Characterizing Teacher Connections in Online Social Media: A Case Study on Pinterest

Hamid Karimi, Kaitlin T. Torphy, Tyler Derr, Kenneth A. Frank, and Jiliang Tang

Michigan State University
{karimiha,torphyka,derrtyle,kenfrank,tangjili}@msu.edu

ABSTRACT

Increasingly many teachers are turning to online social media to supplement educational resources and meet students' needs in the classrooms. The diffusion of information from online social media to the classroom is significantly faster than traditional curriculum-based approaches. However, this is contingent upon how well teachers across an online social media network are connected. To understand this, we perform a thorough and large-scale investigation of teacher connections in online social media, which is lacking in the literature. To make this feasible, we construct a large dataset of teachers on Pinterest, an image-based popular online social media. Our dataset includes 540 teachers across 5 states and 48 districts, thousands of connections they have established (either with their peers or some other Pinterest users), and all the resources they have shared in their accounts. Then, taking into account some crucial teacher-related attributes (e.g., their districts, grade levels, etc), we characterize direct and indirect teacher connections. Moreover, we compare the physical (face to face) and virtual (Pinterest) network of our surveyed teachers using several graph-related metrics. The finding in this study can serve as a basis to investigate teachers on social media in a deeper manner.

Author Keywords

Teachers; Social Media; Pinterest

CCS Concepts

•**Applied computing** → **Education;** •**Information systems** → *World Wide Web;*

INTRODUCTION

Online social media platforms have connected billions of people across the globe and their analysis plays a crucial role in many applications [2, 3, 6, 9]. Within education, many teachers utilize online social media to enhance their educational activities. One of the primary drivers of teachers to turn to online social media is to supplement their instructional and educational resources. In the classroom, many teachers encounter needing additional pedagogical resources to improve their students' learning. Traditional means of educational resource curation (e.g., asking a colleague) is time-consuming and not scalable. In contrast, seeking out educational resources from other teachers in online social media is readily accessible.

Although, there is a substantial amount of evidence showing the usefulness of online social media for teachers seeking additional resources [4, 5, 7, 10, 11, 12], a social media service like Pinterest has millions of users and billions of posts. It is likely that teachers and their educational resources are buried. Therefore, it is unclear how teachers that use online social media for professional career development are connected to one another. Thus, it is of great importance to understand and consequently facilitate teacher connections which can have numerous benefits including diffusion of resources in a faster and more efficient manner. Aiming at understanding teacher connections, in this paper, we take two important steps. First, we construct a large dataset of teachers on Pinterest comprising 540 teachers across 5 states and 48 different districts. Second, we thoroughly explore and analyze the data to characterize teachers' connections on Pinterest from multiple perspectives. In particular, we investigative likelihoods for a social tie by considering a set of teacher-related attributes that might relate to the connection a teacher establishes with another one. Importantly, we observe that many teachers on Pinterest are just locally connected to their peers in the same district and the same state. Moreover, we compare the virtual network (i.e., Pinterest) with and the physical network of teachers (i.e., the face to face social network) and delineate their differences and similarities.

DATASET

We surveyed 540 teachers across 5 states, 48 districts, and 99 schools. 428 teachers are females, 13 males, and 99 unspecified. More than 82% of teachers are teaching grades K to 6. For a teacher teaching multiple grades, 12 of them, we considered the highest grade they teach. Table 1 shows the statistics of the Pinterest data. For all teachers in our dataset, we obtained their Pinterest handles (usernames). Then through the API (application programming interface) provided by Pinterest, we collected their data from Pinterest. We retrieved all resources pinned (saved) by our surveyed teachers. In total, we downloaded 1,205,631 pins shared by the end of Feb 2019. Next, for each teacher, we acquired the list of their followers and followees. A follower is someone who follows that specific teacher while a followee is someone who is followed by that teacher. For all followees and followers, we retrieved their pins as well as the list of their connections (i.e., their own followers/followees). Once all connections are determined, we

L@S '20, August 12–14, 2020, Virtual Event, USA.

