Project no. 34721

TAGora

Semiotic Dynamics in Online Social Communities

http://www.tagora-project.eu

Sixth Framework Programme (FP6)

Future and Emerging Technologies of the Information Society Technologies (IST-FET Priority)

D3.4: Methods for Tracking Tag Emergence

Period covered: from 01/06/2008 to 31/08/2009
Start date of project: June 1st, 2006
Due date of deliverable: October 15th, 2009
Distribution: Public

Date of preparation: 15/09/2009
Duration: 39 months
Actual submission date: September 15th, 2009
Status: Final

Project coordinator: Vittorio Loreto
Project coordinator organisation name: Sapienza Università di Roma
Lead contractor for this deliverable: UNI-SOTON

Contents

1 Methods for Tracking Tag Emergence **3**

1.1 Spreading 4

1.1.1 User Tagging across Multiple Folksonomies 4

1.1.2 Regular vs Clusterized Tag Occurrences 5

1.2 Context 7

1.2.1 Folksonomy Specialization 7

1.2.2 Tag Specialization 8

1.3 Community Influence 9

1.3.1 Trend Detection in Bibsonomy 9

1.3.2 Group Membership and Tag Similarity 10

1.4 Synopsis 14

Chapter 1

Methods for Tracking Tag Emergence

Facebook, Delicious, YouTube, Flickr, Twitter, etc., are often regarded as individual folksonomy sites, specialising in slightly different areas (e.g. bookmarking, videos, photos, messaging). However, what is often overlooked is the fact that many users of such sites are also active users of several other sites. It is very common for individuals to have multiple accounts in several Web 2.0 sites, hence creating bridges between those sites. Therefore, what takes place in one of those folksonomies is very likely to spread and influence what happens in the other folksonomies through their mutual users.

Previously in D3.5, we introduced the idea of cross-folksonomy integration, and provided some motivating examples as to why it might useful. We also describes some approaches for realising this integration, with some analysis and results. In this deliverable D3.4, we describe our work that relates to tracking the emergence of tags, and the factors that influence that behaviour.

Tracking tag emergence and evolution across folksonomies is a considerably difficult task, and much more research is require to better understand and model all its dynamics. In this report, we describe various pieces of works that form our initial steps towards reaching that ambitious goal.

Studying how tags evolve and emerge requires an investigation of several areas; spreading of tags across groups and folksonomies, context in which those tags have been used, similarity of those contexts, influence of the community structure on tag usage and evolution, etc. In section 1.1 we describe our methods for tracking the use of tags across Delicious and Flickr by individuals who actively use both systems. The aim was to investigate trends in people's tagging patterns across folksonomies, and their emergent distribution.

In Section 1.2 we extend the analysis above to learn about the type of tags that are likely to emerge from each community or folksonomy. Different folksonomies may have slightly different focuses than others, which influences the usage of tags in those communities. Understanding this behaviour, and the categories of tags that are likely to appear in each folksonomy, is important for building better tag emergence tracking tools. In Section 1.2, we also describe our work on tag context similarity, where we learn about the tendency for related tags to appear in the same context. Modeling this phenomena allows for tag emergence to be better predicted, by monitoring the emergence of their related tags.

The choice of tags by individuals could be influenced by the tagging activities of the rest of the community. Section 1.3 revisits our work which was presented in D3.1, where we measure tag popularity over time within a community, taking into account the popularity of the tagged resources, as well as the popularity of the taggers themselves; the individuals who used the particular tag. This study is an important step towards building a tag prediction method, which is sensitive to the influence of the whole community.

1.1 Spreading

1.1.1 User Tagging across Multiple Folksonomies

Using data collected from the tagging activity of 502 individuals with Delicious and Flickr accounts, who we correlated using the Google Social Graph API[1], we were able to determine the tags used in both systems and explore their usage patterns (Szomszor et al., 2008b). Out of a dataset containing $1,639,639$ Delicious Posts and $4,694,161$ Flickr Posts, we obtained a sample tag set with $83,851$ distinct Delicious tags and $149,529$ distinct Flickr Tags. The total number of tags used in both Delicious and Flickr was $28,550$ (or about 10%) as depicted in Figure 1.1. By defining the intersection weight of a tag as the sum of users who have used the tag in both Delicious and Flickr, we are able to construct an intersection tag-cloud (as shown in Figure 1.2) where higher tag intersection weights are depicted using a larger font. One can observe that high-level classifications are popular (such as `architecture`, `design`, `food`), as well as dates (`2006` and `2007`), functional descriptors (`shopping`, `cooking`), locations (`nyc`, `sanfrancisco`), and events (`Christmas`, `conference`).

