JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 5(3), 42—59. http://dx.doi.org/10.18608/jla.2018.53.4

Learn From Your (Markov) Neighbour: Co-enrollment, Assortativity, and Grade Prediction in Undergraduate Courses

Josh Gardner,[1] Christopher Brooks,[2] Warren Li[3]

Abstract

In this paper, we evaluate the complete undergraduate co-enrollment network over a decade of education at a large American public university. We provide descriptive and exploratory analyses of the network, demonstrating that the co-enrollment 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 improve GPA-based student performance predictors. We model the university-wide undergraduate co-enrollment network as an undirected graph, and implement multiple network-augmented approaches to student grade prediction, including an adaption of the structural modelling approach from (Getoor, 2005; Lu & Getoor, 2003a). We compare the performance of this 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 modelling strategies that are likely to be most effective for grade prediction.

Notes for Practice

- The co-enrollment networks evaluated here demonstrate power-law degree distributions common to many other types of networks.
- Structural models and, in particular, multi-view ensembles of structural models and traditional "flat" models can improve over the performance of non-network models for student grade prediction in residential higher education courses.
- Assumptions of independence between network and non-network features in statistical models may be strongly violated, which can reduce the performance of models that cannot model or otherwise incorporate this dependence.

Keywords

Co-enrollment, graph mining, predictive modelling

Submitted: 23.06.2018 — **Accepted:** 01.09.2018 — **Published:** 11.12.2018

[1]*Corresponding author Email: jpgard@umich.edu Address: School of Information, The University of Michigan, Ann Arbor, MI, USA, 105 S State St, Ann Arbor, MI 48109*

[2]*Email: brooksch@umich.edu Address: School of Information, The University of Michigan, Ann Arbor, MI, USA, 105 S State St, Ann Arbor, MI 48109*

[3]*Email: liwarren@umich.edu Address: School of Information, The University of Michigan, Ann Arbor, MI, USA, 105 S State St, Ann Arbor, MI 48109*

1. Introduction

Co-enrollment 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 co-enrollment networks to 1) analyze the properties of these networks at a scale not previously examined, and 2) explore the effectiveness of grade prediction techniques that utilize co-enrollment networks. The motivation for this approach is twofold: First, prior research suggests that there are aspects of university social networks that may be relevant to student performance (Gasevic 2013), but such research has not examined co-enrollment networks at scale, nor has it used these networks for predictive modelling. Our exploratory analysis reveals important properties of co-enrollment 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 that are otherwise assumed to be independent. We find that structural models using student-level co-enrollment 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 methods.

These results demonstrate the effectiveness of network-based models in learning analytics, and the importance of co-enrollment 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, course selection tools, and other applications that 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 co-enrollment network and the classification task used in this work. We present the co-enrollment 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 (co-enrollments). For each node $o_i \in \mathbb{O}$, we know a set of *object attributes* $OA(o_i)$ that here include student- and course-specific attributes such as gender, ethnicity, cumulative GPA, subject, and major (these are the attributes normally available to institutions from student information systems, self-reports on admissions applications, etc.). 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 neighbours (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:

> **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 in which we wish to predict grades, where all object attributes $OA(o_i)$ are known for each $o_i \in \mathbb{O}$.

2. Previous Work

2.1 Social Learning and Graph-Based Modelling 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 (Gašević, Zouaq, & Janzen, 2013) and lies at the core of many modern pedagogical approaches, including social constructivism (Adams, 2006) and co-operative learning (Johnson & Johnson, 2009). A variety of empirical and experimental research has demonstrated positive relationships between social interactions in courses and learning outcomes (Gašević et al., 2013), including elevated cognition (Schrire, 2006) and self-regulation (Hadwin & Järvelä, 2011).

Network analysis has seen limited application in higher education research, where the influence of other students on a given student's performance is known as a "peer effect," but analysis of co-enrollment networks has been limited in both quantity and scale. In a thorough overview, (Biancani & McFarland, 2013) identifies at least 56 social network-based studies that use individual university students as the unit of analysis, 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 (Baker, Mayer, & Puller, 2011), friendship networks (Wimmer & Lewis, 2010), and demographic networks based on geographic background (Lee, Scherngell, & Barber, 2011), and networks based on institutional factors such as major and class year (Traud, Kelsic, Mucha, & Porter, 2011). 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 (Y. Jiang & Golab, 2016), faculty authorship (M. E. Newman, 2001), and citation networks (Redner, 1998).

