

# Modeling Innovation Diffusion in an Online Tutoring Network

Anderson, Robin  
robina@stanford.edu

Lang, David  
dnlang86@stanford.edu

Lee, Hana  
leehana@stanford.edu

December 11, 2017

## 1 Introduction

Professional development for schools and institutions is often ineffective, and costly (Gates Foundation, 2014). When teachers are permitted to choose their own professional development, their satisfaction and adoption of new pedagogies increases. Due to the nature of personal preferences and a focus on individualized professional development to increase teacher learning, multiple pedagogies will likely be adopted into these teacher networks. A new pedagogical practice may enter a network due to the preference of only a small number of teachers, yet spread quickly due to network structures. Providers of professional development, often school districts or other governing agencies, capitalize on the influential ability of a few teachers. Approximately half of research on professional development study programs that utilize a cascade model of teacher training (Popova et. al. 2016). Some of these programs employ a “train the trainer” approach, where a teacher leader is guided through the professional development as a learner by an expert, then trained on how to guide other teachers through the same experience (Borko et al. 2017). In other programs, teachers are selected by the administration to attend conferences or workshops and are asked to report back on their experience (Day, 1999). While research has found that cascade models promote the transfer of skills rather than values (Solomon & Tresman, 1999), they are often still the chosen form of professional development due to time and monetary constraints.

To the extent that practice can be changed by treating only a few teachers, through a cascade model of professional development, it may make more sense to focus on identifying the most central figures in a professional social network, rather than expose all teachers to a required professional development. Alternatively, if practices can be transmitted across different vectors (e.g. via assigning students to teachers), then this too could inform the design of educational intervention at scale.

In this project, we explore how adoption of a digital whiteboard spreads across two distinct types of network: a chatroom graph and a tutoring session

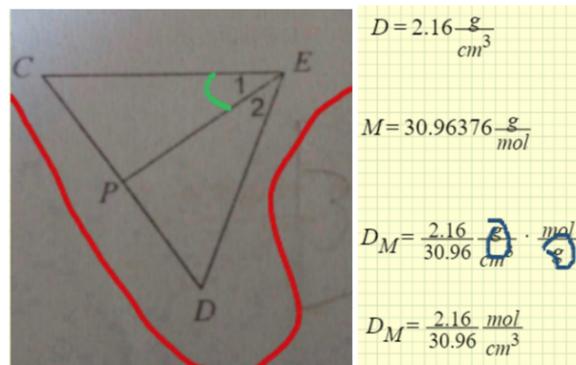


Figure 1: Examples of the whiteboard tool.

graph. The tool facilitates a tutor’s ability to draw on a student’s work, typeset equations, or overlay a grid such that graphing examples and other formulae can be written in a more accessible fashion (Figure 1).

This project has two goals:

1. Investigate the adoption of a whiteboard technology among tutors in an online tutoring application using a threshold model.
  - RQ 1: Does peer adoption predict tutor adoption of the whiteboard?
2. Investigate cascade models of professional development through the case study of the adoption of whiteboard technology within a network of tutors.
  - RQ 2: Who should attend professional development within a cascade model?
  - RQ 3: When should a cascade model of professional development be used?

## 2 Related Work

### 2.1 Diffusion of innovation

Diffusion of innovations theory primarily posits that new ideas and practices spread through social

contact. Thomas Valente (2005) describes several different mathematical and network models used to study the diffusion of innovations through communities of individuals. As an example of a network model, Valente uses the results of a previous study conducted in Cameroon that attempted to determine if friendship ties were associated with contraceptive choice. The study compared diffusion in the real-life social network with diffusion in a randomly generated network of the same size and density, but with randomly allocated links, and found that social ties did influence the rate of diffusion through the network.

Overall, Valente concludes that the results of the various studies summarized in his paper indicate that social network influences on behavior are an important way that innovation spreads through networks. However, he states that this conclusion is primarily supported by qualitative evidence rather than a quantitative mathematical model. Most people acknowledge that they are influenced in their decisions by those in their social network, and the rate of acceleration of diffusion in a real social structure is empirically higher than in a network with random links. However, precisely modeling the effect of social ties on adoption is difficult because of the many variables and nonlinearities present.

