Coenrollment Networks and their Relationship to Grades in Undergraduate Education

Josh Gardner
School of Information
The University of Michigan
jpgard@umich.edu

Christopher Brooks
School of Information
The University of Michigan
brooksch@umich.edu

ABSTRACT

In this paper, we evaluate the complete undergraduate coenrollment network over a decade of education at a large American public university. We provide descriptive properties of the network, demonstrating that the coenrollment networks evaluated follow power-law degree distributions similar to many other large-scale networks; that they reveal strong performance-based assortativity; and that network-based features can significantly improve GPA-based student performance predictors. We then implement a network-based, multi-view classification model to predict students' final course grades. In particular, we adapt a structural modeling approach from [19, 34], whereby we model the university-wide undergraduate coenrollment network as an undirected graph. We compare the performance of our predictor to traditional methods used for grade prediction in undergraduate university courses, and demonstrate that a multi-view ensembling approach outperforms both prior "flat" and network-based models for grade prediction across several classification metrics. These findings demonstrate the usefulness of combining diverse approaches in models of student success, and demonstrate specific network-based modeling strategies which are likely to be most effective for grade prediction.

ACM Reference format:
Josh Gardner and Christopher Brooks. 2018. Coenrollment Networks and their Relationship to Grades in Undergraduate Education. In *Proceedings of International Conference on Learning Analytics and Knowledge, Sydney, NSW, Australia, March 7–9, 2018 (LAK '18),* 10 pages.
https://doi.org/10.1145/3170358.3170373

1 INTRODUCTION

Coenrollment networks, or networks of students enrolled in the same courses at an institution, represent a powerful source of information about student performance and about broader patterns of student engagement in higher education. In this work, we conduct a large-scale analysis of coenrollment networks to (a) analyze the properties of these networks at a scale not previously examined, and (b) explore the effectiveness of grade prediction techniques which utilize coenrollment networks. The motivation for this approach is twofold: First, prior research suggests that there are aspects of university social networks which may be relevant to student performance (i.e., [18]), but such research has not examined coenrollment networks at scale, nor has it used these networks for predictive modeling. Our exploratory analysis reveals important properties of coenrollment networks in a large American university, and confirms that network properties are indeed related to student performance in this data at a scale not previously evaluated.

Second, we hypothesize that graph-based prediction models should perform better than "flat" models by capturing relationships between observations in data which are otherwise assumed to be independent. We find that structural models which use student-level coenrollment networks outperform traditional "flat" models, and that network-based features only provide performance improvements when used as a part of such structural models (they do not improve traditional flat models by simply being added to the input feature set). Furthermore, we demonstrate that by combining network-based and flat models using a multi-view blended ensemble, we achieve predictive performance superior to prior link-based classification work.

These results demonstrate the effectiveness of network-based models in learning analytics, and the importance of coenrollment networks to student success prediction in higher education. Such predictive models have potential applications in dropout prediction and "early warning" models, for broader instructor- and student-facing support systems, and course selection, as well as other applications which require accurate predictions of student performance before academic data from a course is available.

1.1 Network Notation and Classification Task

In this section we briefly introduce the notation used to describe the coenrollment network and the classification task used in this work. We present the coenrollment network as a graph $\mathbb{G} = (\mathbb{O}, \mathbb{L})$ where $\mathbb{O}$ is the set of n objects (students), and $\mathbb{L}$ is the set of undirected links (coenrollments). For each node $o_i \in \mathbb{O}$, we know a set of *object attributes* $OA(o_i)$ which here include student- and course-specific attributes such as gender, ethnicity, cumulative GPA, subject, and major (these are the attributes normally used for tasks such as grade prediction). In addition, for each of these nodes, we derive a set of *link attributes*, $LA(o_i)$. Link attributes are aggregations of class labels across o_i's neighbors (our process for constructing $\mathbb{G}$ and extracting $LA(o_i)$ is detailed in Section 3.2).

In general, we use blackboard bold letters to refer to vectors or collections of items (i.e., $\mathbb{O}$ for the set of objects, or nodes) and indexed, lowercase letters to refer to individual elements of those vectors (o_i for an individual node).

Our experiment in Section 5 addresses the following task:

LAK '18, March 7–9, 2018, Sydney, NSW, Australia

ACM ISBN 978-1-4503-6400-3/18/03…$15.00
https://doi.org/10.1145/3170358.3170373

The link-based classification task (LBCT): Learn a model from a student graph, $\mathbb{G}$, such that the accuracy of the class predictions $\hat{C}^{(*)}$ on a disjoint future graph $\mathbb{G}^{(*)}$ is maximized.

Note that in this task, the disjoint graph $\mathbb{G}^{(*)}$ is a future academic semester we wish to predict grades in.

2 PREVIOUS WORK

2.1 Social Learning and Graph-Based Modeling in Higher Education

Social learning theory provides a theoretical foundation for the impact of social networks on learning. Social interaction among peers has been recognized as a core component of the learning process for several decades [18] and lies at the core of many modern pedagogical approaches, including social constructivism [1] and cooperative learning [28]. A variety of empirical and experimental research has demonstrated positive relationships between social interactions in courses and learning outcomes [18], including elevated cognition [49] and self-regulation [22].

Network analysis has seen limited application in higher education research, where the influence of other students on a given students' performance is known as a "peer effect", but analysis of coenrollment networks has been limited in both quantity and scale. In their thorough overview, Biancani and McFarland identify at least 56 social network-based studies which use individual university students as the unit of analysis [8], noting that this research has been primarily descriptive and explanatory, not predictive. Prior work has explored other networks in higher education, including dorm roommate networks [3], friendship networks [59], and demographic networks based on geographic background [32], and networks based on institutional factors such as major and class year [56]. Many of these analyses use external data from social networking sites such as Facebook to construct models of social ties.[1] Network analyses have also been used to describe relationships between departments competing for undergraduate co-op placements [27], faculty authorship [40], and citation networks [45].

Prior work specifically on the influence of social networks on undergraduate student achievement exists, but its findings have been mixed, limited in scope, and largely exploratory. Prior evidence has suggested that grades are correlated within friendship networks [2] and dorm roommate networks [48, 51]; [60, 65] find roommate effects only for the middle 70% of performers; while [25] finds significant effects for roommates and dorm-mates, but not classmates and [9] observes roommate peer effects which are dependent on major, with large, positive peer effects in the hard sciences and much smaller and ambiguous effects in the humanities and social sciences. Cohort-based coenrollment and team grouping are found to generate networks which are predictive of student grades in a small $N = 250$ MBA cohort in [4]. Other research has found no peer effects in roommate [16, 38, 50], cohort [36], and friendship networks [16]. Finally, we note that "curving" (enforcing specific grade distributions) is also a common practice at American universities, which can cause students' grades to explicitly depend upon those of their peers.