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 (Antrobus, Dobbelaer, & Salzinger, 1988) and dorm roommate networks (Sacerdote, 2000; Stinebrickner

[1] See (Biancani & McFarland, 2013) for a thorough survey of research on social networking sites in this context.

ISSN 1929-7750 (online).

& Stinebrickner, 2006). Some researchers (Zimmerman, 2003; Winston & Zimmerman, 2004) find roommate effects only for the middle 70% of performers; while others (Hoel, Parker, & Rivenburg, 2005) find significant effects for roommates and dorm-mates, but not classmates. Still others (Brunello, de Paola, & Scoppa, 2010) observe roommate peer effects that are dependent on major, with large, positive peer effects in the hard sciences and much smaller, ambiguous effects in the humanities and social sciences. Cohort-based co-enrollment and team grouping are found to generate networks that are predictive of student grades in a small $N = 250$ MBA cohort (Baldwin, Bedell, & Johnson, 1997). Other research has found no peer effects in roommate (McEwan & Soderberg, 2006; Foster, 2006; Siegfried & Gleason, 2006), cohort (Lyle, 2007), and friendship networks (Foster, 2006). Finally, we note that "curving" (enforcing specific grade distributions) is also a common practice at American universities, which can cause student grades to depend explicitly upon those of their peers.

There is also evidence suggesting that analysis of co-enrollment might be particularly informative for learning analytics research. For example, (Kossinets & Watts, 2006) demonstrates that co-enrollment is strongly related to social tie formation, finding that co-enrollment makes individuals three times more likely to interact (relative to non-co-enrolled peers) if they also share an acquaintance, and 140 times more likely if they do not share an acquaintance, in a university-scale email network $N = 43,553$. The impact of co-enrollment on performance in a small ($N = 505$) online masters program has been demonstrated (Gašević et al., 2013), but the interactions that take place online are different from the in-person interactions in residential higher education. Exploration of co-enrollment-based effects on performance at scale, across an entire university network, has not verified this result.

2.2 Network-Based Modelling 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 each observation from all others. This motivates a network-based modelling approach to account for dependence between observations.

Link-based modelling 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 neighbouring nodes are already classified. Models in these contexts can exploit the information contained in the labels of neighbouring 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 student's final grades are known (but all student and course attributes are known). We therefore adapt a within-network model to use a proxy labelling approach described in Section 3.2. Other modelling algorithms that have been applied to network classification tasks include conditional random fields (Lafferty, McCallum, & Pereira, 2001), relational Markov networks (Taskar, Abbeel, & Koller, 2002), and probabilistic relational models (Koller, 1999; Friedman, Getoor, Koller, & Pfeffer, 1999).

Link-based classification techniques have been applied to a diverse array of domain-specific tasks, including hypertext categorization (Chakrabarti, Dom, & Indyk, 1998; Oh, Myaeng, & Lee, 2000; Zhang, Popescul, & Dom, 2006), blog classification (Bhagat, Rozenbaum, & Cormode, 2007), user classification in targeted advertising (Hill, Provost, & Volinsky, 2006; Provost, Dalessandro, Hook, Zhang, & Murray, 2009; Tang & Liu, 2011), spam detection (Becchetti, Castillo, Donato, & others, 2008), customer valuation (Domingos & Richardson, 2001), and fraud detection (Cortes, Pregibon, & Volinsky, 2001; Fawcett & Provost, 1997; Pandit, Chau, Wang, & Faloutsos, 2007). Many of these methods are based on the "Markov assumption" that conditional distributions within each class can be approximated using near neighbours 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 neighbouring *labels* only, and that models separating link-based and object-based attributes often perform best. In many cases, incorporating object attributes (as opposed to class labels) from neighbours actually *decreases* classification accuracy while incorporating information about neighbouring classes *increases* classification accuracy (Chakrabarti et al., 1998; Getoor & Diehl, 2005; Lu & Getoor, 2003b). The structural model implemented here (described in Section 5.1) follows this finding by using only the proxy labels of neighbouring nodes, but not other features of these nodes, to construct $LA(\mathbb{O}, \mathbb{L})$(Getoor, 2005; Getoor & Mihalkova, 2011; Lu & Getoor, 2003a, 2003b).