## 2.2 Diffusion through heterogeneous social networks

The study of diffusion of innovation has attempted to model both biological contagion and social influence. Delre, Jager, and Janssen (2007) argue that previous attempts to model social contagions, like product adoption by consumers, fall short due to the assumption of homogeneity in the sample population. They argue that a model of social diffusion of innovation must take into account both network structure and personal preferences. Through their discussion, the authors highlight that a model must both act like an epidemic model, which draws on the characteristics of space within the network, and also a threshold model, which models adoption based on influence of others behaviors close to the adopter in the network.

Through multiple simulations of modeling different types of networks, Delre, Jager, and Janssen show that this model does extend the understanding the epidemic framework to social contexts with the adherence to two social concepts: 1) social influence in personal networks, and 2) heterogeneity in decision-making. These simulation of the models found that the spread of an idea through social networks is faster through highly clustered networks, which supports the analysis of graphs done by Cen-

tola (2010), which found that adoption of new technology spread faster through clustered-lattice graphs (neighbors are neighbors) versus random graphs.

## 3 Dataset

We used chatroom transcript and session log data from an educational technology company called Yup. Yup provides tutoring to students via text message and multimedia message. When a student requests a session, they are randomly assigned to one of a set of available tutors.

### 3.1 Tutoring sessions

Our first data set consists of interactions between tutors in chat rooms. These chat rooms consist of messages between tutors, Yup employees, and corporate officers. The nature of these chatrooms focus on a variety of topics including pedagogy, shift coordination, technical support, reporting student behavior, and recreational content. The largest of these chatrooms has more than 70 participants and has more than 100,000 unique messages. Each message is associated with a timestamp, its contents, and the user who sent that particular message. In this model, each participant is a node in the network. We created edges between participants in a chat room if they sent messages within a 30 minute interval of each other. Because length of attendance in a chat room is not available per individual, we used temporal proximity within the chatrooms to account for an interaction. We chose a 30 minute interval because that is the minimum shift length of a tutor and tutors most often participate in the chat rooms during their shifts.

### 3.2 Chatrooms

Our second data set consists of tutoring session data. This data includes the associated timestamp of the session and which tutor and student engaged in the session. There are also several aggregate measures of session quality including its length, student satisfaction, whether or not the student purchased services, and whether or not the tutor used the piece of educational technology in question (the digital whiteboard). In this model, we treated the network as a bipartite graph of tutors and student interactions. Each node corresponds to either a student or a tutor and each edge constitutes a tutoring session between the two. In total, there were 8,954 sessions across 62 tutors and approximately 1800 unique students in these sessions. We collapsed this data into a tutor-tutor network where an edge is created between two tutors if they tutored the same student.

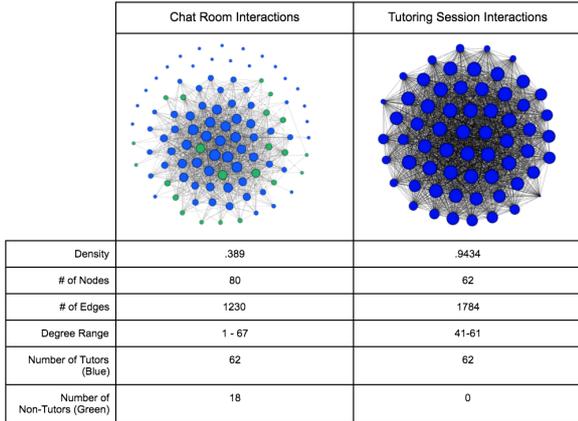


Figure 2: Visualization and features of the two networks.

### 3.3 Feature analysis

We hoped to gain more insight from analyzing both of these networks than we could from analyzing either network alone. The chatroom network encodes direct relationships between tutors, with edges formed between tutors who have actually interacted with each other. The tutoring session network encodes indirect relationships formed between tutors who have tutored the same students. The relationship of direct interaction from the chatroom network has a more common analogue in most professional development settings, where teachers are more likely to interact with each other than they are to share the same students. The networks differ significantly from each other in terms of features like density and degree range (Figure 2). The tutor network is a near complete, connected graph, whereas the chatroom graph has more variance in the degree distribution. Additionally, the chatroom graph contains isolated nodes.