ACM ISBN 978-1-4503-7951-9/20/08 ...$15.00.
http://dx.doi.org/10.1145/3386527.3405941

constructed the entire network of our Pinterest users including the surveyed teachers and their followers/followees. Following the terminology of the social network analysis and graph theory, *nodes* in our network are Pinterest users, and a *link* between two nodes exists if one follows another one i.e., all follower and followee connections are considered undirected links.

Table 1: Statistics of the Pinterest network

# Teachers	540	# links between teachers	1,059
# Other users	98,667	# Links between teachers and others	117,169
# Total users	99,207	# Total links	6,119,338

TEACHER CONNECTIONS CHARACTERIZATION

To deepen our understanding of teacher connections on Pinterest, we characterize their connections while considering some teacher-related attributes. For these attributes, we characterize direct and indirect teacher connections, which are discussed next. Then we introduce the teacher-related attributes and finally, a comparison between physical and virtual networks is presented.

Direct connection characterization

Given the network of teachers on Pinterest, we evaluate several conditional probabilities in the following format.

$$\mathrm{P}\ (\mathrm{T_i}\ \textbf{attribute}\ \mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}) \quad (1)$$

where $\mathrm{T_i}$ and $\mathrm{T_j}$ are two surveyed teachers, **attribute** denotes an attribute we consider to investigate teacher-teacher connections (will be discussed later in this section), and $\mathrm{T_i} — \mathrm{T_j}$ denotes a connection between teachers $\mathrm{T_i}$ and $\mathrm{T_j}$. Through Eq (1), we attempt to evaluate how likely two teachers have a certain attribute (i.e., $\mathrm{T_i}$ **attribute** $\mathrm{T_j}$) given that the two teachers are directly connected on Pinterest (i.e., $\mathrm{T_i} — \mathrm{T_j}$).

Indirect connection characterization

To better characterize teacher connections on Pinterest, we extend the characterization formulated in Eq (1) beyond a direct connection between two teachers and consider the case when two teachers are indirectly connected by an intermediate user who can be another teacher or a Pinterest user unknown to us. Hence, for some attributes between two teachers, we evaluate a conditional probability in the following format.

$$\mathrm{P}\ (\mathrm{T_i}\ \textbf{attribute}\ \mathrm{T_j} \mid \mathrm{T_i} — \mathrm{U} — \mathrm{T_j}\ \mathrm{AND}\ \mathrm{T_i} - \times - \mathrm{T_j}) \quad (2)$$

where U denotes a Pinterest user bridging teachers $\mathrm{T_i}$ and $\mathrm{T_j}$ and $\mathrm{T_i} - \times - \mathrm{T_j}$ signifies that there is no direct connection between $\mathrm{T_i}$ and $\mathrm{T_j}$.

Teacher-related Attributes

We evaluate teacher connections (either in a direct or indirect manner as explained) by considering two sets of teacher-related attributes, namely **geographical** and **professional** attributes. For geographic attributes, we consider *school*, *district*, *state*, and *physical link*. The reason for including the latter is physical connections are manifested in a face to face social network among teachers, which is obviously bound to geographic constraints. For professional attributes, we consider *grade levels* and number of *shared resource(s)* between two teachers. A shared resource is defined as a pin that is saved (pinned) by both teachers. We consider a binary case whether any number of resources has been shared or none.

Physical network versus virtual network

To help characterize teacher connections on Pinterest in a better way, we compare physical and Pinterest (virtual) networks of teachers as well. Figure 1a and Figure 1b illustrate these two networks, respectively. In the physical network, a link is established between two teachers if one has sought teaching advice from the other. Moreover, we de-identify teachers and denote them in the format **T***N* where *N* is a random number in range [1,540] assigned to each teacher . Note that physical edges are only available for a subset of teachers, namely 104 out of the 540 total teachers.

