



Towards Improving Sample Representativeness of Teachers on Online Social Media: A Case Study on Pinterest

Hamid Karimi^(✉), Tyler Derr, Kaitlin T. Torphy, Kenneth A. Frank, and Jiliang Tang

Michigan State University, East Lansing, USA
{karimiha,derrtyle,torphyka,kenfrank,tangjili}@msu.edu
<https://www.teachersinsocialmedia.com>

Abstract. Many teachers utilize online social media to supplement their students' needs and enhance their professional activities, curating millions of educational resources. In fact, during the Coronavirus pandemic, online curation of resources provides teachers a repository of materials to provide students in online space. Teachers' engagement online then provides the ability to learn more about how teachers are addressing students' learning needs and potentially improve the quality of the resources they share. Historically, to perform such a study, we often survey some teachers and then leverage their shared resources to investigate education-related research questions. However, this can lead to problems including *sample representativeness* where surveyed teachers may not be representative of the population of teachers in social media. In this paper, we attempt to improve the sample representativeness of teachers on Pinterest. We first survey 541 teachers in the United States as *seed* samples and then collect their online data and social connections on Pinterest. Then, we devise a heuristic that automatically identifies other Pinterest accounts that are likely to be teachers thus improving the sample representativeness. Finally, we evaluate our heuristic with advanced machine learning techniques.

Keywords: Teachers in social media · Sample representativeness · Pinterest

1 Introduction

Unlike traditional resource curation (e.g., asking a colleague or turning to a one's district or state department of education), the diffusion of information from social media to the classroom is significantly more efficient and scaleable. Hence, increasingly, teachers use social media to supplement their instructional resources [7, 8, 11, 16, 17]. According to a survey from RAND Corporation [11], more than 87% of elementary school teachers and 62% of secondary school teachers use Pinterest for professional purposes. Furthermore, since the coronavirus

pandemic in 2020, instructional resources and homeschooling are in the top three of the most frequent searches within Pinterest (personal communication, April 1, 2020). Given the breadth of online activities by teachers, particularly within social media, and its direct diffusion into classrooms, it warrants understanding how teachers harness social media to diffuse their classroom ideas, lessons, and practices to a community of colleagues.

To study teachers in social media, researchers usually follow a bottom-up data collection approach where they first survey teachers *offline* and then retrieve their *online* data [2, 3, 12–14]. However, this common bottom-up data collection method is restrictive since there are potentially many other teachers online that are not included in our sampled and surveyed teachers. In other words, the surveyed teachers may not be representative of the population of teachers in online social media. Moreover, the survey process is usually costly and time-consuming. In this paper, we complement bottom-up data collection approaches by offering a scalable top-down approach. More specifically, we first survey 541 teachers across 5 U.S. states and 48 different districts and then using the surveyed teachers as the *seed* samples, we acquire their Pinterest data (a bottom-up approach). We then propose a top-down approach, building a heuristic that automatically identifies new likely teachers on Pinterest beyond our surveyed teachers. Finally, we use advances in machine learning and social network analysis to evaluate the performance of our heuristic.

2 Related Work

Research shows teachers use various online platforms for educational engagement including Facebook, Twitter, and Pinterest. Steinbrecher and Hart [14] showed that in addition to personal usage, teachers use Facebook for some professional activities such as “classroom support and strategy idea generation”. Authors in [3] explored Twitter usage by K-16 educators and discovered that many educators use Twitter for professional development. In similar studies [1, 4], it was shown that pre-service teachers use Twitter for some professional career development purposes such as resource sharing and connecting to other teachers. Carpenter et al. [2] indicated that teachers use Pinterest to promote educational materials. In particular, they discovered that many individuals were sharing resources curated in TeachersPayTeachers.com, a crucial virtual resource pool where teachers can sell/buy various educational resources. We have discovered similar results for TeachersPayTeachers.com and Pinterest. Some research has endeavored to identify who is curating educational resources. The authors in [13] explored the characteristics of teachers contributing to TeachersPayTeachers.com and attempted to identify the profile of resource curators. Similarly, Schroeder et al. [12] showed that teachers mostly utilize Pinterest to look for educational resources according to their classroom needs. Frank et al. [6] thoroughly analyzed the role of social networks and in particular Pinterest in providing emerging beneficial opportunities for education. Torphy et al. [16] examined the diffusion of educational resources on Pinterest. Their results indicated that direct connection between teachers spurs resource curation. Other work has examined teachers’ social media. The interested reader can refer to [8] for a survey on how to incorporate online social media in educational research.

3 Automated Teacher Identification

Dataset. We surveyed 541 PK-12 teachers across 5 states, 48 districts, and 99 schools. 432 teachers are females, 13 males, and 69 unspecified. For all teachers in our dataset, we acquired their followers and followees (their connections) which resulted in a network with 89,190 nodes (Pinterest users) and 4,379,592 links. Also, for all 89,190 users, we collected their Pinterest data i.e., their shared pins.