[1]See [8] for a thorough survey of research on social networking sites in this context.

There is also evidence suggesting that analysis of coenrollment might be particularly informative for learning analytics research. For example, [30] demonstrates that coenrollment is strongly related to social tie formation, finding that individuals are 3 times more likely to interact if they also share an acquaintance, and about 140 times more likely if they do not in a university-scale email network $N = 43,553$. Coenrollments' impact on performance in a small ($N = 505$) online masters program is demonstrated in [18], but the interactions that take place online are different from the in-person interactions in residential higher education. Exploration of coenrollment-based effects on performance at scale, across an entire university network, has not verified this result.

2.2 Network-Based Modeling and Multi-View Learning

If peer effects do exist and if the features relevant to these effects are observable, a sufficiently flexible predictive model should be able to capture these effects, even if the underlying relationships are complex and vary by course or subject. However, traditional supervised learning techniques are often unable to account for relationships between observations. Indeed, a core assumption of most supervised learning methods is the independence of data. This motivates a network-based modeling approach to account for dependence between observations.

Link-based modeling is one such approach, and consists of tasks where $\mathbb{G} = (\mathbb{O}, \mathbb{L})$ is fully known, as are all object attributes $OA(\mathbb{G})$. The objective is to label each node $o_i \in \mathbb{O}$ by predicting $c_i \in \mathbb{C}$, the class label (final course grade A, B, C, D) for each node $o_i \in \mathbb{O}$. Cases where $\mathbb{C}$ is also (at least partially) known are called "within-network classification", because the classification takes place within a network where at least some neighboring nodes are already classified. Models in these contexts can exploit the information contained in the labels of neighboring nodes. The LBCT as specified here is *not* a within-network classification task, because in the prediction scenario, $\mathbb{C}$ is entirely unknown: our goal is to make predictions for all students at the beginning of a semester, when no students' final grades are known (but all student and course attributes are known). We therefore adapt a within-network model to use a proxy-labeling approach described in Section 3.2. Other modeling algorithms that have been applied to network classification tasks include conditional random fields [31], relational Markov networks [54], and probabilistic relational models [17, 29].

Link-based classification techniques have been applied to a diverse array of domain-specific tasks, including hypertext categorization [10, 42, 64], blog classification [7], user classification in targeted advertising [24, 44, 53], spam detection [5], customer valuation [14], and fraud detection [12, 15, 43]. Many of these methods are based on the "Markov assumption" that conditional distributions within each class can be approximated using near neighbors instead of full graph; this assumption is central to the structural approach used in the experiment in Section 5.

Previous research has found that feature extraction in network-based models should focus on neighboring *labels* only, and that models which separate link-based and object-based attributes often perform best. In many cases, incorporating object attributes (as

opposed to class labels) from neighbors actually *decreases* classification accuracy while incorporating information about neighboring classes *increases* classification accuracy [10, 20, 35]. The structural model implemented here (described in Section 4) follows this finding by using only the proxy labels of neighboring nodes, but not other features of these nodes, to construct $LA(\mathbb{O}, \mathbb{L})$[19, 21, 34, 35].

Finally, the use of multi-view supervised learning, where models are trained on nonoverlapping feature sets and are ensembled to produce a single, more robust prediction, has previously been applied to learning analytics research in other contexts. In particular, multi-view learning has been used to model the complex phenomena contributing to MOOC dropout [26, 33], but it has not been used in residential grade prediction to the authors' knowledge.

3 DATA

3.1 Student-Course Dataset

The data used in this analysis were drawn from the University of Michigan Learning Analytics Data Architecture (LARC), built from the University of Michigan enrollment and student information systems. LARC includes student-, semester-, and course-level data similar to the records retained by many institutions: student demographic, performance, and registration information; course details such as subject, enrollment, credit hours, meeting days and times; and facility information for the course location, such as instructional technology.[2] The data utilized for this experiment were drawn from winter semesters (January-April) between 2005 and 2015, and represent all undergraduate course records at the University of Michigan in these semesters. The number of records from each semester ranged from 198,544 (Winter 2005) to 232,509 (Winter 2015) before preprocessing.

3.1.1 Data Preprocessing. We perform data preprocessing and filtering for several reasons: (i) institutional factors suggested that grade prediction models would be substantially different between certain student populations (i.e., across student populations – undergraduate vs. graduate vs. professional – or across departments – biochemistry vs. economics vs. English); (ii) computational and modeling factors limited the types of data we were able to consider (i.e., high-cardinality categorical variables); and (iii) practical factors limited types of records for which any model would be able to make predictions (i.e., only for courses in subjects previously observed; only information which is known at the time of registration).

We therefore perform the following filtering and preprocessing steps: First, we only include records for undergraduate students who received a valid grade (no auditors, dropouts, withdraws, or other special cases, as we do not attempt to predict these completion states). Next, we filter the predictors, dropping those not known at beginning of semester or those we do not want model to depend on (i.e., date of most recent SAT/ACT test) and those with $\geq 10\%$ missing data (most modeling algorithms used below require complete cases with no missing data; dropping highly sparse columns is preferable to dropping many observations at modeling time). We create indicator values for missingness in any remaining categorical variables, and drop any categorical variables with > 20 levels as many predictive algorithms limit the cardinality of categorical predictors allowed, and exploratory analysis suggested that the 20-level cutoff retained most variables while only excluding the "long tail" of very high-cardinality predictors. We also chose not to binarize all categorical predictors because there were dozens of categorical fields with hundreds of values each; binarization would have led to an explosion of dimensionality. Finally, we keep only the remaining complete cases (which was > 96% of the remaining data at this step).

After filtering and preprocessing, the training dataset (compiled from Winter 2005 - 2014 records) included 985,291 observations of 116 variables, and the testing dataset (Winter 2015) included 106,265 observations of these same variables.

3.2 Network-Based Features

From the raw tabular dataset, we construct a network $\mathbb{G} = (\mathbb{O}, \mathbb{L})$. Each object (or node) $o_i \in \mathbb{O}$ is a student (these are given from the raw data, as are their individual attributes $OA(\mathbb{O})$), and undirected links (or edges) $\mathbb{L}$ are formed based on student coenrollment and link attributes $LA(\mathbb{O}, \mathbb{L})$ are constructed. Building the network dataset $\mathbb{G}$ consists of two main tasks: *network construction* and *network feature extraction*; we discuss approaches to both tasks here.