Table 2: Results of direct teacher connections characterization according to Eq. 1

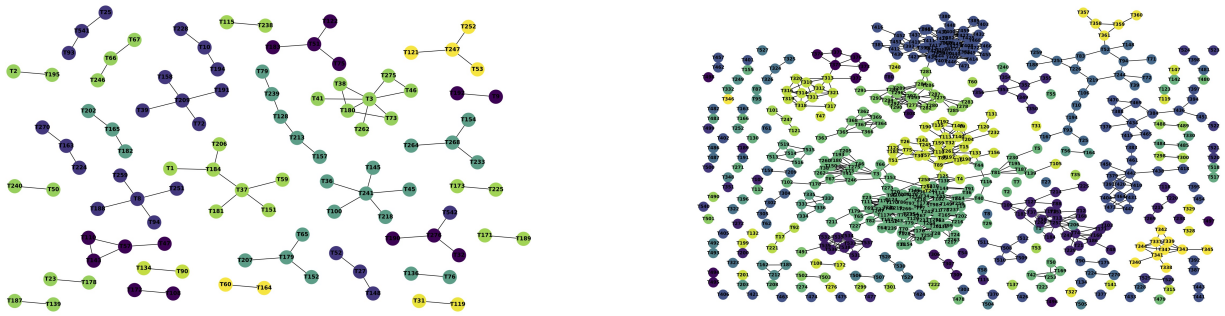
	Attribute	Probability
Geographic	P ($\mathrm{T_i}$ ***the same school*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	557/1059 = 52.13%
	P ($\mathrm{T_i}$ ***different school*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	507/1059 = 47.87%
	P ($\mathrm{T_i}$ ***the same district*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	1016/1059 = 95.94%
	P ($\mathrm{T_i}$ ***different district*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	557/1059 = 4.06%
	P ($\mathrm{T_i}$ ***the same state*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	1056/1059 = 99.71%
	P ($\mathrm{T_i}$ ***different state*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	3/1059 = 0.29%
	P ($\mathrm{T_i}$ ***physical link*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	31/81 = 38.27%
	P ($\mathrm{T_i}$ ***no physical link*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	50/81 = 61.73%
Professional	P ($\mathrm{T_i}$ ***the same grade level*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	230/895 = 25.69%
	P ($\mathrm{T_i}$ ***different grade level*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	665/895 = 74.31%
	P ($\mathrm{T_i}$ ***shared resource*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	1059/1059 = 100%
	P ($\mathrm{T_i}$ ***no shared resources*** $\mathrm{T_j} \mid \mathrm{T_i} — \mathrm{T_j}$)	0/1059 = 0%

Table 3: The results of indirect teacher connections characterization according to Eq. 2

	Attribute	Probability
Geographic	P ($\mathrm{T_i}$ ***the same school*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	878/28040 = 3.13%
	P ($\mathrm{T_i}$ ***different school*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	27162/28040 = 96.87%
	P ($\mathrm{T_i}$ ***the same district*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	5005/28040 = 17.84%
	P ($\mathrm{T_i}$ ***different district*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	23035/28040 = 82.16%
	P ($\mathrm{T_i}$ ***the same state*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	10210/28040 = 36.41%
	P ($\mathrm{T_i}$ ***different state*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	17830/28040 = 63.59%
	P ($\mathrm{T_i}$ ***physical link*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	25/1251 = 2.00%
	P ($\mathrm{T_i}$ ***no physical link*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	1226/1251 = 98.00%
Professional	P ($\mathrm{T_i}$ ***the same grade level*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	3473/24296 = 14.29%
	P ($\mathrm{T_i}$ ***different grade level*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	20823/24296 = 85.71%
	P ($\mathrm{T_i}$ ***shared resource*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	28040/28040 = 100%
	P ($\mathrm{T_i}$ ***no shared resources*** $\mathrm{T_j} \mid \mathrm{T_i} - \mathrm{U} - \mathrm{T_j}$)	0/28040 = 0%

RESULTS AND DISCUSSIONS

In this section, we present the results for teacher connections characterization as well as discussions. Table 2 shows the results for the characterization of direct teacher-teacher connections (i.e, Eq (1)) while Table 3 demonstrates the results for indirect connections between teachers (i.e., Eq (2)). Note



(a) Physical network

(b) Virtual network (Pinterest)

Figure 1: Illustration of physical network and virtual (Pinterest) network. Physical links are not available for all teachers. Node colors represent districts a teachers teaching in (better be seen in color).

that, for each attribute, we have included the conditional probability of its negation as well. Further, Table 4 shows some basic graph-related properties of both the virtual and physical networks. Next, we discuss the results by answering several crucial questions.