Using the tutoring session data, we examined whether or not each tutor adopted the whiteboard in their tutoring sessions. We define adoption as a tutor using the whiteboard at least once during the period of data collection (September 1st, 2016 to December 10th, 2016). At the start of the data collection period, 21 of the 62 tutors had already used the whiteboard and counted as infected. By the end of the data collection period, 56 of the 62 tutors adopted the technology. There are no new adopters after the tenth week (Figure 3).

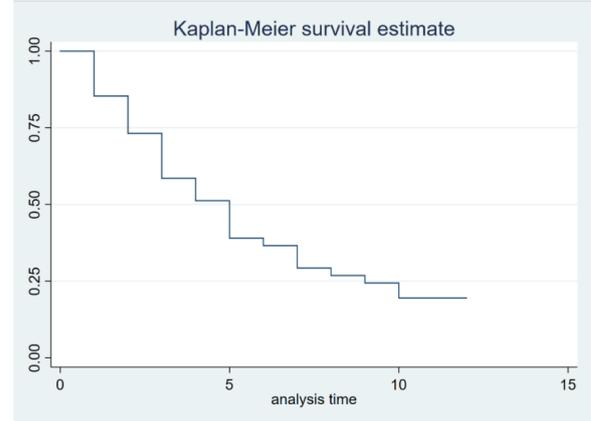


Figure 3: Proportion of non-adopters week-to-week.

## 4 Analysis

### 4.1 Threshold model with unweighted edges

We first considered a peer adoption model where tutors try out the whiteboard in the subsequent week if a certain percentage of their colleagues utilize the tool. Over the 14 weeks of data collection, we adopted the rule that tutors are always whiteboard users if they use the whiteboard in at least one tutoring session in a prior week. We then assessed various potential tipping point thresholds at which an uninfected user  $u$  will adopt the whiteboard if at least  $x\%$  of  $u$ 's peers adopt the technology. In one of the prior problem sets for CS224W, we evaluated a variant of these tipping point models where individuals voted for a candidate if a plurality of their neighbors ( $x\%$ ) preferred a candidate.

In the tutor-student network, we estimated this model using each week of data. Figure 4 displays the associated receiver-operator characteristic (ROC) curve, a curve describing the relative trade-off between rules that accurately identify whiteboard users and accurately identify non-whiteboard users. The specific curve corresponds to predicting whiteboard adoption in a tutor's next week of sessions based on his peers' adoption of whiteboard in the current week. If we wish to maximize the accuracy of this rule, then the associated tipping point from our training set is approximately 50% of peers adopting the technology. This model accurately classifies individuals in each week 75.6% of the time. The associated model fit statistics yields an AUC of 0.60. As a baseline comparator, guessing that all users adopt the whiteboard in each period results in an accuracy rate of 73.6%<sup>1</sup>.

<sup>1</sup>For these we also utilized a train and test sample corresponding to a roughly 60% training dataset and a 40% test

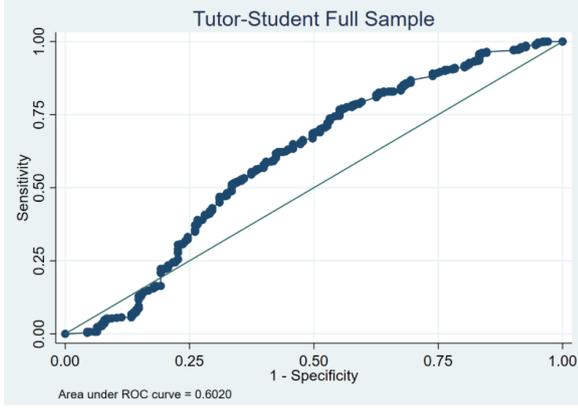


Figure 4: ROC curve for tutoring session network with unweighted edges.