Network Construction: This is the procedure for adding links $\mathbb{L}$ to the graph. There are at least two reasonable methods for building $\mathbb{L}$ in the context of a coenrollment network. Consider that there typically exist *multiple* records for a given student in a single semester, each representing one course the student is enrolled in. We could either construct a network where a students' records are linked to all students they are enrolled in *any* courses with, which leads to a larger coenrollment network that accounts for all potential links across classes; or we could construct a network where each record is only linked to other students in the same class, which leads to a narrower, course-specific coenrollment network. We refer to the methods for building these networks from the raw data as *network-building* functions or simply *network builders*; the two network-building functions are shown in Table 1. [3] Network builders define the Markov neighborhood over which we extract link-based features.

Feature Extraction: This is the procedure for generating link attributes, $LA(\mathbb{O}, \mathbb{L})$, once the network has been constructed. The appropriate method to extract or aggregate features across a nodes' neighborhood is not obvious and may depend on contextual factors; prior research has also indicated that it can substantially affect the performance of predictive models using $LA(\mathbb{O}, \mathbb{L})$ [19]. We may think that the absolute *number* of connections in various performance groups (i.e., the number of links to 'A' students, 'B' students, etc) may be relevant to a given students' performance; or, perhaps the *proportion* of links to each type of student may be a better predictor. We explore four different *link feature aggregation functions* (or link feature aggregators), to perform this task. These are *mean-link*, *count-link*, *binary-link*, and *proportion-link*. Each function takes a set of neighboring nodes and their attributes as its input, and

[2] For a more detailed description of the LARC dataset, see https://enrollment.umich.edu/data-research/learning-analytics-data-architecture-larc .

[3] In prior research, these are often called "link types" (i.e. [19]); we find the terminology of "link types" and "link models" to be unnecessarily abstruse and adopt the more descriptive "network builder" function.

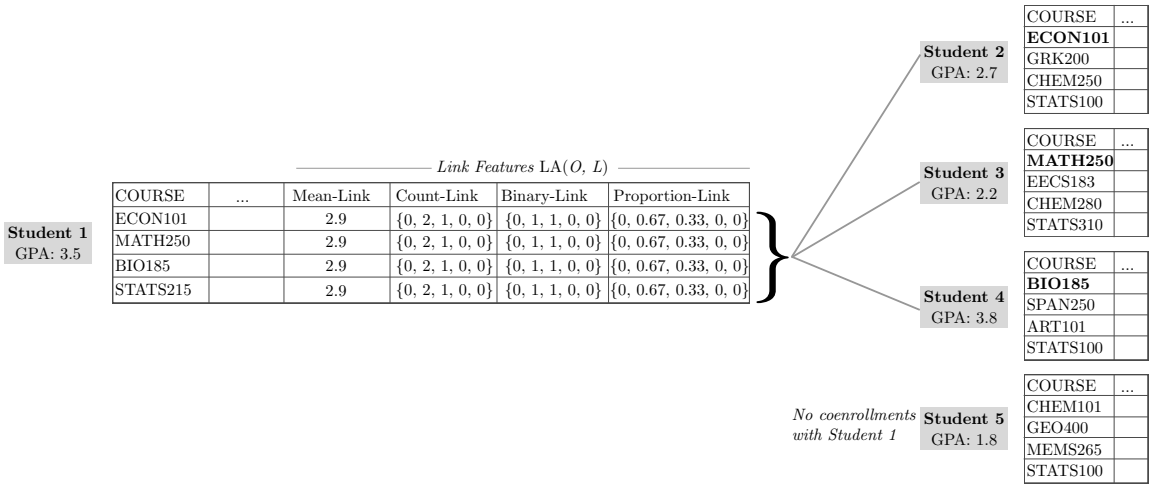

Figure 1: Example of coenrollment-based link features, LA(O, L). In the coenrollment graph shown above, students connected by an edge are coenrolled in a class together. Link features for Student 1 are shown. This example shows student-link features, where each observations' neighborhood includes and students coenrolled in any course with the target student. In course-link features (not shown here), the neighborhood only uses links between records within the same course (links to students in other courses are excluded, considering a more restricted neighborhood for each observation).

returns a single feature vector representing the aggregate features for a given students' coenrollment neighborhood. Definitions of the four feature aggregators used here are shown in Table 1; all but proportion-link are replicated from [19, 34] (proportion-link was added to explore the potential of proportions, not counts, as useful network features).

As we will discuss below, each structural model uses at least one network builder to first define the edges $\mathbb{L}$ of $\mathbb{G}$, and then applies a link feature aggregation function type to aggregate information about class labels over each nodes' neighborhood in to generate link attributes $LA(\mathbb{O}, \mathbb{L})$. In this experiment, we test one set of models using student-link types only, and one set of models which uses *both* course-link and student-link features, following the use of multiple link types in [19, 34] which led to more stable performance across datasets. See Section 5 for more information on the model specifications in this experiment.

A final note on network features: a key difference between our task and many other link-based classification tasks is that here, we do not know the labels of *any* nodes at the time of prediction; $C^{(*)}$ is entirely unknown, that is, we do not know any students' neighbors' final grades at the time we want to predict that students' grade; ideally at the beginning of a course. Therefore, we are unable to directly generate link attributes $LA(\mathbb{O}, \mathbb{L})$ from each node's neighboring class labels as most link-based models do: the input for a feature aggregation function (neighboring nodes' final course grades) would not be known at the beginning of the course. However, in our data (and in most educational settings where this model would be applied), we have a feature which can serve as a strong *proxy label*: students' cumulative GPA in prior semesters. Cumulative GPA is known for every student who is not in their first semester at the institution, and this measure of past performance is strongly correlated to future performance. We thus use students' prior cumulative GPA as a proxy for C in order to generate network features. The concept of proxy labeling has been used in other predictive tasks in education [58], but to our knowledge has not been applied to network-based prediction. This allows us to train and test models exactly as they would be in the real-world version of the LBCT: training on historical data, and predicting on new, disjoint networks, for which all object and link attributes (but no labels) are known. Without this proxy-labeling approach, we would have no link attributes for $\mathbb{G}^{(*)}$ and would not be able to make beginning-of-semester predictions with the structural models defined below.

4 THE COENROLLMENT NETWORK

The exploration, description, and analysis of large and complex networks is an area of open and active research in computer science, statistics, and related fields. In this section, we address this task for the coenrollment network be exploring network properties of degree distribution and assortativity to motivate our modeling approach. Degree distribution is the distribution of the number of edges (called the *degree*) of nodes across $\mathbb{G}$, and is commonly examined by exploring a histogram of the node degree values for each $o_i \in \mathbb{O}$. This is a useful exploration in the case of coenrollment networks because (i) it provides a novel analysis of a previously-unexplored property of coenrollment networks on a full-scale network; (ii) it provides evidence about whether the coenrollment possesses properties similar to other general network types; (iii) it can specifically confirm whether the coenrollment network is sufficiently similar to a document citation network so as to justify the use of the link-based classification model applied to citations; and (iv) it can provide initial evidence of potential differences in network properties based on student performance.