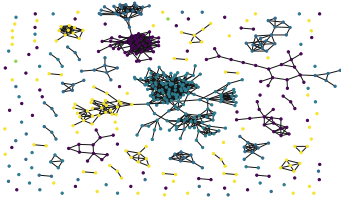


Fig. 1. Network of surveyed teachers on Pinterest where colors represent districts

to connect to their peers and engage in professional career development activities e.g., sharing resources. Second, in general, the network consists of several disjointed sub-networks (components). This disrupts the diffusion of information amongst teachers on Pinterest which plays an essential role in improving the quality of teaching [16]. Third, around 95% of teachers are connected to their peers in the same district, which defies the main strength of social media i.e., breaking physical constraints. Hence, we conclude that we should adamantly attempt to obtain a better picture of the network of teachers on Pinterest as explained in the following.

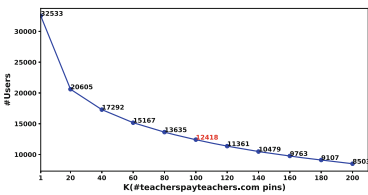


Fig. 2. Number of TPT pins vs the number of users

esult perfectly match these two platforms. Finally, quite often content producers in TPT are teachers/educators who join Pinterest and advertise/share their resources [2]. We also discovered that TPT is the dominant source of resources shared by our surveyed teachers comprising around 50% of the top 5 pin domains. Hence, we hypothesize that the existence of TPT pins in an account is a strong indication that the account belongs to a teacher/educator.

With the above discussion in mind, for all 88,649 other users in our dataset, we process their pins and if for a user the number of his/her TPT pins is more

Top-Down Teacher Identification. As mentioned before, the bottom-up data collection where we first survey teachers offline and then project them into online social media may not properly capture the representativeness of teachers in that online space. To grasp the idea, we visualize the network of our surveyed teachers in Fig. 1. First, we can see there is a considerable number of teachers without any connection to others (96 teachers or around 17% of our surveyed teachers). This is an undesirable property as we expect teachers

In line with previous studies [2, 13], we discovered that the predominant source of educational resources among teachers is TeachersPayTeachers.com (hereafter referred to as TPT). There are several reasons behind this. First, TPT is the largest marketplace of educational resources offering millions of high-quality PK-12 educational resources. Second, image-oriented characteristics of TPT resources and image-based nature of Pinterest

than a threshold K , we mark that user as a teacher. Figure 2 shows the number of users whose K pins' domain is TPT where K changes from 1 to 200. We set K to 100 through which we can mark more than 12,000 users as likely teachers which is almost 23 times larger than the number of surveyed teachers. Note that not necessarily all those marked users are school teachers since they can be other types of educators such as educational organizations, home teachers, parents, and so on. However, as long as their footprint on Pinterest is concerned, they are similar to our surveyed teachers and we keep referring them as *teachers*.

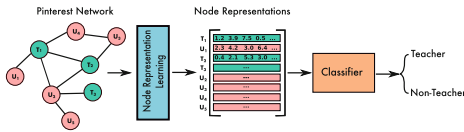


Fig. 3. Evaluating the automated teacher identification

network is an effective approach and used in different applications [5, 9, 10, 18]. In this paper, we adopt the method proposed by Tang et al. [15] known as LINE (Large-scale information network embedding). The size of the representation for each node is 64. Second, on top of learned node representations, we carry out two classifications using Random Forest as the classifier. Both classifiers are tested against 100 surveyed teachers and 100 non-teachers. The first classifier is trained on the rest of 441 surveyed teachers and 441 identified non-teachers. The second classifier is trained on 5500 teachers and 5500 non-teachers (identified using our heuristic). For *non-teacher* samples, we include those having no TPT pin and no educational pin where being *educational* is marked by the Pinterest internal pin labeling system. The accuracy of the first classifier is just 58% while the second one achieves 76%. Hence, we can conclude that our heuristic for automated teacher identification is reliable as it significantly improves the performance of the teacher classification problem.

In the future, we plan to compare the two datasets, i.e., surveyed teachers and the augmented version, from the perspective of structural properties of the two networks as well as behavioral attributes of teachers. Further, we intend to make sense of the diffusion of information among teachers and characterize resources through their diffusion.

Acknowledgment. Research reported in this paper was supported by the Defense Advanced Research Projects Agency under project ID number (10332.02 RaChem PHI) and the Center for Business and Social Analytics at MSU 2017 and 2020, the National Science Foundation, and the William T. Grant Foundation under award numbers (NSF REAL- 1420532, WT Grant - 182764).