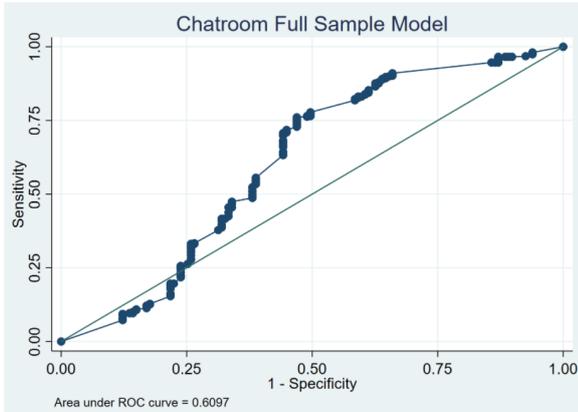


Figure 5: ROC curve for chatroom network with unweighted edges.

We then performed the same procedure on the chatroom network, where tutors who communicated within a 30 minute time period of each other share an edge. Of those tutors who adopt, we assumed the same threshold model; if  $x\%$  of a tutor’s neighbors have adopted the tool, then the tutor will adopt the tool in the subsequent period. In our full sample model, the most accurate tipping point is approximately 52%, and the associated accuracy rate is 77.9%. The AUC measure here is slightly better at 0.61, but this value is still generally considered a poor model fit for predictive purposes. As a baseline comparator, guessing that each user would adopt the tool in the subsequent period results in a 77.3% model accuracy.

In general, the model-fit statistics perform relatively poorly based on the AUC scores. These statistics suggest that the models for both networks fit their network equally well but, on the whole, do a

dataset. Parameter estimates and model accuracy did not vary substantially and are omitted for brevity.

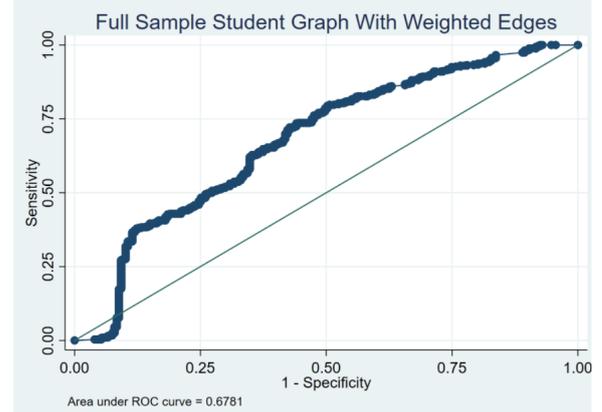


Figure 6: ROC curve for tutoring session network with weighted edges.

poor job. The model for the first network is only 2% more accurate than simply guessing that all users adopt the whiteboard, and the model for the second network has an even smaller gain in accuracy.

## 4.2 Threshold model with weighted edges

After analyzing the accuracy of a simple threshold model on the unweighted networks, we experimented with weighting edges in the two graphs. In a practical sense, we know that not all connections between tutors are equal. In the chatroom network, some tutors may engage in frequent conversations with one another, whereas others might interact only a handful of times. In analogous fashion, some pairs of tutors have greater overlap in terms of students in common than other pairs.

For the chatroom network, we weighted edges according to a simple count of how many times that particular tutor-tutor pair interacted with each other in the chatroom. For the tutoring session network, we weighted edges according to how many unique students were taught by both tutors at least once. We then re-analyzed the accuracy of the simple threshold model on the weighted network, considering each weighted edge between two tutors as a collection of  $x$  edges between the two tutors, where  $x$  is the weight of the edge.

For our tutor-student network, we find that a weighted model of edges greatly improves model accuracy. While the overall accuracy rate doesn’t experience much of a change (74.3% in the weighted model), the area-under-the-curve (AUC) statistic increases to 0.68. This shift suggest that incorporating edge weights can improve the prediction of whiteboard adoption.

In contrast, our weighting schema for the tutor-

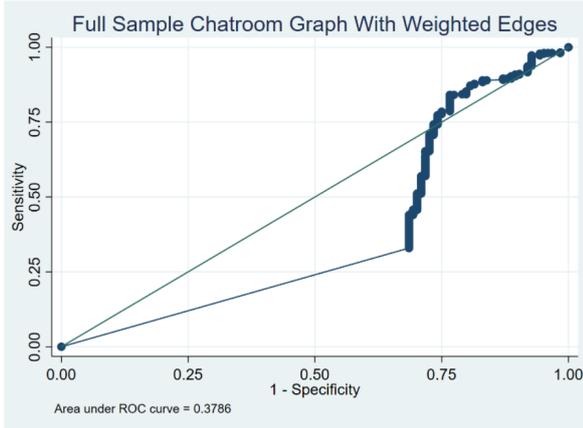


Figure 7: ROC curve for chatroom network with weighted edges.

chatroom graph proved less valuable. In particular, the AUC for this curve was .38, suggesting that this model would actually perform worse than random chance at many of the possible values for a peer adoption threshold.