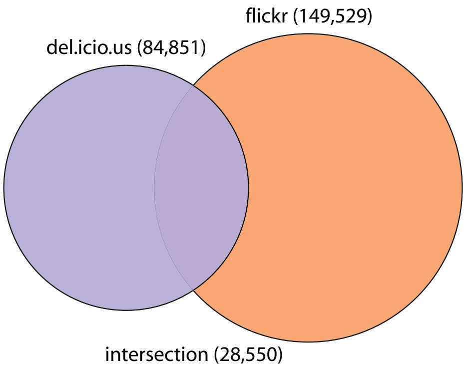

Figure 1.1: Delicious ∩ Flickr Tag Intersection Statistics

2006 2007 advertising airport animals apple architecture art baby beer bike blog book books building california camera canada car cars cat cats chicago christmas church coffee color comics computer conference cooking cute design desktop diy dog drawing email family fashion film fire firefox flash flickr food football france free friends fun funny gallery game games geek google graffiti green guitar halloween history home hotel house humor illustration internet iphone ipod itunes japan kids laptop library light linux logo london mac magazine map maps me media microsoft mobile money movie museum music nature news newyork nintendo nyc office osx painting paper party people phone photo photography photoshop podcast politics radio religion restaurant sanfrancisco school screenshot sculpture security shoes shop shopping software street subway sun technology television tools toys traffic train travel tree tv uk urban usa video wallpaper water weather web wedding windows wine work writing yahoo

Figure 1.2: Delicious ∩ Flickr Intersection Tag Cloud.

Figure 1.3 contains a histogram of the tags found in Delicious that also appear in Flickr vs those that do not. Tags are grouped on the x axis according to their frequency (the most frequent tags on the left, tags that appear only once on the right). The group containing tags with a Delicious frequency between $10,000$ and $100,000$ (i.e. the most common terms - far left of x-axis) is almost entirely represented in Flickr (except for 1 tag). If a tag is used less frequently in Delicious, it is less likely to appear in Flickr. If we assume that more frequently used tags correspond to more general concepts (since they are used to describe a larger collection of objects), then the results show that while folksonomies are likely to share a common set of *high-level* tags that are used frequently, a significant number of tags in the long-tail are not found in both. These less frequently used tags can give some insight into the focus of the folksonomy. For example, the Delicious tags `web2.0`, `Software`, and `Programming` are popular, but do not appear in the Flickr set.

[1] `http://code.google.com/apis/socialgraph/`

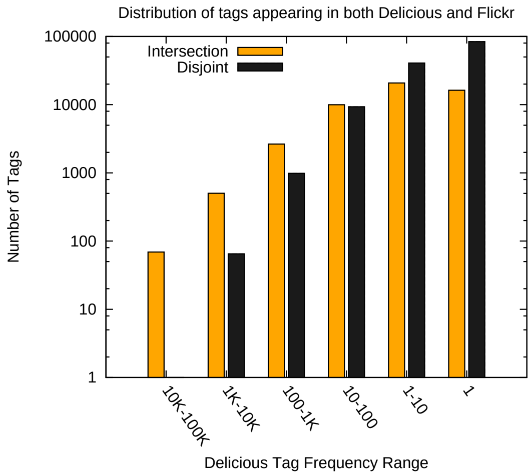

Figure 1.3: A Histogram depicting the distribution of tags appearing in Delicious and Flickr (orange) vs those that appear only in Delicious (black).

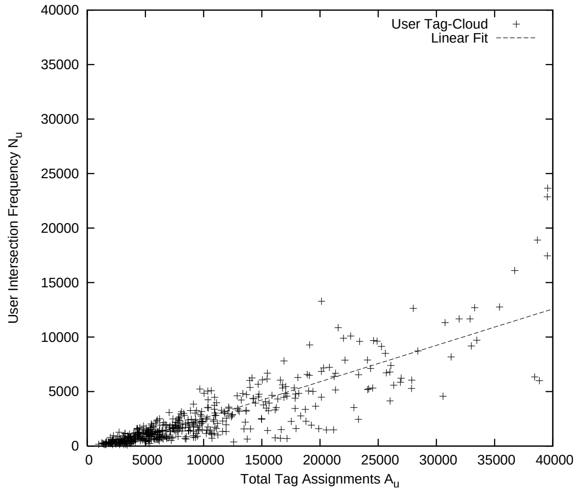

Figure 1.4: Each point represents a user in our sample set. The Intersection Frequency (y-axis) is plotted against the total number of tags assignments made (x-axis).

Conversely, the tags `nokia3660`, `sunny`, and `joseph` appear frequently in Flickr, but do not appear in Delicious. The idea of folksonomy specialisation is discussed further in Section 1.2.

To investigate individual's tagging patterns across folksonomies, we define the user intersection frequency for every user in our sample as the sum of all tag frequencies for tags used in both Delicious and Flickr. Figure 1.4 contains a plot of all user tag-clouds in our sample set with the x-axis representing the total number of tag assignments made, and the y-axis denoting the user intersection frequency. Essentially this plot shows us that as users tag more resources in Flickr and Delicious, their intersection frequency will increase. This tells us that to some degree, a person tagging in both Flickr and Delicious is likely to use some of the same tags, probably those that correlate to topics of interest, events, and places.

1.1.2 Regular vs Clusterized Tag Occurrences

The distribution of tagging with respect to time can differ radically between users and tags. At one end of the scale, some tags are used at a regular interval and distributed evenly in time. These tags are likely to correspond to concepts that an individual has a continued interest in. For example, a computer programmer may regularly tag items with the term `api`. At the other end of the scale, some tags are used only in a short time interval. These bursts of activity are likely to correlate to concepts that a user has only a short lived interest in. For example, after a vacation a user may tag pictures with the place names they visited.