References

1. Carpenter, J.: Preservice teachers' microblogging: professional development via twitter. *Contemporary Issues Technol. Teacher Educ.* **15**(2), 209–234 (2015)
2. Carpenter, J., Abrams, A., Dunphy, M.: Educators' professional uses of pinterest. In: *Society for Information Technology & Teacher Education International Conference*, pp. 1925–1930. Association for the Advancement of Computing in Education (AACE) (2016)

Evaluation. The evaluation process of our automated teacher identification is demonstrated in Fig. 3. First, we use the entire constructed network of Pinterest users (89,190 nodes and 4,379,592 links), and extract some features for nodes in an *unsupervised* manner. Feature extraction from a

network is an effective approach and used in different applications [5, 9, 10, 18]. In this paper, we adopt the method proposed by Tang et al. [15] known as LINE (Large-scale information network embedding). The size of the representation for each node is 64. Second, on top of learned node representations, we carry out two classifications using Random Forest as the classifier. Both classifiers are tested against 100 surveyed teachers and 100 non-teachers. The first classifier is trained on the rest of 441 surveyed teachers and 441 identified non-teachers. The second classifier is trained on 5500 teachers and 5500 non-teachers (identified using our heuristic). For *non-teacher* samples, we include those having no TPT pin and no educational pin where being *educational* is marked by the Pinterest internal pin labeling system. The accuracy of the first classifier is just 58% while the second one achieves 76%. Hence, we can conclude that our heuristic for automated teacher identification is reliable as it significantly improves the performance of the teacher classification problem.

In the future, we plan to compare the two datasets, i.e., surveyed teachers and the augmented version, from the perspective of structural properties of the two networks as well as behavioral attributes of teachers. Further, we intend to make sense of the diffusion of information among teachers and characterize resources through their diffusion.

Acknowledgment. Research reported in this paper was supported by the Defense Advanced Research Projects Agency under project ID number (10332.02 RaChem PHI) and the Center for Business and Social Analytics at MSU 2017 and 2020, the National Science Foundation, and the William T. Grant Foundation under award numbers (NSF REAL- 1420532, WT Grant - 182764).

References

1. Carpenter, J.: Preservice teachers' microblogging: professional development via twitter. *Contemporary Issues Technol. Teacher Educ.* **15**(2), 209–234 (2015)
2. Carpenter, J., Abrams, A., Dunphy, M.: Educators' professional uses of pinterest. In: *Society for Information Technology & Teacher Education International Conference*, pp. 1925–1930. Association for the Advancement of Computing in Education (AACE) (2016)

3. Carpenter, J.P., Krutka, D.G.: How and why educators use Twitter: a survey of the field. *J. Res. Technol. Educ.* **46**(4), 414–434 (2014)
4. Carpenter, J.P., Krutka, D.G.: Engagement through microblogging: educator professional development via twitter. *Professional Dev. Educ.* **41**(4), 707–728 (2015). <https://doi.org/10.1080/19415257.2014.939294>
5. Derr, T., Karimi, H., Liu, X., Xu, J., Tang, J.: Deep adversarial network alignment. arXiv preprint [arXiv:1902.10307](https://arxiv.org/abs/1902.10307) (2019)
6. Frank, K., Lo, Y.J., Torphy, K., Kim, J.: Social networks and educational opportunity. In: Schneider, B. (ed.) *Handbook of the Sociology of Education in the 21st Century*, pp. 297–316. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76694-2_13
7. Greenhow, C., Cho, V., Dennen, P.V., Fishman, B.J.: *Education and Social Media: Research Directions to Guide a Growing Field*, vol. 121. Teachers College Record (2019)
8. Karimi, H., Derr, T., Torphy, K., Frank, K., Tang, J.: A roadmap for incorporating online social media in educational research. *Teachers College Record Year Book* (2019)
9. Karimi, H., Derr, T., Brookhouse, A., Tang, J.: Multi-factor congressional vote prediction. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 266–273 (2019)
10. Karimi, H., VanDam, C., Ye, L., Tang, J.: End-to-end compromised account detection. In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 314–321. IEEE (2018)
11. Opfer, V.D., Kaufman, J.H., Thompson, L.E.: *Implementation of k-12 state standards for mathematics and english language arts and literacy* (2016)
12. Schroeder, S., Curcio, R., Lundgren, L.: Expanding the learning network: how teachers use pInterest. *J. Res. Technol. Educ.* **51**(2), 166–186 (2019). <https://doi.org/10.1080/15391523.2019.1573354>
13. Shelton, C.C., Archambault, L.M.: Who are online teacherpreneurs and what do they do? a survey of content creators on teacherspayteachers.com. *J. Res. Technol. Educ.* **51**(4), 398–414 (2019)
14. Steinbrecher, T., Hart, J.: Examining teachers' personal and professional use of facebook: recommendations for teacher education programming. *J. Technol. Teacher Educ.* **20**(1), 71–88 (2012)
15. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: large-scale information network embedding. In: *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077 (2015)
16. Torphy, K., Sihua, H., Liu, Y., Chen, Z.: *Examining the Virtual Diffusion of Educational Resources Across Teachers Social Networks Over Time*, vol. Special issue, Teachers College Record (2018)
17. Torphy, K.T., Drake, C.: Educators meet the fifth estate: the role of social media in teacher training. *Teachers College Record* **121** (2019). <https://www.tcrecord.org/Content.asp?ContentId=23066>
18. Wang, Z., Derr, T., Yin, D., Tang, J.: Understanding and predicting weight loss with mobile social networking data. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1269–1278 (2017)