Tutor-student graphs appear to be more predictive when rated, suggesting that repeated interactions with the same student may influence a tutor’s practices accordingly. In contrast, the amount of communication between tutors doesn’t prove very predictive. This makes sense when one considers the fact that we did not filter the chatroom edge lists based on whether or not the messages contained pedagogical content, nor could we completely identify whether or not an individual was closely following a conversation that was occurring, only whether or not the individual participated in it at some point. For these reasons, it makes sense that the chatroom network may be less predictive of whiteboard adoption.

### 4.3 Influence maximization

While the simple threshold model produced relatively poor results in terms of how network structure predicts the diffusion of the whiteboard innovation across the networks, we will nonetheless move on to addressing our second research question: who should attend professional development within a cascade model? If we assume that educational interventions do work and can diffuse through a network, then we can turn to influence maximization as a way to answer this question.

Before beginning influence maximization, we make another assumption: nodes do not all have equal influence in the graph even if they have identical degree. They should be re-weighted based on the number of sessions taught by each tutor. The rationale behind reweighting nodes is that through profes-

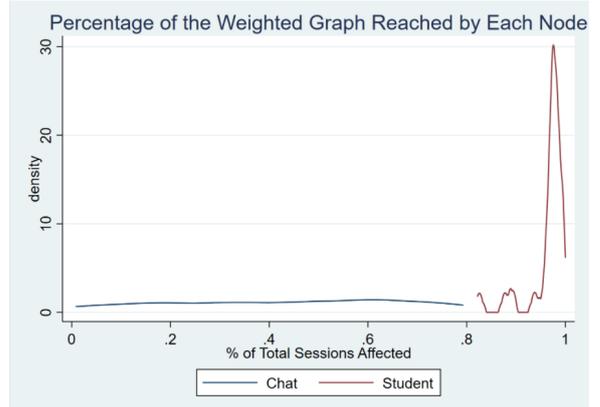


Figure 8: Distribution of influence for each node in both networks.

sional development, we are not only trying to reach as many tutors as possible, but also attempting to maximize the number of students affected by new pedagogical innovations. By reweighting nodes according to the number of sessions taught (i.e. the number of students reached), we can choose the “maximally influential” nodes based not only on the raw number of other tutor nodes they are connected to, but how many “high activity” (many sessions) nodes they are connected to.

In the tutor-student graph, we find that the choice of tutor to influence is relatively unimportant both in terms of degree as well as the number of sessions. Choosing the highest influence node results in reaching 100% of the graph with a single person influencing their immediate neighbors. Choosing the least influential node still results in reaching 90% of the graph. In short, choosing any node at random would likely suffice if we believed a practice/behavior would diffuse through this graph.

In contrast, the chatroom graph has a lower range and much more variability in its ability to influence the graph. Of particular note, there are tutors who are in the tutoring session graph but are inaccessible/did not participate in the chatroom graph. Choosing the maximal influence node in the chatroom network would reach about 80% of the weighted tutors. In contrast, the minimally influential node would influence 0% percent of the graph because the individual is disconnected from this network.

Since both of these networks represent the underlying structure of relationships between tutors, the tutoring session graph appears to yield much more information about these connections. Therefore when we are calculating influence sets in order to maximize our influence over the graph by choosing only a few nodes, we should either use a combined

network of edges from both of the networks, or only take into account the tutoring session network. The chatroom network alone doesn't capture enough of the tutor-tutor connections to be useful for influence maximization.

The third research question we attempt to address is whether or not professional development models should use a cascade model. Most teacher professional development models are based on the assumption of a single-cascade (e.g. take one teacher at a high school and have them teach their peers). Normally, there are reasons to think that the second generation of instruction is less effective than the first: the teacher who received instruction is not as much of a domain expert as the original content provider; the second form of professional development is often abbreviated; and some information is lost in transmission.