Table 1: Feature aggregators. These represent the method for aggregating link attributes $LA(\mathbb{O}, \mathbb{L})$ over the neighborhood (referred to as *link models* in prior work) [19, 21, 34, 35]. All GPAs in the dataset range from 0 through 4.35. A 4.0 generally represents an 'A' average, a 3.0 a 'B' average, etc.

Feature Aggregator	Definition
Count	Count of neighbors with cumulative GPA in letter grade-level buckets: (4.0, ∞], (3.0, 4.0], (2.0, 3.0], (1.0, 2.0], (0.0, 1.0].
Mean	Mean cumulative GPA of all students in neighborhood of target student.
Binary	Binary Indicator for having neighbors in (4.0, ∞], (3.0, 4.0], (2.0, 3.0], (1.0, 2.0], (0.0, 1.0].
Proportion	Proportion of neighbors in (4.0, ∞], (3.0, 4.0], (2.0, 3.0], (1.0, 2.0], (0.0, 1.0].

Table 2: Network-builder functions. These represent the method for defining the neighborhood over which a given feature aggregation function (Table 1) is applied (referred to as *link types* in prior work).

Network-Builder	Definition
Course	Generate links only to other students in the same course. This generates unique feature values for each individual course, but considers a narrower coenrollment network.
Student	Generate links to any student coenrolled with target student, even those in other courses. This generates identical feature values for each record for a given student, as it builds the coenrollment network across all courses (these are the links shown in Fig. 1).

We examined the cumulative degree distribution (proportion of nodes with degree > n) for the coenrollment network for each of the 12 semesters evaluated in this analysis. A visualization of the cumulative degree distribution of a coenrollment network is shown in Figure 2a and is compared to the cumulative degree distribution of a citation network from [39, 45] in Figure 2b. The similarity suggests a power-law distribution, or *scale-free* network, which means that coenrollment networks are similar in shape to many network types, including social networks, web-page networks, internet nodes, and document citation networks[39].

This experiment is particularly concerned with uncovering and modeling relationships between the coenrollment network and student performance. As such, we also examined the *assortativity* of each coenrollment network. Assortativity measures the tendency for nodes in networks to be connected to other nodes that are like (or unlike) them in some way [41]. We calculate the assortativity coefficient

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b} \tag{1}$$

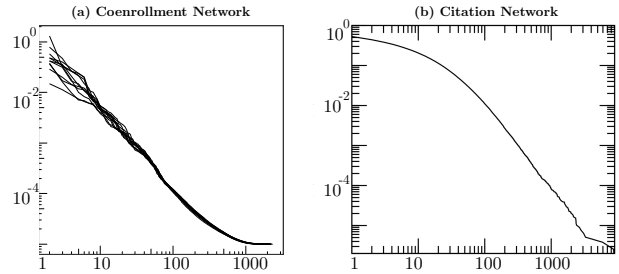

Figure 2: Cumulative degree distribution of (a) coenrollment networks for the 11 semesters evaluated, representing approximately 22,000 nodes per network, compared to the (b) cumulative degree distribution of a citation network from [39, 45]. The similarity of shape on the log scale demonstrates that both networks conform to power-law degree distributions.

as described in [41] using students' previous cumulative GPA (the same outcome used to generate network-based features) to examine the relationship between performance and connectedness in the network. The assortativity coefficient measures the strength of association between a property of nodes in the network and those they are connected to. In this case, the assortativity coefficient measures the association between students' GPA and the GPA of students they are connected to in the coenrollment network. Essentially, the assortativity coefficient measures whether "birds of a feather flock together," "opposites attract," or neither (with respect to academic performance) in the coenrollment network. As a correlation coefficient (r), assortativity ranges from -1 to 1, with positive r values indicating a positive association (high-GPA students are more likely to connect to higher-performing peers), negative r values indicating negative association, and values near zero indicating no association.

Results of this analysis are shown in Table 3. For each of the 12 semesters examined, we find strong evidence of positive assortativity, with $\bar{r} = 0.622$ and $SD(r) = 0.07$; t-testing indicates that these results are highly significant ($p \leq 10^{-16}$). These results demonstrate strong, consistent performance-based assortativity at scale across the entire undergraduate coenrollment network at a major U.S. university, revealing at scale a result that has previously been only suggested by smaller-scale coenrollment network analysis (i.e. [18]). Furthermore, this result suggests that there are performance-based network dynamics across the student coenrollment network that an effective predictive model might capture and use for grade prediction.

Having demonstrated some graphical properties of the network which suggest parallels to networks used in previous link-based predictive models, we finally conducted an initial exploration of the predictiveness of network-based features to verify whether the features we extracted above actually provide additional predictive power in simple models. We construct three simple linear models for each dataset: (a) a model with only the network-based features described above; (b) a model with only previous cumulative GPA; (c) a model with both network features and GPA. Each model is a simple ordinary least squares linear regression model with only

Table 3: Assortativity coefficients [41] for $\mathbb{G}$ across each semester and t-test of Pearson's correlation coefficient. These results demonstrate strong and consistent performance-based assortativity across the entire undergraduate coenrollment network.

Semester	Assortativity Coefficient	t-statistic	$p \leq 10^{-16}$
W2005	0.69	178.9	*
W2006	0.69	180.2	*
W2007	0.69	184.4	*
W2008	0.69	184.3	*
W2009	0.69	182.0	*
W2010	0.68	182.7	*
W2011	0.59	142.4	*
W2012	0.58	139.5	*
W2013	0.54	128.7	*
W2014	0.54	129.5	*
W2015	0.54	129.8	*

Table 4: Initial predictive analysis using simple OLS regression. R^2 values, which demontrate the proportion of variance in the outcome explained by the predictors, are shown for network (all link-based features), GPA (only cumulative GPA), and a combined model including both for each semester. ANOVA results compare the GPA-only model to the combined model, and test whether the network-based features explain a statistically significant additional proportion of the variance in outcome over the GPA-only model. "All Train" represents all training semesters (W2005-W2014).

Semester	Network	GPA	Combined	ANOVA F-Statistic	$p \leq 10^{-16}$
W2005	0.068	0.218	0.246	168.1	*
W2006	0.083	0.210	0.247	230.4	*
W2007	0.083	0.230	0.265	227.2	*
W2008	0.083	0.215	0.246	205.8	*
W2009	0.073	0.220	0.248	184.6	*
W2010	0.082	0.236	0.266	205.5	*
W2011	0.084	0.231	0.264	234.9	*
W2012	0.067	0.203	0.228	178.2	*
W2013	0.069	0.208	0.239	220.1	*
W2014	0.067	0.203	0.234	223.3	*
All Train	0.075	0.218	0.246	1866.6	*
W2015	0.067	0.198	0.233	251.3	*
All	**0.074**	**0.217**	**0.245**	**2063.0**	*

first-order terms, and predicts a continuous outcome of student grade for each record (i.e., 4.0 = A; 3.5 = B+, etc.).