Q1: Do geographical attributes affect teacher connections on Pinterest?

The results in Table 2 show that being at the same school does not have much bearing on two teachers being connected on Pinterest where a noticeable number of teachers from different schools are connected to each other. It is promising that teachers are being connected to their peers outside their schools. However, the results are opposite for attributes state and district where we can observe that a large majority of connections are between teachers who are coming from the same state and even the same district. In other words, this shows that whether two teachers are connected is predominately based on if they are coming from the same district/state or not. Thus, the likelihood of teachers being connected outside their states and districts is very small. Interestingly, the results are opposite for intermediate connections where, according to Table 3, we can observe that the likelihood of two teachers from different schools, districts, and states being bridged by a third person/user (conditioned on that they are not directly connected) is significantly high. Hence, this analysis provides an affirmative answer to the crucial question that *are teachers mostly connected to their peers in the same district/state?* The graph-related properties presented in Table 4 corroborate this as well where the number of components of the virtual network is high. A component is a connected sub-network e.g., T101, T247, and T121 in Figure 1b. We see on average each district has 3.2 components and the average component size is 3.48 teachers. Also, Figure 1b demonstrates the localized nature of teacher-teacher connections on Pinterest where nodes within a component tend to have the same color (the same district).

Another noticeable phenomenon is the effect of attribute *physical link*. According to results presented in Tables 2 and 3, the physical connections have a low likelihood to be reflected on Pinterest. We speculate this is because the teachers might not feel to seek further advice from their colleagues online and mostly physical interactions suffice them. We believe following up on this subject deserves further investigation and will leave it for the future.

Q2: Are professional attributes of teachers related to their connections on Pinterest?

As for the grade level, we can observe from Tables 2 and 3 that not necessarily teaching at the same grade drives teacher connections. This is promising as teachers are not confined to their peers at the same grade level and connections are driven the broader notion of teachers rather than a specific grade level. Note that we removed teachers with *unspecified* grade level from this analysis. Moreover, we can observe that for both direct teacher and indirect connections, the presence of a shared resource is strongly related to the connection. For a direct teacher-teacher connection, this is not very surprising since after all teachers connect to their peers to acquire resources. For indirect connections, nevertheless, this is quite interesting and asks for further explanation. First, note that such a resource does not necessarily need to be curated by either of the two teachers and it is possible that they both acquired it from the same source. Second, regardless of the producer or the source of the resources, the value of 100% for the shared resources attribute in Table 3 signifies that the resources are in the interest of both teachers while they have likely been diffused to them via some third party. This tells us that we are dealing with a situation where two teachers are using the same resource diffused through Pinterest connections to them while they are not aware of each other.

Q3: How does the virtual network differ from the physical network?

Now, utilizing several graph-related metrics, we compare the physical network with the virtual (Pinterest) network. From Figures 1a and Figure 1b, we note that both physical and virtual connections are localized regarding the districts that teachers belong to. Numerically, this can be inferred from Table 4 as well where the number of components for both networks is high. Even if we normalize the number of components to the number of teachers and districts, both networks exhibit similar behavior (see the last two rows of Table 4). The density of edges in the virtual network (i.e., Pinterest) is smaller than that in the physical network. Density is defined as the ratio

of the number of connections to the total number of potential connections that could exist. The low value of connection density for the virtual network, again, is due to the scarcity of direct connections between teachers. The maximum number of connections for teachers in the physical network is 7 while this number for the virtual network is 28 where indicates more flexibility is establishing connection online. We also calculate the maximum centrality for both networks. The centrality of a node in a network measures the importance or influence of that node in the network e.g., the number of connections a node has can be considered as its influence. In our analysis, we use Eigenvector centrality [1] since while computing the centrality of a node, it recursively incorporates the centrality of its neighbors as well. We can observe that the two networks have similar maximum centrality values.

Table 4: Graph-related metrics of physical and virtual networks.