A possible way to discover at which class a tag pertains for a given user, we could use its distribution of inter-arrival times in that user's time ordered stream of tag. Tags with an exponential (poissonian) inter-arrival time distribution are being used with regularity with a characteristic inter-arrival time that can be inferred by the distribution itself. Bursty tags, instead, are characterized by a fat-tailed (generally a power-law) distribution. There might be tags that do not belong to either class, but nevertheless show a clusterized usage, with an inter-arrival time distribution neither poissonian nor fat-tailed. Unfortunately, this method of inspecting the inter-arrival time distribution of tags to discern their usage class is not very feasible since it requires the analysis of a large number of graphs. We present below an alternative and equivalent method capable of detecting the usage class of a tag by means of a single scalar value. We apply this method to the streams of tags of a selected user, known to take part both in Delicious and Flickr.

This method makes use of the two-point tag-tag correlation function defined in (Cattuto et al., 2007). In this work, we restrict to only one particular tag "Tag":

$$C_{\mathrm{Tag}}(\Delta t) = \frac{1}{Lp_{\mathrm{Tag}}^2} \sum_{t=1}^{T} \delta(\mathrm{Tag}(t+\Delta t), \mathrm{Tag}(t)) \,, \tag{1.1}$$

Where $\delta(\mathrm{Tag}(t+\Delta t), \mathrm{Tag}(t))$ is the usual Kronecker delta function, taking the unity value when the tag "Tag" occurs at times t and $t+\Delta t$, and zero otherwise; $T = L - \max(\Delta t)$ is the length of the stream L minus the maximum value of Δt considered; p_{Tag} is the frequency of the tag "Tag" in the stream. A typical picture of the two-point tag-tag correlation function is shown in Figure 1.5 for the tags `filmmaking` and `nytimes` for a fixed Delicious user. A value of C_{Tag} around the unity is a symptom of an uncorrelated situation (a poissonian inter-arrival time distribution), where the probability of finding the tag at distance Δt depends only on the square of that tag's frequency. We define now the scalar quantity:

$$\mathcal{C}_{\mathrm{Tag}} = \frac{1}{50} \sum_{\Delta t=1}^{50} C_{\mathrm{Tag}}(\Delta t) \tag{1.2}$$

With $\mathcal{C}_{\mathrm{Tag}} \simeq 1$ indicating a regularly used tag and $\mathcal{C}_{\mathrm{Tag}} \gg 1$ a tag used in a clusterized way. In Figure 1.5 the tag `filmmaking` is being used in a bursty / clusterized way (see bottom part of the figure), while the tag `nytimes` is being used more regularly in time.

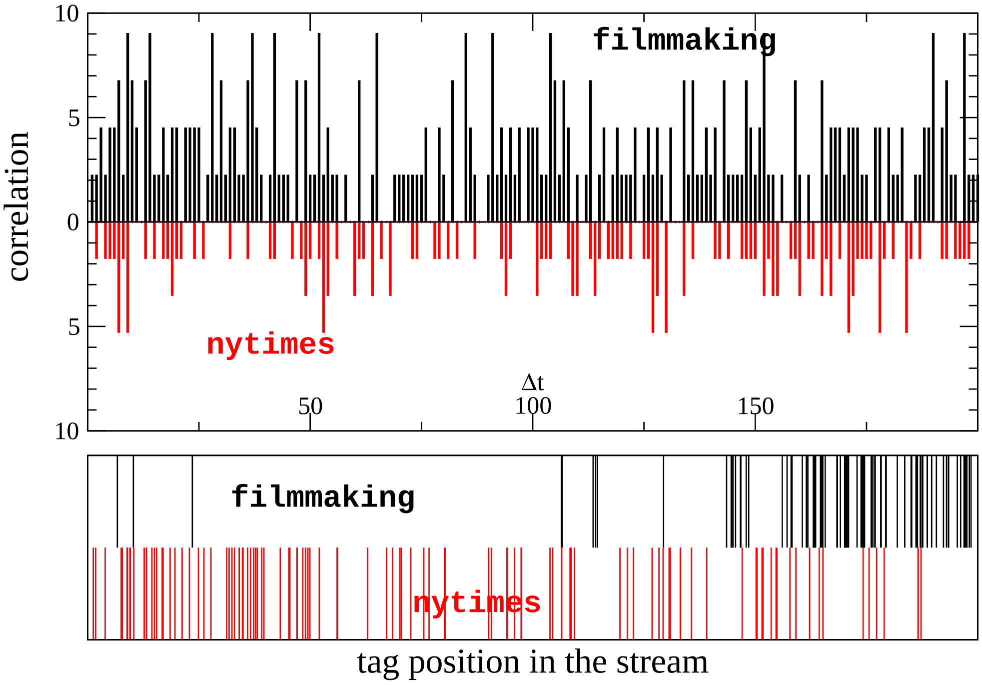

Figure 1.5: Upper figure: Two-point correlation function $C_{\mathrm{Tag}}(\Delta t)$ calculated for a fixed Delicious user and for the tags `filmmaking` (in black) and `nytimes` (red). Lower figure: occurrences of those tags in the Delicious time ordered tag stream of the chosen user. The tag `nytimes` is used more regularly in time than `filmmaking`, which is also indicated by its lower correlation function values.

