Casting a Web of Trust over Wikipedia: an Interaction-based Approach

Silviu Maniu
Télécom ParisTech - LTCI
Paris, France
maniu@telecom-paristech.fr

Talel Abdessalem
Télécom ParisTech - LTCI
Paris, France
abdessalem@telecom-paristech.fr

Bogdan Cautis
Télécom ParisTech - LTCI
Paris, France
cautis@telecom-paristech.fr

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Algorithms, Experimentation

1. INTRODUCTION

We are witnessing today the rapid emergence of large online communities that contribute and share content on the Web. Examples of applications include online encyclopedias (Wikipedia, Knol), photo sharing sites (Flickr) or rating sites (Epinions). An important trend in such platforms aims at exploiting user relationships, links between users (e.g., social links), in order to improve core functionalities in the system. For instance, search, recommendation or access control can benefit from socially-driven approaches. This is especially the case when links can be viewed as being *signed*, indicating a *positive* or *negative* attitude; possible meanings for positive links could be trust, friendship or similarity, while for negative links they could be distrust, opposition or antagonism. In settings where explicit relationships do not exist, are sparse or are inadequate indicators of one's attitude towards fellow members of the community, it becomes thus important to uncover *implicit* user inter-connections, positive or negative links, from relevant user activities and their interactions.

We present in this short paper a study of the Wikipedia network of contributors. For a collection of 320 articles from the politics domain, starting from the revision history, we investigate mechanisms by which relationships between contributors - in the form of signed directed links - can be inferred from their interactions. We take into account *edits* over commonly-authored articles, activities such as *votes* for adminship, the *restoring* of an article to a previous version, or the assignment of *barnstars* (a prize).

Our model for user relationships is a local one: for a given ordered pair of members of the online community - called in the following the link *generator* and the link *recipient* - it will assign a positive or negative value, whenever such a value can be inferred.

Our preliminary study provides valuable insight into principles underlying a signed network of Wikipedia contributors that is captured by social interaction. We look into whether this network, denoted in the rest of the paper as WikiSigned, represents indeed a plausible configuration of link signs. First, we assess connections to social theories such as *structural balance* and *status*, which have already been considered in online communities [6]. Second, we evaluate on WikiSigned the predictive accuracy of a learning approach for *edge sign prediction*. Equipped with the learning techniques that have been applied in [5] on three explicit signed networks, we obtain good accuracy over the WikiSigned network. By cross training-testing we obtain strong evidence that our network does reveal an implicit signed configuration and that these networks have similar characteristics at the local level.

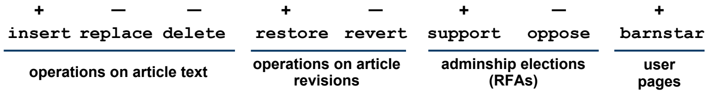
+
insert
−
replace
−
delete
operations on article text
+
restore
−
revert
operations on article revisions
+
support
−
oppose
adminship elections (RFAs)
+
barnstar
user pages

Figure 1: The interaction vector (from generator to recipient)

There are many opportunities for exploiting such a network at the application level, and we briefly discuss one application that also impacts the readers, namely the classification of Wikipedia articles by importance and quality. The intuition here is that such article features depend on how contributors relate to one another.

A core contribution of this paper is a thesis: user interactions in online social applications can provide good indicators of implicit relationships and should be exploited as such.

Main related work. To the best of our knowledge, this is the first study on inferring a signed network (a "web of trust") directly from user interactions. The work that is closest in spirit to ours uses a semi-supervised approach and existing links to build a predictor of trust-distrust from interactions in Epinions [7]. Several papers deal with edge sign prediction using an existing network, among which [4, 5] (see also the references therein). These approaches use the explicit signed network, either for verifying the accuracy of the predictor or as a basis for the inference of new links. In [3, 2], a contributor reputation system and a measure of trustworthiness of text are derived based on their interactions over Wikipedia content.