Results of this initial analysis are shown in Table 4 and demonstrate several relevant results. First, network-based features alone explain between 6-8% of the variance in student performance across each semester. Second, even when accounting for GPA (the strongest overall predictor of future student performance), these network-based features explain a statistically significant additional proportion of the variance. The proportion of additional variance explained by network features over a GPA-only model is remarkably consistent at around 3% and is highly statistically significant ($p < 10^{-16}$ for each semester evaluated). This provides further evidence that network-based features can indeed be effective predictors of student performance, and that a network-based model may perform better than a student-only model by accounting for the performance of students' coenrolled peers, motivating the more complex modeling approach in the following experiment.

5 PREDICTION EXPERIMENT

In this section, we describe our methodology for building a network-based structural classification model, beginning with the extraction of network-based features used to construct the model. We implement a version of the structural logistic regression proposed by [19, 21, 34, 35]. Then, we test alternative model specifications, including a single "flat" model which trains a single discriminative classifier on the union of all object and link features, and two multi-view "blended" ensembles.

5.1 Structural Models

The procedure used in this experiment first builds two separate models, a node-based model constructed using each node's object attributes, $OA(\mathbb{O})$ (hereafter called the *student model*) and a network-based model constructed using each node's link attributes $LA(\mathbb{O}, \mathbb{L})$ (hereafter called the *coenrollment model*). Recall that object attributes, $OA(\mathbb{O})$, are the features derived from the student information system; link attributes, $LA(\mathbb{O}, \mathbb{L})$, are the network-based features described in Table 2. These two models are trained separately on disjoint feature sets: the student model on all 116 object attributes; the coenrollment model on link attributes only (between 1 and 5 features, depending on the link feature aggregator used. From these two models, a single model (called the *structural model*) is built by combining the predicted probabilities for each record under an assumption of independence between the two models (see Figure 3 and Table 5).

The procedure for implementing this model is as follows. First, we fit a typical "flat" object-based model, the student model, to estimate the probability of each class label $P(c|OA(\mathbb{O}))$. In the original implementation, this is a penalized logistic regression; however, we instead use a random forest for several reasons: (a) random forests allowed us to consider the wide object attribute feature space without having to perform potentially expensive feature selection or manage multicollinearity, (b) random forests can capture complex interactions between variables, while in a logistic regression these interactions would need to be manually specified, (c) random forests still capture the benefits of discriminative classifiers noted in [20], and (d) prior research suggested that random forests are effective for network-based modeling [57].

We then fit a network-based model, the coenrollment model, to capture the dependence between nodes using the proxy labeling technique described above, estimating $P(c|LA(\mathbb{O}, \mathbb{L})$ using a multinomial logistic regression. Recall that, unlike other implementations of the structural model, this model uses link attributes generated from the cumulative GPA of neighbors, which is known at the time of prediction (this is the *proxy labeling*), not from the true labels

Table 5: Structural model specifications for the student-link model (top) and course-link model (bottom). C represents the class label, in this case the course grade. The course-link model uses both types of links, following [20].

Structural Model with Student Links
$\hat{C}(X) = argmax_{c \in C} \frac{P(c\|OA(\mathbb{O})) \cdot P(c\|LA_{st}(X))}{P(c)}$
Structural Model with Course and Student Links
$\hat{C}(X) = argmax_{c \in C} \frac{P(c\|OA(\mathbb{O})) \cdot \prod_{t \in cr,st} P(c\|LA_t(X))}{P(c)}$

(final course grade) which are not known at the beginning of a semester.

Finally, following the original implementation, this procedure makes the (useful but almost certainly violated) assumption that these two models are independent in order to calculate a joint predicted probability for each observation and each potential outcome class, normalizing by the prior class probability (estimated from the training data) to generate the *structural model*. The specifications for the structural models tested in this experiment are shown in Table 5. The structural modeling approach thus uses two types of models – a student (flat) model, and a coenrollment (network) model – to predict student grades. Note in Table 5 that the structural modeling approach with student-link features utilizes *both* course- and student-links separately, following previous formulations of these models [20]. This allows each link type to have different model parameters.

We fit one model per course subject– i.e., separate models for courses in Mathematics, Statistics, Spanish, etc. – because of previous experience suggesting that instructional approaches as well as course grading policies (curving) were most often formed on a subject/department level. This allowed the models for each subject to have different parameters instead of assuming any similarity in effect across subjects.

5.2 Blended Ensemble Models

In addition to evaluating the structural models, whose independence assumptions were likely to be violated, we also explored ensembling these models. An ensemble can directly account for the dependence between the different models' predictions, learning the relationship between each models' predictions in order to produce a more accurate final prediction [13, 52]. We implemented a model "blending" approach from [6, 55], which is an alternative to the more common stacked generalization technique [61]. Instead of using the complex cross-validation/prediction scheme required by stacked generalization, which is prone to data leakage, blending uses a "probe set", similar to a validation set, which is the only dataset used to fit the ensemble on the predictions of the base learners (here, the base learners consist of the flat model plus each of the network models). For the ensembles, 75% of training data was used for training the base classifiers (the flat model, plus the six network

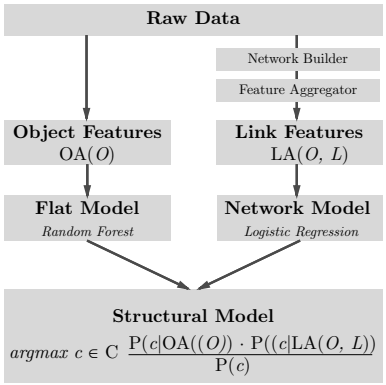

Figure 3: Structural model-building procedure. Separate flat (student) and network (coenrollment) models are constructed from disjoint feature sets. The models are combined to generate a structural model using the predicted probabilities for each class label C and the class prior probability.

models) and 25% was reserved for the probe set, which base models make predictions on (but are not trained on). We explored two meta-learners for the ensemble models – gradient boosted trees (using eXtreme Gradient Boosting, or XGBoost [11]) and a two-layer neural network. XGboost is a common, highly flexible classifier that is often used for ensembling; and a neural network was used in the original blending experiment [6, 55] and was appropriate here because the input data for the ensembles was the predicted class probability for each outcome from each classifier (an $n \times 36$ matrix, with each of the nine models generating four columns – one per class), and therefore contained in the interval [0, 1]. The neural network was a two-layer neural network with 36 nodes per layer and logistic activation function, fit using resilient backpropogation with weight backtracking [47]. The results of these ensemble models are shown in Table 6, and both performed similarly.

