantics*, 4.
- [15] Highfield T. and Leaver T. 2015. A methodology for mapping Instagram hashtags. *First Monday*, 20, 1 (5 January 2015) <http://firstmonday.org/ojs/index.php/fm/article/view/5563/4195> DOI= <http://dx.doi.org/10.5210/fm.v20i1.5563>
- [16] Manikonda, L., Hu, Y., and Subbarao K. Analyzing user activities, demographics, social network structure and user-generated content on instagram. *arXiv preprint arXiv:1410.8099*, 2014.
- [17] Borra, E., 2015. *Instagram Scraper*, Amsterdam, NL: Digital Methods Initiative. Available at: <https://tools.digitalmethods.net/beta/instagram/>
- [18] Zeynep T., 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, at <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MEDES'13, October 29-31, 2013, Neumünster Abbey, Luxembourg.

Copyright © 2013 ACM 978-1-4503-2004-7...\$10.00.

¹⁰ https://commons.wikimedia.org/wiki/Category:Via_Francigena

