Why Do You Follow Him? Multilinear Analysis on Twitter

Yuto Yamaguchi[1] Mitsuo Yoshida[2] Christos Faloutsos[3] Hiroyuki Kitagawa[1]
[1]U. of Tsukuba [2]Toyohashi U. of Tech. [3]CMU
yuto_ymgc@kde.cs.tsukuba.ac.jp yoshida@cs.tut.ac.jp
christos@cs.cmu.edu kitagawa@cs.tsukuba.ac.jp

ABSTRACT

Why does Smith follow Johnson on Twitter? In most cases, the reason why users follow other users is unavailable. In this work, we answer this question by proposing *TagF*, which analyzes the *who-follows-whom* network (matrix) and the *who-tags-whom* network (tensor) simultaneously. Concretely, our method decomposes a coupled tensor constructed from these matrix and tensor. The experimental results on million-scale Twitter networks show that TagF uncovers different, but explainable reasons why users follow other users.

Categories and Subject Descriptors: H.2.8 [Database Management]: Database Applications – Data mining

Keywords: social graph; social tagging; tensor analysis

1. INTRODUCTION

Why does Smith follow Johnson on Twitter? Is it because Johnson posts interesting tweets about politics or just because Smith and Johnson are friends? It is valuable to answer such questions not only for understanding user behaviors, but also for a lot of applications such as user recommendations, ads, and community discovery.

We propose to answer such questions, by doing a joint study, of (a) the *who-follows-whom* network (matrix) and (b) the *who-tags-whom* network (tensor). We show that our proposed method *TagF* tells us the reasons why users follow other users. For example, if we observe that user A follows B and also A tags B by a tag *politics*, we can safely say that the reason why A follows B is that A is interested in B's (interesting) tweets about politics. However, since the who-tags-whom network is extremely sparse, most of who-follows-whom relationships are still not explained. Hence, TagF performs multilinear analysis of the coupled matrix and tensor to find the latent patterns shared on these two types of networks.

Related work. There are only a few studies aiming at explaining the reason why users follow other users [1, 3]. These studies are different from ours in that they use only the who-follows-whom network. There are also a small number of studies analyzing Twitter lists [5, 4]. However, these studies do not focus on the reasons to follow.

Contributions. (a) **Method**: we design a novel method called TagF to uncover *why users follow other users*, which is based on the multilinear analysis of the coupled tensor, (b) **Experiment**: we perform an experiment on the *who-follows-whom* and the *who-tags-whom* networks on Twitter, and (c) **Discoveries**: we qualitatively show from the results that users have different, but explainable reasons to follow other users.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742715.

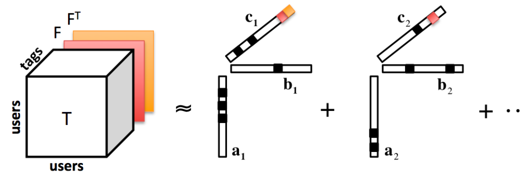


Figure 1: Given user $\times$ user matrices F and F^T, and a user $\times$ user $\times$ tag tensor $\mathcal{T}$, TagF simultaneously factorizes these matrices and tensor into R components, which correspond to latent patterns on the networks.

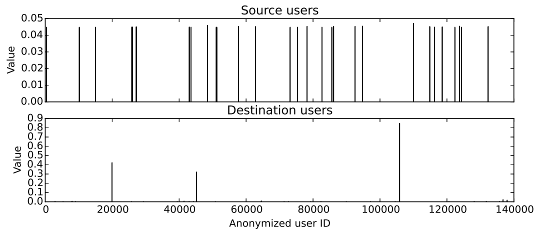
2. OUR APPROACH

We define the problem of uncovering the reason why users follow other users as the coupled tensor analysis problem. We model the who-follows-whom network as a matrix F, and the who-tags-whom network as a 3-mode (user, user, tag) tensor $\mathcal{T}$, where $F_{ij} = 1$ iff user i follows user j, and $T_{ijk} = 1$ iff user i tags user j by tag k. We construct coupled tensor $\mathcal{X}$ by adding matrix F and its transpose F^T into tensor $\mathcal{T}$ as the last two slices along with the third mode of $\mathcal{T}$ (Fig. 1). Note that we employ both F and F^T for multilinear analysis, which enables us to study the forward and mutual relationships between users (see Section 3 for details). Analyzing coupled tensor $\mathcal{X}$, we can find the implicit reasons to follow based on the explicit tagging behaviors of users.