Moreover, we have reason to believe that certain skills such as lesson planning, classroom management, and pedagogical content may be more or less transmissible than other skills. For instance, a language teacher would receive relatively little benefit from conversing with a math teacher who has learned a new way of explaining fractions to their students. Based on this assumption, policy makers will often have a choice of engaging in professional development that either has the potential for spillover effects or not. If peers do not benefit from another peer's training (i.e. the spillover effect is low), then the quality of the initial treatment effects become much more relevant for policy decisions. Alternatively, policymakers often have the choice of focusing development on teachers who are relatively isolated from their peers (e.g. subject-specific teachers at small schools) or individuals who are highly connected but not necessarily involved in a classroom (e.g. math coaches, assistant principals, or curriculum developers).

As such, we construct a cost-benefit ratio where we compare the benefit of directly training a tutor in a skill that has no spillover effect (the training only benefits students that interact directly with the tutor) to the benefit of training that has only spillover effects (the original tutor doesn't benefit from the training, but his peers and his peers' students benefit from the additional content knowledge.) We perform this comparison by computing breakeven rates. We assume an initial treatment effect  $\alpha$  that degrades with each generation of transmission. For example, if  $\alpha = 0.5$ , a teacher going to professional development conveys 50% of the information learned from attending the original training to his peers, those peers convey less than 50% of the information to their peers, and so on. We explicitly define this spillover effect/second generation benefit in reference

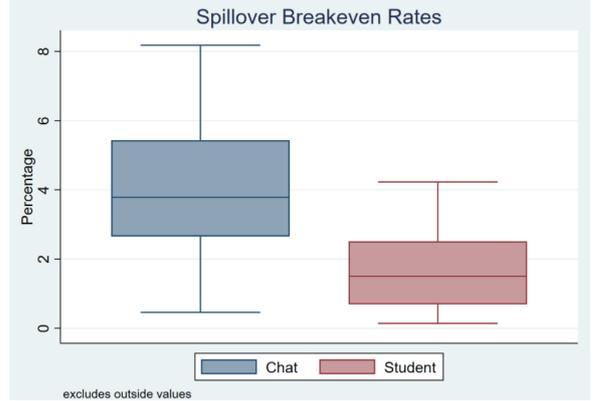


Figure 9: Distribution of spillover breakeven rates for each node in both networks.

to a breakeven rate defined below:

$$Breakeven_i = \frac{Sessions_i}{PeerSessions_i}$$

If the spillover effect is greater than this breakeven rate, a policy maker would prefer training tutor  $i$  as a math coach because the benefits of them teaching their fellow teachers improves more sessions by a higher amount than if tutor  $i$  received the benefits of the original training but remained completely isolated from their peers.

If the spillover effect is less than the breakeven rate, a policy maker would prefer training an isolated tutor  $i$  and keeping them as a tutor, rather than a math coach, because the overall improvement in session quality for that tutor's sessions would be greater than the benefit of making tutor  $i$  a math coach for other tutors.

We computed this breakeven ratio in both graphs for each node, using tutor  $i$ 's session count as well as the sum of  $i$ 's neighbors' session count. In Figure 9, we plot breakeven ratios for the student and chatroom networks. We find that in both graphs, the spillover effects dominate even when they are relatively small. In other words, even if most of the original benefit of professional development is lost when teaching one's neighboring peers, if your social network is sufficiently dense, the spillover effects can still be larger than the original treatment effect.

In the student graph, the median tutor needs a spillover effect of approximately 2% for their spillover effect to be larger than the original treatment's benefit. In the chatroom graph, the median spillover breakeven rate is around 4%.

This information suggests that cascade models could potentially work and be effective at influencing a large number of teachers, even if it loses efficacy rapidly. This finding also suggests that that special effort should be given to identifying teachers that are

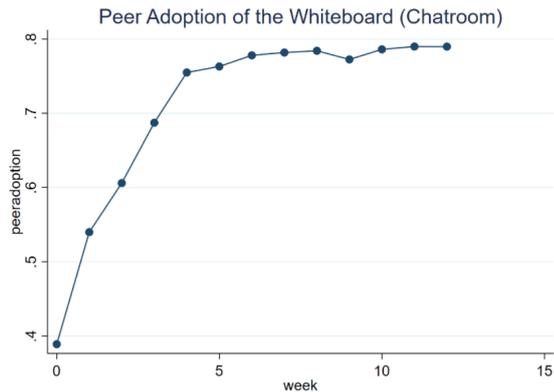


Figure 10: Mean percentage of peer adopters for each node in the chatroom network.