This method is particularly interesting to apply when we know both the Delicious and Flickr tag streams for a selected user. In Figure 1.6, we show each tag with coordinates given by the values of $\mathcal{C}_{\mathrm{Tag}}$ as calculated both in Flickr and Delicious. The position of tags `nyc` and `music` near the unity in both streams indicates that both are being used with regularity in time and therefore may

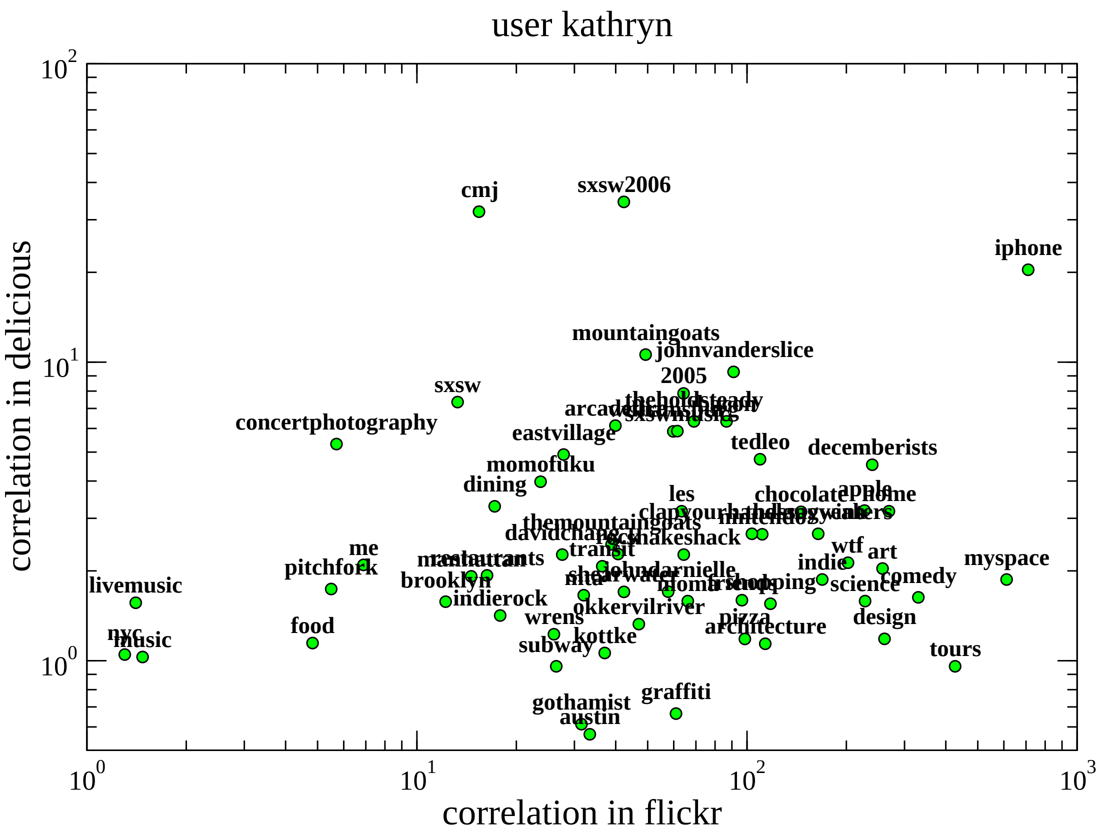

Figure 1.6: Each point in the figure represents a given tag. Its coordinates are given by the values of the indicator $\mathcal{C}_{\mathrm{Tag}}$ in both the Flickr and Delicious tag streams for a fixed user. We might infer that most probably the selected user is a musician living in New York City, since the respective tags (`music` and `nyc`) show coordinates near the unity, therefore are being used regularly in time.

represent the main interests of the user. The higher position of the tag `iphone`, instead, indicates that the interest of the user in this specific product has raised in a short precise moment, probably when the product entered the market.

1.2 Context

Different folksonomies foster different tag dynamics. While it has been realised that broad and narrow folksonomies (typical examples of which are Delicious and Flickr respectively) have different patterns, our research has uncovered evidence that the tags themselves are subject to different properties depending on the context in which they are used. In this Section, we report on ways in which folksonomies and tags exhibit specialisation towards particular semantics, depending on the nature of the items annotated.

1.2.1 Folksonomy Specialization

Over time, as more and more users participate in a particular folksonomy, certain tags begin to dominate. Naturally, these tags tend to highlight the different focus of the folksonomies. However, a simple analysis of the tag strings alone does not provide a complete picture because many of the tags have morphologic variations, synonyms, and polysemes. In work carried out to examine the meaning of tags mined from individual's cross-folksonomy tagging history (Szomszor et al.,

Table 1.1: A Table showing the Wikipedia Categories that describe the most popular tags in our Delicious and Flickr sample dataset.