This specific ensembling approach, in which different models are fit using different, disjoint sets of features and then ensembled into a single model, is referred to as "multi-view learning" because models are trained on different "views" or representations of the data. Multi-view learning is often able to achieve both greater stability and generalization accuracy over traditional single-view machine learning techniques [52, 62]. Multi-view learning has been used in learning analytics for MOOC dropout prediction [26, 33] as well as for other tasks, such as image recognition (where the term "view" applies more naturally); the other boosting algorithms (Adaboost) have been adapted elsewhere successfully for multi-view ensemble learning [63]. Multi-view models are also useful because inspection of learned models can yield insights on which "views" of the data contribute most to effective models, similar to a feature importance analysis, where the sets of features actually represent lower-order models.

Table 6: Structural model performance results on independent test dataset (future semester). All structural models achieve performance above the baselines, but only the ensembles and student-link models exceed the "flat" model which uses no structural features. Ensemble models can utilize the information from both the flat and network models while modeling the dependency between them (instead of assuming their independence, as a structural model does). Reported sensitivity and specificity are mean one-vs-all values measured across each of the four outcome classes (A, B, C, D); AUC value is multiclass AUC, see [23] for details. *: Note that these models replicate the original structural model from [20] (with the exception of proportion-link, which was not previously evaluated).

Model Type	Model	AUC	Accuracy	Kappa	Sensitivity	Specificity
Multi-View Ensemble	Neural Network	0.734	0.660	0.344	0.375	0.838
	XGBoost	0.731	0.658	0.341	NA	0.919
Structural (student-link)	Binary-Link	0.717	0.645	0.332	0.403	0.835
	Mean-Link	0.717	0.648	0.327	0.394	0.833
	Count-Link	0.714	0.646	0.321	0.393	0.831
	Proportion-Link	0.713	0.647	0.322	0.392	0.831
Flat	Flat (no structural features)	0.707	0.653	0.314	0.365	0.828
	Full (with structural features)	0.699	0.652	0.307	0.358	0.826
Baseline	Predict GPA Baseline	0.685	0.522	0.212	0.357	0.813
Structural (course- *and* student-link)*	Mean-Link	0.678	0.642	0.291	0.346	0.823
	Binary-Link	0.675	0.640	0.297	0.356	0.825
	Count-Link	0.669	0.638	0.282	0.347	0.820
	Proportion-Link	0.669	0.639	0.281	0.343	0.820
Baseline	Majority Class Baseline	0.500	0.572	0.000	0.250	0.750

5.3 Results

The results of our prediction experiment are shown in Table 6. We find that multi-view ensembles outperform all other model specifications across all performance metrics considered (Multiclass AUC [23], accuracy, Fleiss' Kappa and average multiclass sensitivity, specificity), demonstrating how sophisticated models may can capture the complex interplay between student and link attributes across the multiple "views" represented by several lower-order models. Models using the student-link network builder function also achieve performance consistently above traditional "flat" models, suggesting that even these models (which are simpler to train than the multi-view ensembles) can improve upon non-network approaches.

However, the more sophisticated structural model, which utilizes both student- *and* course network builders and replicates the original structural model from [20], performs below the flat model, and even below a baseline of simply predicting grades using students' GPA. This finding, that multiple link types do not improve performance, is counter to the findings in [20]. This suggests that the structural model's core assumption – that the object and link features are independent – may be so strongly violated in this data that it mitigates the performance gains from link modeling observed in [20].

Such a conclusion is supported by the superior performance of the ensembles: an ensemble model, which takes the predictions of each (flat + eight network models) as input, is able to directly model and exploit the dependencies between the predictions of each model, instead of assuming their independence. We note that *only* the ensembles utilize this information about dependence between network and flat features in a way that improves future prediction performance. The flat model trained with structural features (i.e., using the union of the disjoint student and network features, $OA(\mathbb{O}) \cup LA(\mathbb{O}, \mathbb{L})$) showed almost no difference in performance relative to the flat model with only student features.

Additionally, we note that all student-link models outperformed all course- plus student-link models. This suggests that the wider neighborhood considered by the course-link models is sufficient to capture relevant information about each students' neighbors in the graph, and that the additional data provided by a student-link model is overwhelmed by the additional assumption of independence between the two link types shown in the formulation in Table 5.

An expanded flat model with structural features, identical to the original flat model but with features for all link-based attributes simply appended to the student features, does not achieve a performance improvement over a model without network features, suggesting that building flat and network models independently, but then modeling the dependence between those models with an ensemble, might be the most performant approach: ensemble models achieved better performance than any of the structural or flat models (see Table 6).

6 CONCLUSIONS AND FUTURE RESEARCH

This investigation makes several contributions to the literature regarding network analysis in higher education, including providing both descriptive and predictive analyses of coenrollment networks, at scale, over 10 years of university records and over 1 million individual records. Our analysis demonstrates that university coenrollment networks display the power-law degree distributions common to many networks, including the citation networks used as the basis for previous link-based predictive models. We also demonstrate strong, consistent assortativity within this network, which reveals an association between student performance and the likelihood of having a connection in the coenrollment graph. We also demonstrate that network-based features explain a statistically

significant additional proportion of the variation in student performance over GPA-only models, contributing around 3% for every semester evaluated.

We also make the novel contribution of developing an extension of link-based classification models used for document classification, modifying these models to predict on future semesters. These structural models outperform traditional "flat" grade prediction models, but only with coenrollment networks constructed with student-level links. Structural models with student- and course-level links, which follow the original implementation [19], perform worse than flat models, likely due to their additional assumptions of independence between the student- and course-level structural models. We construct multi-view blended ensembles of the full set of structural and flat models, which achieve further performance gains, and demonstrates how ensembling can utilize the different "views" in each structural model to achieve further improvements in generalization performance. These performance gains are at least partially due to the ensembles' ability to account for the non-independence between different model predictions, instead of assuming independence, as the most complex structural models do.

These results suggest that network-based features can improve predictive models of student grades, but that structural models which make strong assumptions about independence between object and link attributes may not realize these performance gains. Additionally, the results demonstrate how diverse models with different feature sets and functional forms can be combined in a multi-view ensemble to improve predictive performance by exploiting the ways in which the models err differently. They also demonstrate a first application of model blending to predictive models in education, which can simplify the common challenge of ensembling models across different "views" of students and their performance.

The different predictive performance of the link types considered suggests that different types of student relationships (same-course vs. across-course) might vary in their relevance to grade prediction; we leave a more exhaustive comparison of different link types (as well as second-degree, same-major, and various other types of links that could be constructed from the available data) to a future work.