Metric \ Network	Physical	Virtual
#Connections	78	1059
Max Connection	7	28
Min Connection	1	0
Avg Connection	1.5	3.9
Max Centrality	0.62	0.27
Connection Density	0.014	0.007
#Components	31	155
#Components / #Teachers	0.29	0.28
#Components / #Districts	3.1	3.2
Avg component size in terms of #Teachers	3.35	3.48

It is worth noting that we have developed a systematic and scalable approach to automatically mark unknown users who are likely to be teachers and thus augment the teacher dataset [8].

CONCLUSION AND FUTURE WORK

In this paper, we thoroughly analyzed teachers' connections using 540 teachers on Pinterest. In particular, we characterized direct and indicate connections between teachers while considering several crucial teacher-related attributes. We discovered that many teacher-teacher connections are confined to geographical factors such as state and district while professional attributes have insignificant impact on driving those connections. Moreover, using some graph-related metrics, we highlighted the difference between the physical network and the virtual network of teachers. The findings in this paper open a door to several exciting future works. First, to induce more teacher-teacher connections, we plan to develop a link recommendation method utilizing the rich structure of the Pinterest network. Second, a deep analysis of the diffusion of resources among teachers and characterizing it based on the properties of resources (e.g., their quality) is another worthwhile direction. Finally, it would be of interest and quite valuable to devise a method to identify educational pins related to PK-12 students.

Acknowledgment. The 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 William T. Grant Foundation under award number WT Grant - 182764, and the National Science Foundation under grant numbers IIS1907704, IIS1928278, IIS1714741, IIS1715940, IIS1845081 and CNS1815636.

REFERENCES

[1] Phillip Bonacich. 2007. Some unique properties of eigenvector centrality. *Social networks* 29, 4 (2007), 555–564.

[2] Tyler Derr, Hamid Karimi, Xiaorui Liu, Jiejun Xu, and Jiliang Tang. 2019. Deep adversarial network alignment. *arXiv preprint arXiv:1902.10307* (2019).

[3] Maeve Duggan and Joanna Brenner. 2013. *The demographics of social media users, 2012*. Vol. 14. Pew Research Center's Internet & American Life Project Washington, DC.

[4] Kenneth Frank, Yun-jia Lo, Kaitlin Torphy, and Jihyun Kim. 2018. Social Networks and Educational Opportunity. In *Handbook of the Sociology of Education in the 21st Century*. Springer, 297–316.

[5] Christine Greenhow, Vincent Cho, Vanessa P. Dennen, and Barry J. Fishman. 2019. *Education and Social Media: Research Directions to Guide a Growing Field*. Vol. 121. Teachers College Record.

[6] Hamid Karimi, Tyler Derr, Aaron Brookhouse, and Jiliang Tang. 2019a. Multi-factor congressional vote prediction. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 266–273.

[7] Hamid Karimi, Tyler Derr, Kaitlin Torphy, Kenneth Frank, and Jiliang Tang. 2019b. A Roadmap for Incorporating Online Social Media in Educational Research. *Teachers College Record Year Book* 2019 (2019).

[8] Hamid Karimi, Tyler Derr, Kaitlin T Torphy, Kenneth A Frank, and Jiliang Tang. 2020. Towards Improving Sample Representativeness of Teachers on Online Social Media: A Case Study on Pinterest. In *proceedings of 21th International Conference on Artificial Intelligence in Education*.

[9] Hamid Karimi, Courtland VanDam, Liyang Ye, and Jiliang Tang. 2018. End-to-end compromised account detection. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 314–321.

[10] V Darleen Opfer, Julia H Kaufman, and Lindsey E Thompson. 2016. Implementation of K–12 state standards for mathematics and English language arts and literacy. (2016).

[11] Kaitlin Torphy, Hu Sihua, Yuqing Liu, and Zixi Chen. 2018. *Examining the Virtual Diffusion of Educational Resources Across Teachers Social Networks Over Time*. Vol. Special issue. Teachers College Record.

[12] Kaitlin T. Torphy and Corey Drake. 2019. *Educators Meet the Fifth Estate: The Role of Social Media in Teacher Training*. Vol. 121. Teachers College Record.