Delicious		Flickr	
Wikipedia Category	Total Frequency	Wikipedia Category	Total Frequency
design	69,215	travel	51,674
blogs	68,319	australia	51,617
music	45,063	london	46,623
photography	41,356	festivals	42,504
tools	35,795	music	40,943
video	34,318	cats	38,230
arts	29,966	holidays	37,610
software	28,746	family	37,100
maps	26,912	japan	36,513
teaching	22,120	concerts	35,374
games	21,549	surnames	34,947
how-to	19,533	washington	33,924
technology	18,032	given names	32,843
news	17,737	dogs	32,206
humor	15,816	birthdays	22,290

2008a)(and subsequently generate profiles of their interests), we associated tags to Wikipedia[2] Entries and analyzed the distribution of Wikipedia Categories that describe these Entries. We extracted a ranked list of WIkipedia Categories using a dataset of Delicious and Flickr tagging for $1,392$ users, with $138,028$ Delicious tags, and $307,182$ Flickr tags. For each tag that was matched to a Wikipedia Category, the global frequency was incremented by the number of times that tag was used. Table 1.1 shows the top 15 categories found in Delicious and Flickr. These results are a good indication of the types of interest one can learn from the two different domains. Delicious tells us about the bookmarking habits of the user, and subsequently, the topics they are interested in reading about on the Web. For example, `design`, `software`, and `humor` account for many of the posts made. In Flickr, the tags tell us more about locations, events, and people.

1.2.2 Tag Specialization

Prior work on analyzing collaborative tagging systems has given evidence for emergent semantics (Halpin et al., 2006; Hotho et al., 2006). Cattuto et al. (Cattuto et al., 2008) characterized several measures of tag relatedness. Tag context similarity (whereby each tag was described by its cooccurrence-vector with other tags) provided the most precise semantics hereby. Their analysis was based on a dataset containing the $10,000$ most popular tags from Delicious (crawled[3] in 2006), along with all users and resources connected to at least one of those tags in the folksonomy graph.

To extend the work outlined above, and examine the semantics of tags in different folksonomies, we computed the tag context similarity for the $10,000$ most popular Flickr tags using a crawled dataset[4] and compared them with Delicious. Table 1.2 shows the top 5 tags with the highest ranked tag context similarity for 3 tags (`bug`, `windows`, and `net`) that appear in both folksonomies. The tag `bug` is used in Delicious to describe items about computer bugs, whereas in Flickr, `bug` is used to annotate photos containing insects. Similar patterns can be observed for the tag `windows`

[2] http://wikipedia.org/

[3] http://www.uni-koblenz.de/~goerlitz/datasets/tas_delicious2003-2006.gz

[4] http://www.uni-koblenz.de/~goerlitz/datasets/tas_flickr.gz

Table 1.2: Table showing the tag context similarity for 3 tags (`bug`, `windows`, and `net`) that appear in both the Delicious and Flickr folksonomies. These variations clearly highlight the difference in semantics between the two folksonomies.

tag	folksonomy	1	2	3	4	5
bug	*delicious*	bugs	msie	ie6	ie7	internetexplorer
	flickr	wasp	hoverfly	grasshopper	dragonfly	insecte
windows	*delicious*	utilities	utility	opensource	open_source	freeware
	flickr	structure	roof	facade	window	balcony
net	*delicious*	internet	sites	services	www	service
	flickr	rope	fisherman	fishermen	sunny	wind

(the computer operating system and architectural feature) and `net` (as in the Internet and a net for capturing objects e.g. fish, balls, etc). These results demonstrate that tag disambiguation is an important consideration when performing and kind of analysis between different folksonomies. Future work should concentrate on formulating measures to identify such tags so that meaningful comparisons may be made in cross-folksonomy analysis.

1.3 Community Influence

The influence of the community itself on how tags spread and evolve is another important feature to investigate. Users can be influenced with their choice of tags by their online friends, groups, and communities. This section describes some of our work that researches such behaviour.

1.3.1 Trend Detection in Bibsonomy

There are many parameters that could influence the use of a certain tag in communities. For example, if the tag gets used more often by community *leaders*, or used to tag hub or authoritative resources (e.g. photos, websites, videos, articles, etc), then the popularity of the taggers and the tagged resources can also influence the popularity of the tag itself. Methods for measuring popularity of tags over time within a community or sub-community can help to predict its future use and spread over other communities.

We have already reported on some work on detecting tag trends in year 1 of the project. The work was concerned with analysing the dynamics of a folksonomy to detect tag usage. A summary of that work is given here. Details can be found in D3.1 Extracting Emergent Metadata Statistics and Network Metrics in Social Tagging Systems. We have introduced a trend detection measure which determines the popularity of tags, users, or even resources. By applying this measure at different time intervals, we are able to measure how tag popularity is changing over time. This measure uses FolkRank, which differs from conventional co-occurence-based statistical methods by taking into account various elements related to the focus of interest of a group or folksonomy.

Using FolkRank, we can compute topic-specific rankings of users, tags, and resources, then we can monitor those rankings over time to study their evolution. Using data from Bibsonomy, we were able to identify which particular tags were sharply increasing or decreasing in popularity. Such analysis gave us insight into what were the trends in a the community in terms of tags and more generally; topics. In one experiment, we showed how tags related to politics were gaining and losing in popularity over time. Figure 1.7 shows the evolution of all tags that were among the Top Ten in at least one month for the topic ‘politics’. The diagram shows that the early users of Delicious were more critical/idealistic, as they used tags like ‘activism’, ‘humor’, ‘war’, and ‘bushco’[5]. With increasing time, the popularity of these tags faded, and the tags turned to a

[5] In Delicious, ‘bushco’ was used for tagging webpages about the interference of politics and economics in the U. S. administration.