2. METHODOLOGY AND RESULTS

The Wikipedia dataset. We extracted the full revision history of 320 articles, giving us 442297 revisions by 105177 contributors. A contributor to a revision can do one of the following: edit the text or revert to a previous revision. In the case of text modification, we track several metrics: the amount of text *inserted* near the text of another contributor, the amount of the text *replaced* and the amount of text *deleted*. For each revision, we establish ownership at word level based on the text difference between two consecutive revisions of an article. The interaction thus formed is between the author of the current revision and the owners of the text in the previous revision. For revisions that are restored (reverting to a previous revision), we track the author of the restored (previous) revision and the authors of the revisions that were discarded in the process. For a given pair of contributors, we then track the number of revisions authored that have been *restored* and the number of authored revisions that have been *reverted* (i.e., discarded). For deriving the election votes, we retrieved the adminship elections and the votes

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

triad	count	P(+)	lrn	b/s	triad	count	P(+)	lrn	b/s
t_1	1818176	0.97	0.15	/	t_9	2866505	0.93	0.01	/
t_2	164846	0.90	-0.20	/	t_{10}	98883	0.76	-0.21	/
t_3	1888239	0.94	0.01	/	t_{11}	283149	0.83	-0.01	X/
t_4	99776	0.92	0.08	X/	t_{12}	43036	0.78	-0.14	/
t_5	136277	0.53	-0.33	/	t_{13}	171551	0.91	0.03	X/
t_6	25308	0.40	-0.33	X/	t_{14}	99135	0.80	-0.07	X/
t_7	83154	0.44	-0.41	/	t_{15}	62732	0.83	0.03	X/
t_8	24373	0.54	-0.14	X/	t_{16}	16223	0.58	-0.08	X/X

Table 1: Triad statistics. 'X' marks contradiction with theory.

(positive/negative) submitted by our contributors, thus deriving two measures: the number of *support votes* and the number of *oppose votes* between a pair of contributors.

Building the signed network. We use an *interaction vector* as the basis for inferring signed edges between users. This vector contains measures capturing the four types of interactions: text edits, reverts on article versions, adminship votes and barnstars. Figure 1 shows the components of an interaction vector and the sign interpretation of each (positive or negative). For instance, for edits on text we interpret inserts as positive interactions while replacements and deletions of text as negative and we keep their respective counts. This component will be non-empty only if there were at least three text interactions. Note that these vectors denote directed interactions. We obtained in this way 800057 vectors, in which participate 42631 adminship votes and 2913 barnstars.

To infer a signed edge from these interactions, we adopt the following straightforward heuristic. We decide for each of the four dimensions of interactions whether it is overall a positive or a negative contribution by considering the sign that is more present (in the case of barnstars, it can be either positive or non-existent). Then, at vector level, the result (the sign of the edge) is given by a simple voting mechanism (i.e., the sign of the sum over the dimensions).

The WikiSigned network obtained in this way has 71770 nodes and 463312 edges, of which 85.93% are positive (a link proportion that is very similar to the ones of the existing signed networks).

Validation and results. We first analyze the global properties of WikiSigned, checking whether it represents indeed a plausible configuration of link signs. For that, we consider statistics about the link triads that are present in the network, where a triad represents a composition of a link between two nodes A and B and the possible links between them and third party nodes C. Depending on the direction and sign of the link connecting A, respectively B, with C, there are sixteen such types of triads (for more details on triads see the extended version of this paper [1] and [6]). We find that the distribution of triads in WikiSigned as well as the proportion of positive A-B links are very similar to the ones that can be observed in the Epinions or Slashdot datasets (see also [1]).

Then, using the same methodology as [6], we consider the problem of predicting link signs. More precisely, we run logistic regression learning with 10-fold cross-validation based on a feature set consisting of the number of triads of each type in which the link participates. For that, we use a randomly selected (via reservoir sampling) balanced dataset of links involved in at least 10 triads.

