Supporting the Encouragement of Forum Participation

Aashna Garg
Stanford University
aashna94@stanford.edu

Andreas Paepcke
Stanford University
paepcke@cs.stanford.edu

ABSTRACT
Abstract goes here

1. INTRODUCTION

2. INTRODUCTION

3. DATASET

Since 2011 many courses in the university have used the Piazza forum facility for its internal courses. Our dataset comprises the posts and comments of these usage years. From among 5000 university course offerings that have used the Piazza forum facility since 2011 we selected **** courses, with a total of **** offerings. Most courses were taught multiple times since 2011, so we include longitudinal forum usage data for most courses. One course was taught as a MOOC, and used Piazza; we included that data for comparison. Most MOOCs make use of the delivery platform's built-in forum facility.

We used two criteria for selecting the courses to analyze. We favored those that had comparatively large numbers of student posts, and we tried to cover courses from many schools and departments. We in particular sought to include Humanities courses that made use of Piazza for class discussion. Table ?? summarizes our choice.

From the forum posts of these data we constructed one social graph for each course offering.

4. FROM POSTS TO CONNECTION GRAPH

Social networks are most simply modeled by considering each participant as a node, and interactions initiated by participants as out-directed links. In this case all nodes are of one type, and links are unidirectional. Multiple interaction initiations by one person are captured by weighting the corresponding outgoing links. Many graph analysis tools operate on models of this type, and this is the approach we chose.

However, other strategies exist to cover different goals. For example, [1] additionally consider linkages between forum post topics to include communication content in the model. When networks operate on particular platforms, such as underground forums, which include private 'buddy' connections, such facilities may need to be modeled [4].

For the purpose of identifying candidate time points for encouraging online conversation participation our chosen model

Course	Offering	Number of students	Student with pos
Artificial Intelligence	FALL12	192	106
	FALL13	278	195
	FALL14	435	291
	FALL15	494	357
	FALL16	782	529
	SUMMER13	137	81
Machine Learning	FALL11	356	166
	FALL12	581	283
	FALL13	776	383
	FALL14	820	449
	FALL15	827	489
	FALL16	803	393
	SPRING16	331	189
Computer Vision	FALL11	59	29
	FALL12	120	62
	SPRING16	208	150
	WINTER14	152	101
	WINTER15	180	125
Decision Analysis	FALL15	176	100
	FALL16	156	104
Computational Molecular Biology	FALL11	96	61
	FALL12	101	67
	FALL13	123	80
	FALL14	140	103
	FALL15	147	93
	FALL16	120	86

Table 1: Summary of Courses Used for Analysis

suffices. We are not in this work considering additional measures, such as content quality, for which a richer model would be required.

Many measures are used to quantify various aspects of social graphs [2, 3]. Not all are meaningful in the context of education-related forum interactions. We focus here on two measures: *out-degree*, and *page rank*. Figure 1 illustrates.

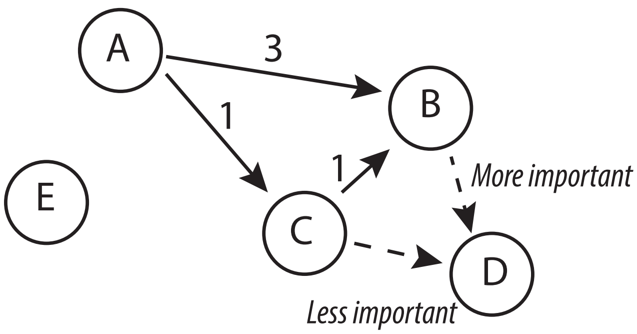

Figure 1: Example social graph induced by forum posts.

Nodes A, B, C, D, and E represent students. The link from A to B is marked with the number 3, because A commented three times on one of more of B's posts. The number of outgoing links is the node's *out-degree*. For example, the *out-degree* of A is 4.

The number of incoming links is called the node's *in-degree*. Node C's in-degree is 1. Node E has no links entering or exiting. The respective student has not participated in the forum.