We adopt the PARAFAC [2] as a tool for our analysis. TagF decomposes coupled tensor $\mathcal{X}$ into the sum of R components as $\mathcal{X} \approx \sum_{r=1}^{R} \boldsymbol{a_r} \otimes \boldsymbol{b_r} \otimes \boldsymbol{c_r}$ (Fig. 1). Three vectors of each component correspond to three modes of original tensor $\mathcal{X}$: $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ correspond to source users, destination users, and tags (or reasons to follow), respectively. Each

(a) Hockey component

Tag	**hockey (0.9988)**, sports (0.0197), media (0.0052)
Follow	F**-element (0.0476)**, F^T-element (0.0009)



(b) Writers component

Tag	**writers (0.9983)**, bloggers (0.0002), artists (0.0002)
Follow	F**-element (0.0416)**, F^T**-element (0.0409)**

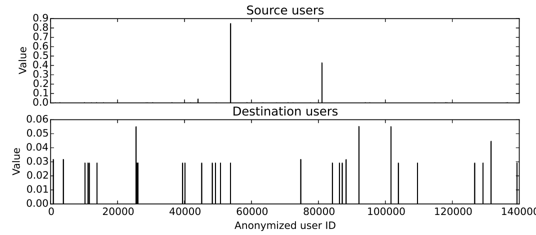


Figure 2: *TagF results make sense*: top 3 tags and F-/F^T-elements in c_r, and top 30 source and destination users in a_r and b_r in hockey component (left) and writers component (right). The reulsts indicate that in the hockey component, a lot of source users unidirectionally follow the representative destination users because only the F-element is large, while in the writers component, two representative source users and a lot of destination users follow each other because both F- and F^T-elements are large.

component can be regarded as a latent pattern in tensor $\mathcal{X}$. For example, in Fig. 1, the first component means that three source users (in $\boldsymbol{a}$) follows one destination user (in $\boldsymbol{b}$) by two reasons (in $\boldsymbol{c}$). The important point is that, in the first component, the last two elements of $\boldsymbol{c}$ corresponding to F and F^T (shown as red and orange in Fig. 1) have large values, meaning this component illustrates the reason why these users follow *each other*. On the other hand, only the red element has a large value in the second component, indicating this component shows just the *one-way* reason why users follow other users. We call these two elements F-element and F^T-element.

Dataset details. We collected the who-follows-whom network directly from Twitter, and also collected *Twitter lists* to construct the who-tags-whom network. Twitter lists can be regarded as the who-tags-whom relationships among users [4] as follows: a user (i.e., tagger) makes a list with a name (i.e., tag) and adds other users (i.e., tagged users) into the list. In this experiment, we use top 50 frequent list names as tags, and discard other lists. As a result, our dataset includes 1,821,432 who-follows-whom relationships and 179,868 who-tags-whom relationships among 147,541 users. Since F is much denser than $\mathcal{T}$, we suppress the effect of F by dividing all the elements of F and F^T by the ratio r of the non-zero elements of F and F^T to that of $\mathcal{T}$ (on this dataset, r is about 20). Our dataset is made available at `http://dx.doi.org/10.5281/zenodo.13966`, and our code is also made available at `https://github.com/yamaguchiyuto/tagf`.

3. RESULTS

Fig. 2 shows two representative components from the result. Each figure shows the top 3 tags and F-/F^T-elements in $\boldsymbol{c_r}$, and the top 30 source and destination users with the largest values in $\boldsymbol{a_r}$ and $\boldsymbol{b_r}$. We can see that each component is represented by the tags *hockey* and *writers*, respectively. These results can be interpreted as follows: (a) in the hockey component (left) a lot of source users unidirectionally follow three representative destination users because these destination users post about hockey, and (b) in the writers component (right) two representative source users and a lot of destination users follow each other because these users are friends in the writers community. The *hockey* component is the one-way pattern because only F-element has a large value, on the other hand, the *writers* component is the mutual pattern because both F- and F^T-element have large values. These patterns of user interactions are indeed observed on the original who-follows-whom network, indicating our TagF can successfully identify the reasons why users follow other users. Although we only show the top 30 source/destination users in Fig. 2, the top 1,000 source users in the *hockey* component have values larger than 0.02, and indeed follow three representative destination users.

4. CONCLUSION

We proposed TagF, which aims at finding the reasons why users follow other users by analyzing the who-follows-whom network and the who-tags-whom network simultaneously. Our contributions are three-fold: (a) method, (b) experiments, and (c) discoveries (see Introduction). It is implied from our discoveries that users have different, but explainable reasons to follow other users.

Acknowledgements: This research was partly supported by the program *Research and Development on Real World Big Data Integration and Analysis* of the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by JSPS KAKENHI, Grant-in-Aid for JSPS Fellows #242322. This material is also based upon work supported by the National Science Foundation (Grants No. IIS-1217559 and CNS-1314632), and by the Army Research Laboratory (under Cooperative Agreement Number W911NF-09-2-0053). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

5. REFERENCES

[1] N. Barbieri, F. Bonchi, and G. Manco. Who to follow and why: link prediction with explanations. In *KDD*, pages 1266–1275, 2014.

[2] R. A. Harshman. Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis. 1970.

[3] A. Tanaka, H. Takemura, and K. Tajima. Why you follow: a classification scheme for twitter follow links. In *HT*, pages 324–326, 2014.

[4] C. Wagner, P. Singer, M. Strohmaier, and B. A. Huberman. Semantic stability in social tagging streams. In *WWW*, pages 735–746, 2014.

[5] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *WWW*, pages 705–714, 2011.