The interest of this prediction analysis is twofold. Using learning coefficients provided by the logistic regression (the signs thereof), we evaluate connections of WikiSigned with the social theories of *structural balance* and *status*, which have already been considered in online communities [6]. In short, balance is grounded in a friend/enemy interpretation (e.g., "enemy of my friend is my enemy"). Status posits that a positive (resp. negative) directed link indicates that the generator views the recipient as having higher (resp. lower) status. We find that at a global level WikiSigned is more tailored to a theory of status interpretation (we observe only one contradiction), in line with the similar study from [6].

	Epinions	Slashdot	Elections	WikiSigned
Epinions	0.926	0.905	0.787	0.727
Slashdot	0.929	0.806	0.792	0.732
Elections	0.922	0.895	0.814	0.733
WikiSigned	0.889	0.844	0.784	0.822

Table 2: Training on the row data, testing on the column data.

type	Importance	Quality
contributors	0.683	0.566
contribs.+links	0.740	0.779
incoming pos+neg	0.683	0.500
outgoing pos+neg	0.740	0.574
inside pos+neg	0.700	0.676
all pos+neg	0.807	0.750

Table 3: Predictive rates for the article importance and quality.

We report the triads statistics and the contradictions between the social theories and our learned model for sign prediction in Table 1.

The predictive accuracy we obtained for the WikiSigned network is of 0.822, with an AUC of 0.899. Furthermore, we have also applied the same learning methodology over the three explicit networks considered in [5], asking the following question: how well a predictor learned on one network performs when applied on another network (see Table 2). First, one can notice that our results that use and apply to explicit networks are almost identical to the ones reported in [6]. Then, WikiSigned performs comparable (slightly better) to the election network, in that prediction on itself is worse than self-prediction over Epinions and Slashdot, while learning the predictor on WikiSigned and applying it on both Epinions and Slashdot yields good prediction rates. All this indicates that these networks have quite similar characteristics at the local level, even though our WikiSigned network is inferred from interactions while the other three are explicitly declared by users.

We also investigate the usefulness of having the signed network in applications, by illustrating how link structure can be exploited in the classification of articles. There are two article labels that are explicit in the Wikipedia politics portal: the quality and the priority. In our dataset, we have articles that span the top 3 article classes (FA, A, GA) and all the importances (T, H, M, L). For each label we have used a balanced random sample of articles.

We use a set of 10 features for each article: the number of *authors* plus three features (total, positive and negative) for each of: *outgoing links* (links from the authors towards other contributors), *incoming links* (the links from other contributors towards the authors) and *inside links* (links from authors to authors).

We report the predictive accuracy we obtained via logistic regression in Table 3. Following the intuition that more important articles have a larger participation and thus more links, we tested the predictive power of these two values (*contributors* and *contribs.+links*). While using knowledge about positive or negative links in separation does not provide better accuracy,btheir combination yields significantly better results (*all pos+neg*). This suggests that the quality of an article is not defined solely by its authors, but also by the relationships between contributors. Possible future extensions for this work are discussed in [1].

Acknowledgements. Work supported by the French ANR project ISICIL and the EU project ARCOMEM FP7-ICT-270239.

3. REFERENCES

[1] http://dbweb.enst.fr/research/WikiSigned-extended.pdf.

[2] B. T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, and V. Raman. Assigning trust to Wikipedia content. In *WikiSym*, 2008.

[3] B. T. Adler and L. de Alfaro. A content-driven reputation system for the Wikipedia. In *WWW*, 2007.

[4] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW*, 2004.

[5] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, 2010.

[6] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *CHI*, 2010.

[7] H. Liu, E. Lim, H. W. Lauw, M. Le, A. Sun, J. Srivastava, and Y. Kim. Predicting trust among users of online communities: an Epinions case study. In *EC*, 2008.