Our future research includes both a more detailed inspection of the results of this and other coenrollment-based predictive models, as well as broader exploration of other modeling techniques. In particular, the use of feature importance analysis, contribution analysis, or local model inspection [46] would further illuminate how different network models contribute to the multi-view ensemble, providing guidance for future work. Modeling techniques of interest for future work include other discriminative classifiers (e.g. SVM) for creating the base student and coenrollment models, and other ensembling techniques (e.g. blend optimization [55]). Additionally, further research into the feature space used to build the coenrollment/network model is required, and should explore different and novel methods for building link features: consider various sizes of neighborhood with different link types; more or less granular bucketing for the count- and binary-link features; using different proxy labeling techniques; building temporal feature sets which extend back over multiple semesters or coenrollment periods; considering pre-requisite restrictions from a course graph (as not all students could be coenrolled with other students); using the predicted grades of neighbors as "bootstrap" estimates and building a more traditional within-network model. These are some of the many ways in which the current method, adapted from hypertext document classification, might be better modified to fit the context of student grade prediction.

Other data sources may also be useful in future analyses. For example, more robust and granular data on student networks and communication could be collected from course discussion fora, or activity-based measurements from learning management systems, and used to augment the current feature set. This experiment should be applied to other institutional datasets, where the network structure might be quite different. Similarly, attendance data, group project data, and even seating data within courses could be valuable in building network models of learners, though the scale of collecting this data is somewhat daunting.

This experiment points to the need for further research about the effect of social networks on learning, particularly coenrollment networks. Further investigation into the potential mechanisms through which coenrollment influences student performance – if such an effect indeed exists – and how different types of coenrollment relationships (course- and student-link) differentially impact performance will provide a stronger theoretical foundation for future predictive modeling efforts.

Finally, future research should investigate and demonstrate the actionable insights supported by such models, and how they can support real-time decisionmaking for instructors, students, and advisors both during course selection and in the early stages of the course itself.

7 ACKNOWLEDGEMENTS

The authors would like to thank the participants and organizers of the 2017 Workshop on Graph-Based Educational Data Mining [37] for their feedback on an initial version of this work, as well as the Michigan Institute for Data Science (MIDAS) for funding under the Holistic Modeling of Education (HOME) grant.

REFERENCES

[1] Paul Adams. 2006. Exploring social constructivism: theories and practicalities. *Education 3-13* 34, 3 (Oct. 2006), 243–257.

[2] J S Antrobus, R Dobbelaer, and S Salzinger. 1988. Social networks and college success, or grade point average and the friendly connection. *Social networks of children, adolescents, and college students* 227 (1988), 260.

[3] Sara Baker, Adalbert Mayer, and Steven L Puller. 2011. Do more diverse environments increase the diversity of subsequent interaction? Evidence from random dorm assignment. *Econ. Lett.* 110, 2 (2011), 110–112.

[4] Timothy T Baldwin, Michael D Bedell, and Jonathan L Johnson. 1997. The social fabric of a team-based MBA program: Network effects on student satisfaction and performance. *Acad. Manage. J.* 40, 6 (1997), 1369–1397.

[5] L Becchetti, C Castillo, D Donato, and others. 2008. Link analysis for web spam detection. *ACM Transactions on* (2008).

[6] James Bennett, Stan Lanning, and Others. 2007. The netflix prize. In *Proceedings of KDD cup and workshop*, Vol. 2007. brettb.net, 35.

[7] Smriti Bhagat, Irina Rozenbaum, and Graham Cormode. 2007. Applying Link-based Classification to Label Blogs. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07)*. ACM, New York, NY, USA, 92–101.

[8] Susan Biancani and Daniel McFarland. 2013. Social Networks Research in Higher Education. *Educ. Stud.* 4 (2013), 85–126.

[9] Giorgio Brunello, Maria de Paola, and Vincenzo Scoppa. 2010. PEER EFFECTS IN HIGHER EDUCATION: DOES THE FIELD OF STUDY MATTER? *Econ. Inq.* 48, 3 (July 2010), 621–634.

[10] Soumen Chakrabarti, Byron Dom, and Piotr Indyk. 1998. Enhanced Hypertext Categorization Using Hyperlinks. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data (SIGMOD '98)*. ACM, New York, NY,