Analogous to Web pages, each node can be assigned a *page rank*. The intuition in this context is that student S_1's presence in the forum is more 'important' than student S_2's if the node representing S_1 has higher page rank than the node that represents S_2. In our context the intuition behind page rank is that a node N is more important (has higher page rank) the more other important nodes comment on N's posts. Imagine a scenario in which student S_1 posts an interesting question, to which many students comment with their opinion, creating a long thread. The node representing S_1 would experience an increase of its page rank with every incoming comment. Node B in Figure **??** is an example for this situation. Its in-degree is 4. If B were to comment on one of D's posts, then D's page rank would increase more than if the low-page-rank node C commented on D.

In terms of evaluating a student's participation in the forum, a high page rank, and high out-degree are positive. Low values are less positive. We chose these two values because of their relatively straight-forward meanings when applied to forum posts, and for their relevance to our goal of identifying potential intervention times.

Some of the fifteen other measures we computed, such as *betweenness* are meaningful for forum scenarios as well, but their usefulness depends on one's analysis goals. For example, [5] include several of those measures for the purpose of prediction analysis. For evaluation contribution quality the contents of posts would need to be considered: students who persistently post irrelevant contents contribute less positively to the forum than constructively participating students. However, for our purposes the two measures of page rank and out-degree provide strong enough signals.

Course	Offering	Nodes	Edges	Weighted OutDeg	Density
Artificial Intelligence	FALL12	192	414	893	0.011289267015
	FALL13	278	1396	4629	0.018128457522
	FALL14	435	1423	4352	0.007537475501
	FALL15	494	2114	4220	0.008680227640
	FALL16	782	3699	7286	0.006056567257
	SUMMER13	137	320	917	0.017174753112
Machine Learning	FALL11	356	601	1237	0.004755499287
	FALL12	581	1306	2663	0.003875600925
	FALL13	776	1823	4649	0.003031260392
	FALL14	820	1943	3610	0.002893177283
	FALL15	827	2413	5553	0.003532415364
	FALL16	803	1510	3140	0.002344698651
	SPRING16	331	657	1548	0.006014831090
Computer Vision	FALL11	59	84	248	0.024547048509
	FALL12	120	195	528	0.013655462184
	SPRING16	208	1018	3722	0.023643626904
	WINTER14	152	601	1618	0.026185081910
	WINTER15	180	702	1639	0.021787709497
Decision Analysis	FALL15	176	265	445	0.008603896103
	FALL16	156	373	1185	0.015425971877
Computational Molecular Biology	FALL11	96	348	653	0.038157894736
	FALL12	101	289	657	0.028613861386
	FALL13	123	490	1279	0.032653605224
	FALL14	140	527	1122	0.027081192189
	FALL15	147	396	797	0.018451216102
	FALL16	120	384	886	0.026890756302

Table 2: Summary of Examined Courses

5. ANALYSIS PROCEDURE

For each course we identified the top 10% of forum-contributing students as measured by their postings across the entire course. For those students we computed the average of our chosen social graph measures. This average was computed for each week of each offering. The results provided a week-by-week time series of postings and evolving pagerank for these students.

Figures 2 and 3 shows a typical time series with weighted out degree on the ordinal, and weeks on the abscissa. The line at the bottom shows the median weighted out degree across time for this Fall 2016 offering of a machine learning class.

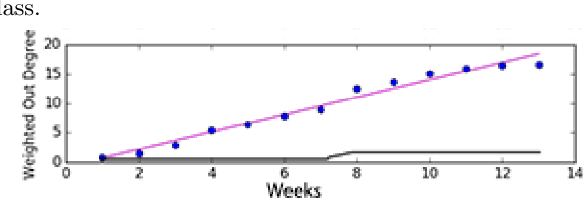

Figure 2: Forum post contributions by top 10% of contributors (machine Learning class).