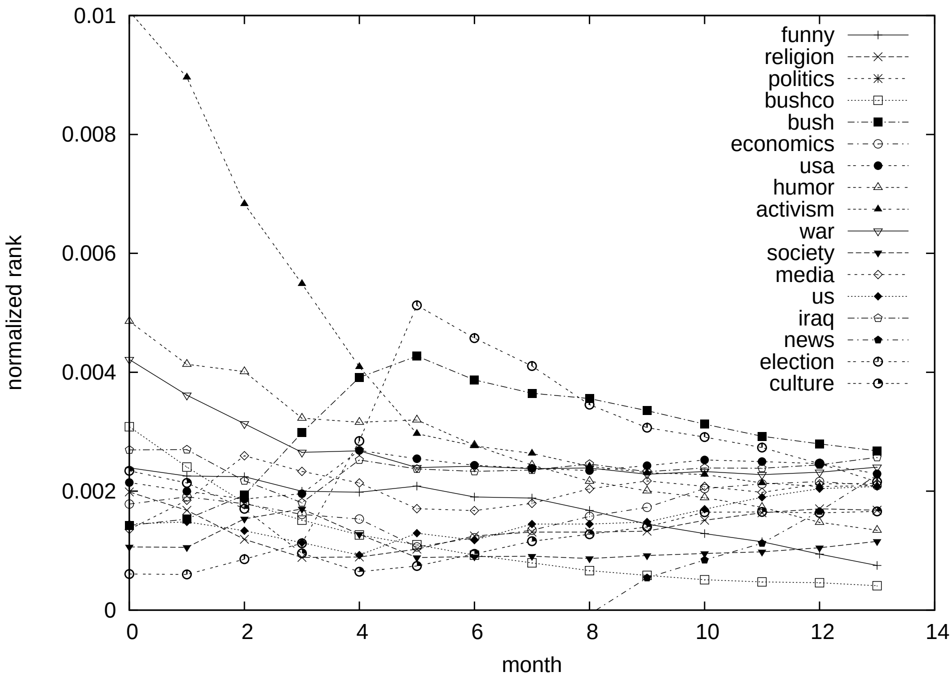

Figure 1.7: Evolution of the ranking of tags related to ‘politics’ over time. ‘Politics’ has value 1.0 due to normalization and is left out for clarity of the presentation of the other values.

more uniform distribution. In particular one can discover the rise of the tags ‘bush’ and ‘election’, both having a peak around the election day, November 2nd, 2004, and remaining on a high level afterwards. This type of analysis demonstrates how tag use could be predicted based on ongoing trends within a community or several communities.

1.3.2 Group Membership and Tag Similarity

The Flickr social network exhibits an explicit social structure, Flickr users express explicitly which peers they know through *contacts*. Unlike most social networks, such relationships are directed. However, contacts are not the only social connection Flickr users can establish, users with common interests can also join ’*groups* to share photos characterized by similar topics, technique, style etc.

The existence of characteristic correlations in the vocabulary of users in social network has been already established in a wide body of previous scientific work. However, it has not been clearly established whether semantical correlations in tagging users arise because of social dynamics taking place within the networks, or of shared background knowledge developed outside the networks. Getting a deeper insight on the reasons behind semantical similarities in folksonomy users is a major goal of the present work.

We have analyzed a dataset that covers about one year (2006) of activity for the Flickr folksonomy, containing 109294825 tag assignments. For each tag assignment, we record the timestamp, author, resource location, and all tags assigned. Besides this posting activity, we have also recorded the explicit Flickr social networks (i.e. contacts and groups memberships). This contact data allows us to build a directed network, where edges go from users to their contacts.

To measure the influence of groups on the tagging patterns of individual users, we measure the semantic similarity in different social networks and for a different definition of semantical similarity. For each network, we have measured the average similarity between the users’ tag clouds over all pairs of neighbors. Since we are dealing with a strongly off-equilibrium systems, drawing conclusions for time-dependent quantities can be problematic. Hence, one has to compare the observed