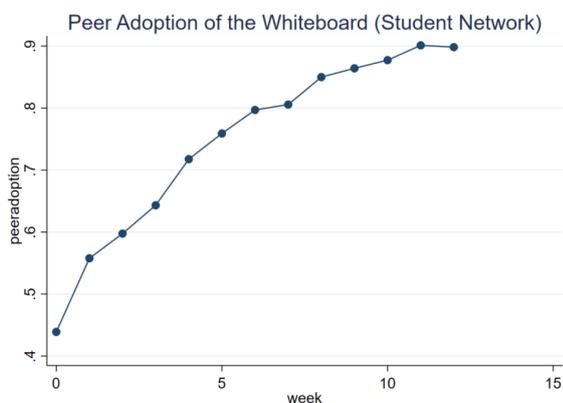


Figure 11: Mean percentage of peer adopters for each node in the tutoring session network.

isolated within a graph, since they will not benefit from the spillover effect.

## 5 Limitations

We theorize that there are a few different reasons why our threshold model has relatively low predictive validity. For one, there is relatively little variation in terms of peer adoption as the network evolves. In the chatroom network, the network reaches its saturation level in week 5 (Figure 10) and doesn't increase very substantially from that point on. In the tutoring session network (Figure 11), we see more variability from week to week, but for both networks, the fact that we do not have data from the time the whiteboard was introduced makes analysis difficult.

Another possible reason for poor model fit is that our definition of adoption is rather obtuse. Because we are focused on exposure to the whiteboard technology (defined as trying it at least once) rather than retention, we define adoption as a person using the

tool at least once in their lifetime. An alternative definition of adoption, such as the relative frequency of whiteboard utilization, may be better suited for modeling.

For future work, instead of a simple threshold model, we may be better off modeling adoption as a repeated exposure of users to a particular behavior rather than seeing it occur in peers only once. It may produce more significant results if we model utilization as a function of a tutor's past usage of the tool and their peer's utilization of the tool:

$$Utilization_{i,t+1} = f(Utilization_{i,t}, PeerAdoption_{i,t})$$

## 6 Contributions

Robin: Background and relevant literature review, feature analysis, network visualization in Gephi, poster creation

David: Procured dataset, parsed and built tutor network data, constructed other non-network figures, and evaluated influence maximization, and breakeven analysis

Hana: Building networks from data, designing threshold model, proofreading and typesetting report

## 7 References

- [1] Bill & Melinda Gates Foundation (2014). Teachers know best: Teachers' views on professional development. Seattle, WA: Bill & Melinda Gates Foundation.
- [2] Centola, D. (2010) The Spread of Behavior in an Online Social Network Experiment. Science. Delre, S. A., Jager, W., & Janssen, M. A. (2007). Diffusion dynamics in small-world networks with heterogeneous consumers. Computational and Mathematical Organization Theory, 13(2), 185–202.
- [3] Day, C. (1999). Developing teachers: The challenges of lifelong learning. London: Falmer Press.
- [4] Delre, S. A., Jager, W., & Janssen, M. A. (2007). Diffusion dynamics in small-world networks with heterogeneous consumers. Computational and Mathematical Organization Theory, 13(2), 185–202.
- [5] Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 78(6), 1360-1380.

- [6] Popova, A.Evans, D. Arancibia, V. (2016). Training teachers on the job : what works and how to measure it. Policy Research working paper; no. WPS 7834; WDR 2018 background paper. Washington, D.C. : World Bank Group.
- [7] Solomon, J. & Tresman, S. (1999). A model for continued professional development: Knowledge, belief and action, *Journal of In-service Education*, 25, pp. 307-319.
- [8] Valente, T. (2005). Network models and methods for studying the diffusion of innovations. *Models and Methods in Social Network Analysis*. 98-116. 10.1017/ CBO9780511811395.006.
- [9] Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315.