USA, 307–318.
[11] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 785–794.
[12] Corinna Cortes, Daryl Pregibon, and Chris Volinsky. 2001. Communities of Interest. In *Advances in Intelligent Data Analysis*. Springer, Berlin, Heidelberg, 105–114.
[13] Thomas G Dietterich. 2000. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*. Springer, Berlin, Heidelberg, 1–15.
[14] Pedro Domingos and Matt Richardson. 2001. Mining the Network Value of Customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '01)*. ACM, New York, NY, USA, 57–66.
[15] Tom Fawcett and Foster Provost. 1997. Adaptive Fraud Detection. *Data Min. Knowl. Discov.* 1, 3 (Sept. 1997), 291–316.
[16] Gigi Foster. 2006. It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *J. Public Econ.* 90, 8–9 (2006), 1455–1475.
[17] Nir Friedman, Lise Getoor, Daphne Koller, and Avi Pfeffer. 1999. Learning probabilistic relational models. In *IJCAI*, Vol. 99. robotics.stanford.edu, 1300–1309.
[18] Dragan Gašević, Amal Zouaq, and Robert Janzen. 2013. "Choose your classmates, your GPA is at stake!" The association of cross-class social ties and academic performance. *Am. Behav. Sci.* 57, 10 (2013), 1460–1479.
[19] Lise Getoor. 2005. Link-based Classification. In *Advanced Methods for Knowledge Discovery from Complex Data*. Springer London, 189–207.
[20] Lise Getoor and Christopher P Diehl. 2005. Link Mining: A Survey. *SIGKDD Explor. Newsl.* 7, 2 (Dec. 2005), 3–12.
[21] Lise Getoor and Lilyana Mihalkova. 2011. Learning Statistical Models from Relational Data. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD '11)*. ACM, New York, NY, USA, 1195–1198.
[22] A F Hadwin and Sanna Järvelä. 2011. Introduction to a special issue on social aspects of self-regulated learning: Where social and self meet in the strategic regulation of learning. *Teach. Coll. Rec.* 113, 2 (2011), 235–239.
[23] David J Hand and Robert J Till. 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Mach. Learn.* 45, 2 (Nov. 2001), 171–186.
[24] Shawndra Hill, Foster Provost, and Chris Volinsky. 2006. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Stat. Sci.* 21, 2 (2006), 256–276.
[25] Jessica Hoel, Jeffrey Parker, and Jon Rivenburg. 2005. Peer effects: do first-year classmates, roommates, and dormmates affect studentsâĂŹ academic success. In *Higher education data sharing consortium winter conference, Santa Fe, NM*. Citeseer.
[26] Feng Jiang and Wentao Li. 2017. Who Will Be the Next to Drop Out? Anticipating Dropouts in MOOCs with Multi-View Features. *International Journal of Performability Engineering* 13, 2 (2017).
[27] Yuheng Jiang and Lukasz Golab. [n. d.]. On Competition for Undergraduate Co-op Placements: A Graph Mining Approach. ([n. d.]).
[28] David W Johnson and Roger T Johnson. 2009. An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educ. Res.* 38, 5 (June 2009), 365–379.
[29] Daphne Koller. 1999. Probabilistic Relational Models. In *Inductive Logic Programming*. Springer, Berlin, Heidelberg, 3–13.
[30] Gueorgi Kossinets and Duncan J Watts. 2006. Empirical analysis of an evolving social network. *Science* 311, 5757 (Jan. 2006), 88–90.
[31] J Lafferty, A McCallum, and F Pereira. [n. d.]. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ([n. d.]).
[32] Conrad Lee, Thomas Scherngell, and Michael J Barber. 2011. Investigating an online social network using spatial interaction models. *Soc. Networks* 33, 2 (2011), 129–133.
[33] W Li, M Gao, H Li, Q Xiong, J Wen, and Z Wu. 2016. Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In *2016 International Joint Conference on Neural Networks (IJCNN)*. ieeexplore.ieee.org, 3130–3137.
[34] Qing Lu and Lise Getoor. 2003. Link-based classification. In *ICML*, Vol. 3. vvvvw.aaai.org, 496–503.
[35] Q Lu and L Getoor. 2003. Link-based classification using labeled and unlabeled data. In *ICML 2003 workshop on The Continuumfrom Labeled to Unlabeled Data in Machine Learning and Data Mining*.
[36] David S Lyle. 2007. Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Rev. Econ. Stat.* 89, 2 (April 2007), 289–299.
[37] Collin F Lynch, Tiffany Barnes, Linting Xue, and Niki Gitinabard. 2017. Graph-based Educational Data Mining (G-EDM 2017). In *Proceedings of the 10th International Conference on Educational Data Mining*, Xiangen Hu, Tiffany Barnes, Arnon Hershkovitz, and Luc Paquette (Eds.). 472–473.
[38] Patrick J McEwan and Kristen A Soderberg. 2006. Roommate Effects on Grades: Evidence from First-Year Housing Assignments. *Res. High. Educ.* 47, 3 (May 2006), 347–370.
[39] M Newman. 2003. The Structure and Function of Complex Networks. *SIAM Rev.* 45, 2 (Jan. 2003), 167–256.
[40] M E Newman. 2001. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2 (Jan. 2001), 404–409.
[41] M E J Newman. 2003. Mixing patterns in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67, 2 Pt 2 (Feb. 2003), 026126.
[42] Hyo-Jung Oh, Sung Hyon Myaeng, and Mann-Ho Lee. 2000. A Practical Hypertext Catergorization Method Using Links and Incrementally Available Class Information. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*. ACM, New York, NY, USA, 264–271.
[43] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos. 2007. Netprobe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 201–210.
[44] Foster Provost, Brian Dalessandro, Rod Hook, Xiaohan Zhang, and Alan Murray. 2009. Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 707–716.
[45] S Redner. 1998. How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* 4, 2 (July 1998), 131–134.
[46] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. dl.acm.org, 1135–1144.
[47] M Riedmiller and H Braun. 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In *IEEE International Conference on Neural Networks*. ieeexplore.ieee.org, 586–591 vol.1.
[48] Bruce Sacerdote. 2000. Peer Effects with Random Assignment: Results for Dartmouth Roommates. (Jan. 2000).
[49] Sarah Schrire. 2006. Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Comput. Educ.* 46, 1 (Jan. 2006), 49–70.
[50] John J Siegfried and Michael A Gleason. 2006. Academic roommate peer effects. *Unpublished manuscript, Vanderbilt Univ. , Nashville* (2006).
[51] Ralph Stinebrickner and Todd R Stinebrickner. 2006. What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *J. Public Econ.* 90, 8–9 (2006), 1435–1454.
[52] Shiliang Sun. 2013. A survey of multi-view machine learning. *Neural Comput. Appl.* 23, 7-8 (Dec. 2013), 2031–2038.
[53] Lei Tang and Huan Liu. 2011. Leveraging social media networks for classification. *Data Min. Knowl. Discov.* 23, 3 (Nov. 2011), 447–478.
[54] Ben Taskar, Pieter Abbeel, and Daphne Koller. 2002. Discriminative Probabilistic Models for Relational Data. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI'02)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 485–492.
[55] Andreas Töscher, Michael Jahrer, and Robert M Bell. 2009. The bigchaos solution to the netflix grand prize. *Netflix prize documentation* (2009), 1–52.
[56] A Traud, E Kelsic, P Mucha, and M Porter. 2011. Comparing Community Structure to Characteristics in Online Collegiate Social Networks. *SIAM Rev.* 53, 3 (Jan. 2011), 526–543.
[57] Anneleen Van Assche, Celine Vens, Hendrik Blockeel, and Sašo Dzeroski. 2004. A random forest approach to relational learning. In *Workshop on Statistical Relational Learning*.
[58] Jacob Whitehill, Kiran Mohan, Daniel Seaton, Yigal Rosen, and Dustin Tingley. 2017. Delving Deeper into MOOC Student Dropout Prediction. (Feb. 2017). arXiv:cs.AI/1702.06404
[59] Andreas Wimmer and Kevin Lewis. 2010. Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *Am. J. Sociol.* 116, 2 (2010), 583–642.
[60] Gordon Winston and David Zimmerman. 2004. Peer effects in higher education. In *College choices: The economics of where to go, when to go, and how to pay for it*. University of Chicago Press, 395–424.
[61] David H Wolpert. 1992. Stacked generalization. *Neural Netw.* 5, 2 (1992), 241–259.
[62] Chang Xu, Dacheng Tao, and Chao Xu. 2013. A Survey on Multi-view Learning. (April 2013). arXiv:cs.LG/1304.5634
[63] Zhijie Xu and Shiliang Sun. 2010. An Algorithm on Multi-View Adaboost. In *Neural Information Processing. Theory and Algorithms (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 355–362.
[64] Tong Zhang, Alexandrin Popescul, and Byron Dom. 2006. Linear Prediction Models with Graph Regularization for Web-page Categorization. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 821–826.
[65] David J Zimmerman. 2003. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Rev. Econ. Stat.* 85, 1 (Feb. 2003), 9–23.