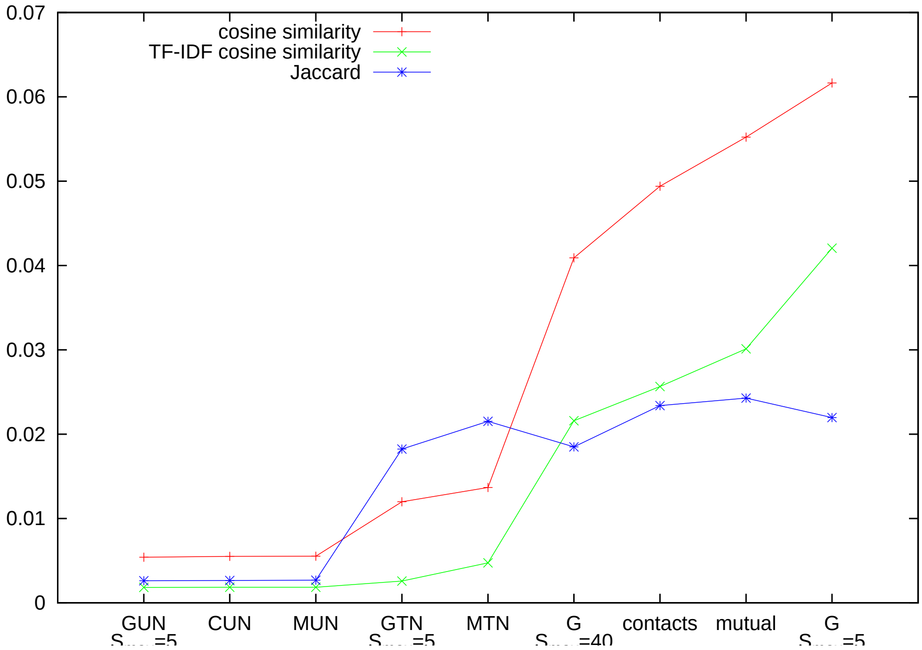

Figure 1.8: The average semantic similarity for three different definition and some Flickr social networks: groups ($S_{max} = 5$) user-based null model, contacs user-based null model (CUN), mutual contact user-based null model (MUN), groups ($S_{max} = 5$) tag-based null model, mutual contact tag-based null model (MTN), groups ($S_{max} = 40$), contacts, mutual contacts, groups ($S_{max} = 5$).

signal with the same measurement performed in a suitable null model.

The first null model we consider is based on the randomization of users. Instead of the real social network, one computes the quantity of interest on a fictitious one, where nodes representing real contacts refer now to random users. Note that the topology of the network has to be mantained in the null model since, in large social networks, the number of neighbors of a node can vary a lot. Therefore, some “hub” node may strongly influence the statistics, and this effect has to be considered in the null model too.

The second null model is based on the randomization of the tag stream, and leaves the social network unchanged. In each time interval, we build the global list of tags with their multiplicity, where each tag appears the total number of times it has been used in the time interval. Then, for each user, each distinct tag is replaced by a random one drawn with uniform probability from the global list of tags, which is assigned the frequency of the real tag.

The results are shown in figure 1.8. To explore the different connection in small and large groups, we have computed the TF-IDF cosine similarity in group-based social networks for different values of S_{max}, showing that assortativity is stronger among founders of a new group, as reported in figure 1.9.

These picture suggests three main conclusions. First, the overlap between users’ tag cloud is very small in all studied cases. But this is hardly surprising: Flickr, unlike other folksonomies such as del.icio.us, CiteULike or Bibsonomy, is a *narrow* folksonomy, where users tend to annotate mainly their own resources. Therefore, their activity is substantially focused on idyiosinchratic topics and encourages the usage of individual tags rather than popular ones.

Second, all networks display assortative mixing, since the average semantical similarity of neighbors is much larger than in the corresponding null model. However, the degree of assortative mixing varies across the different social network considered. Links in the $S_{max} = 5$ social network

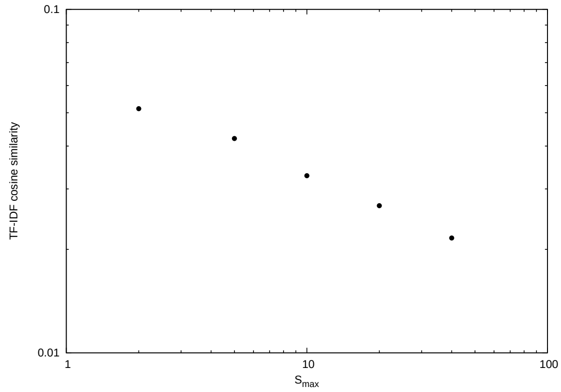

Figure 1.9: The average TF-IDF cosine similarity for social networks based on groups for different values of S_{max}.

have the strongest semantical significance, since the tag clouds of those nodes are, in average, more aligned; at the opposite end, users sharing memberships in larger groups have a looser semantical similarity, and the mutual and directed contact networks display intermediate results, with link contact reciprocity corresponding to a stronger semantical similarity.

Third, and final, the outcomes for different similarity definitions are consistent, and the conclusion above is robust with respect to changes in the measurement method, with two minor exception. This suggests that the problem is correctly defined. In the following section, we will examine only the TF-IDF-based cosine similarity for simplicity. The results obtained for other similarity definitions will not differ significantly.

Alignment dynamics

We focus now our attention on the process through which such similarity is developed. In figure 1.10, we show how the semantic similarity evolves. We have computed the average tag cloud similarity over all pairs of neighbor users in the different social networks as a function of time. The semantic similarity is computed by taking into account the tag cloud of each user from the initial time (cumulative tag cloud) but, at each measurement time (every fifteen days), the average is computed only on users who have been active in the considered period.

In all examined networks, the average neighbors' tag cloud similarity increases in time, whereas in the considered null models the same quantity remains well below. The growth of the similarity changes its rate and stabilyzes roughly after a period of about three months, where it reaches approximatedly it stationary value. The above observation would push toward the conclusion that social neighbors gradually develop a shared vocabulary throughout time, by influencing one to each other in the different manner mentioned above.