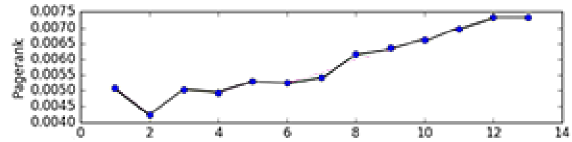

Figure 3: Page rank by top 10% of contributors (machine Learning class).

6. CONCLUSION AND FUTURE WORK

Here is an example chart of the correct size.

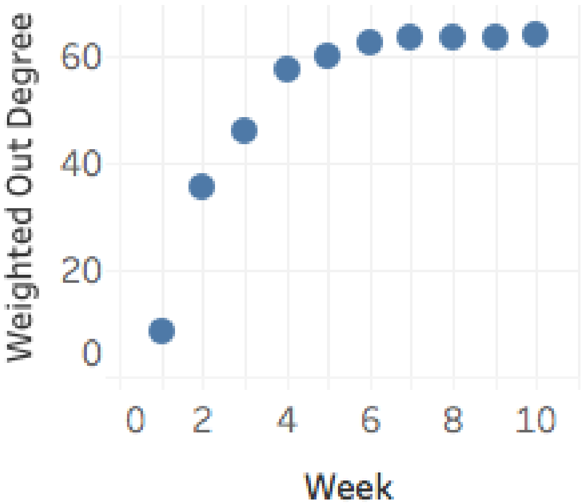

Figure 4: A MOOC forum post comparison: Women's Global Health.

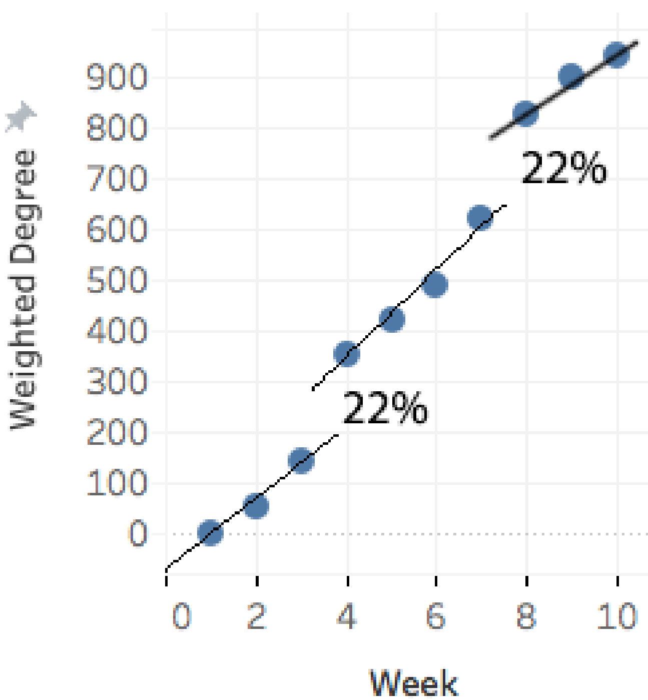

Figure 5: Wrong numbers, just an example. See /tmp/forumPromptsTableauChartsSample.twb

- Consider student post content quality
- Consider consistency: contribute throughout course
- Consider influence on others
- Draw instructor attention to dense topic clusters, which might indicate confusion, or student excitement to harness.

<

7. REFERENCES

[1] T. Anwar and M. Abulaish. Modeling a web forum ecosystem into an enriched social graph. In M. Atzmueller, A. Chin, D. Helic, and A. Hotho, editors, *Ubiquitous Social Media Analysis: Third International Workshops, MUSE 2012, Bristol, UK, September 24, 2012, and MSM 2012, Milwaukee, WI, USA, June 25, 2012, Revised Selected Papers*, pages 152–172. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[2] R. A. Hanneman and M. Riddle. *Introduction to social network methods*. University of California, Riverside, 2005.

[3] J. Leskovec, A. Rajaraman, and J. D. Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2014. Chapter 10.

[4] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker. An analysis of underground forums. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 71–80, New York, NY, USA, 2011. ACM.

[5] D. Yang, T. Sinha, D. Adamson, and C. P. Rosé. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, volume 11, page 14, 2013.