But such a picture is contradicted by a further observation. As shown in figure 1.11, we have measured the tag cloud similarity within the time intervals considered in the previous analysis. In other words, the neighbor tag cloud similarity is now measured by including only the tagging activity of the last three months (snapshot similarity). The picture we observe shows that there is no social dynamics within the considered social networks, and in each time interval the average neighbor similarity remains quite stable. However, in each snapshot the similarity in neighbors' tag cloud remains well above the null model benchmark. Thus, social interaction and semantic similarity are indeed positively correlated, but this influence is determined more by the shared background knowledge rather than the social dynamics taking place within the network.

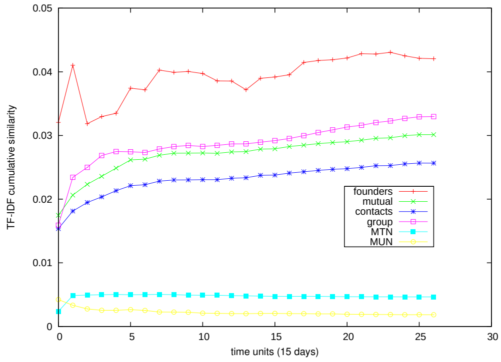

Figure 1.10: The time evolution of the cumulative TF-IDF cosine similarity for different social networks and null models.

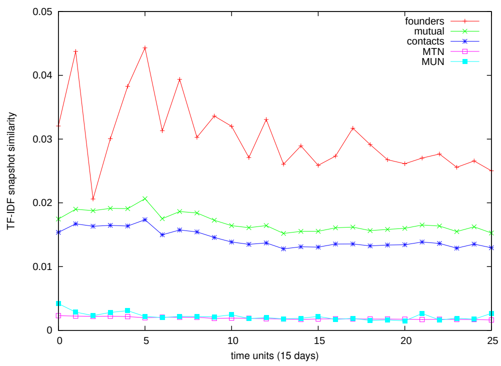

Figure 1.11: The time evolution of the cumulative TF-IDF cosine similarity for different social networks and null models.

Conclusions

We have analyzed a dataset reporting the activity of users in the Flickr network during a period of one year (2006), their social connections and group memberships. We have studied whether social relations in virtual communities may increase semantic similarity between users. We have shown that semantic similarity is larger between socially interacting users, taking it maximum value between group starters.

To uncover the semantic properties of the Flickr social network we have compared it to two possible null models. One is based on the randomization of users in the social network, whereas the other is based on a random re-assignment of tags. Both null model display a pattern of lower similarity, with respect to the real social networks. The dynamics of the tag cloud alignment, however, shows that the observed similarity is not the result of the social interaction within the Flickr groups; rather, it is determined by the existence of a shared background knowledge, and the interaction taking place in the Flickr social network appears to have little effect on the semantics of the folksonomy.

1.4 Synopsis

We have start our ambitious and long research journey towards building comprehensive models and tools for tracking and predicting the emergence of tags, within and across communities. This report presented several of our initial steps in that direction. Several of the ideas and works describes in this document are currently being extended into fully-fledged project proposals, which will hopefully allow us to continue with this important research beyond the end of TAGora.

Bibliography

Ciro Cattuto, Vittorio Loreto, and Luciano Pietronero. Semiotic Dynamics and Collaborative Tagging. *Proceedings of the National Academy of Sciences (PNAS)*, 104:1461–1464, 2007.

Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In Amit P. Sheth, Steffen Staab, Mike Dean, Massimo Paolucci, Diana Maynard, Timothy W. Finin, and Krishnaprasad Thirunarayan, editors, *The Semantic Web – ISWC 2008, Proc.Intl. Semantic Web Conference 2008*, volume 5318 of *LNCS*, pages 615–631, Heidelberg, 2008. Springer. URL `http://dx.doi.org/10.1007/978-3-540-88564-1_39`.

H. Halpin, V. Robu, and H. Shepard. The Dynamics and Semantics of Collaborative Tagging. In *In Proceedings of 1st Semantic Authoring and Annotation Workshop held during ISWC-2006*, 2006.

Andreas Hotho, Robert Jaschke, Christoph Schmitz, and Gerd Stumme. Emergent Semantics in BibSonomy. In *Informatik 2006 - Informatik fur Menschen. Band 2*, volume P-94 of *Lecture Notes in Informatics*, Bonn, oct 2006. Gesellschaft für Informatik. URL `http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006emergent.pdf`. Proc. Workshop on Applications of Semantic Technologies, Informatik 2006.

Martin Szomszor, Harith Alani, Ivan Cantador, Kieron O'Hara, and Nigel Shadbolt. Semantic Modelling of User Interests based on Cross-Folksonomy Analysis. In *submitted to Int. Semantic Web Conf., Karlsruhe, Germany*, 2008a.

Martin Szomszor, Ivan Cantador, and Harith Alani. Correlating User Profiles from Multiple Folksonomies. In *Proc. Int. Conf. Hypertext (HT08), Pittsburgh, PA, USA*, 2008b.