Finally, the use of multi-view supervised learning, where models are trained on non-overlapping 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 (F. Jiang & Li, 2017; Li et al., 2016), 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 pre-processing.

3.1.1 Data Pre-processing

We perform data pre-processing and filtering for several reasons: 1) 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); 2) computational and modelling factors limited the types of data we were able to consider (i.e., high-cardinality categorical variables); and 3) 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 that is known at the time of registration).

We therefore perform the following filtering and pre-processing 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 the model to depend on (i.e., date of most recent SAT/ACT test) and those with $\geq 10\%$ missing data (most modelling algorithms used below require complete cases with no missing data; dropping highly sparse columns is preferable to dropping many observations at modelling 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 pre-processing, 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 co-enrollment 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 three reasonable methods for building $\mathbb{L}$ in the context of a co-enrollment 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. Within a given semester, we could construct a network where a student's records are linked to all students they are enrolled in *any* courses with, which leads to a larger co-enrollment 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 co-enrollment network. We could also look back to a *previous* semester, to observe which students had completed courses together in the prior semester and use these links to connect observations in the target semester. We refer to the methods for building these networks from the raw data as *network-building* functions or simply *network builders*; the three network-building functions are shown in Table 1.[3] Network builders define the Markov neighbourhood 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 node's neighbourhood 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

[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" (Getoor, 2005); we find the terminology of "link types" and "link models" to be unnecessarily abstruse and adopt the more descriptive "network builder" function.

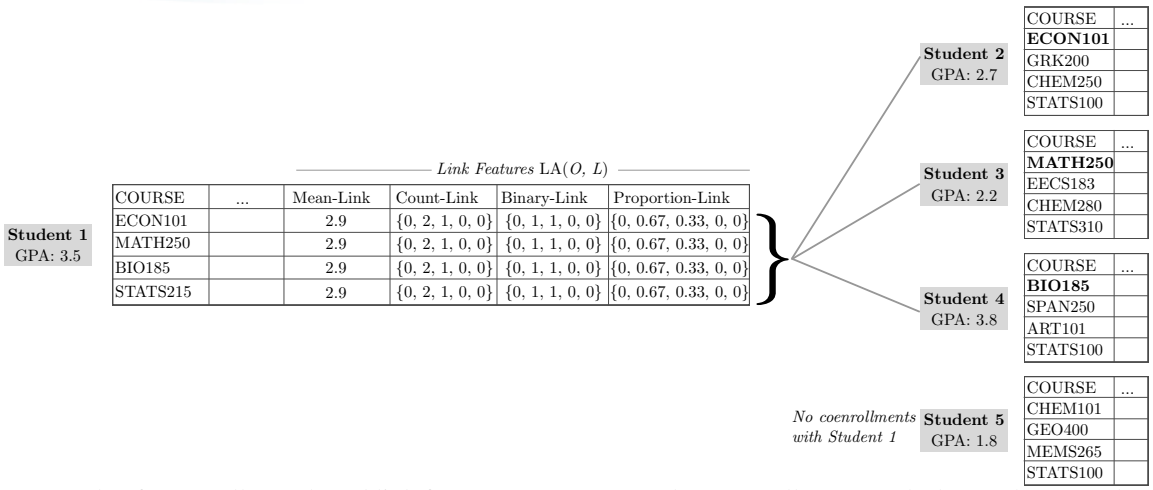

Figure 1. Example of co-enrollment-based link features, LA(O, L). In the co-enrollment graph shown above, students connected by an edge are co-enrolled in a class together. Link features for Student 1 are shown. This example shows student-link features, where each observation's neighbourhood includes students co-enrolled in any course with the target student. In course-link features (not shown here), the neighbourhood only uses links between records within the same course (links to students in other courses are excluded, considering a more restricted neighbourhood for each observation); temporal-link features use student-links from the previous semester.

$LA(\mathbb{O},\mathbb{L})$ (Getoor, 2005). 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 student's 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 neighbouring nodes and their attributes as its input, and returns a single feature vector representing the aggregate features for a given student's co-enrollment neighbourhood. Definitions of the four feature aggregators used here are shown in Table 1; all but proportion-link are replicated from (Getoor, 2005; Lu & Getoor, 2003a) (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 node's neighbourhood in to generate link attributes $LA(\mathbb{O},\mathbb{L})$. In this experiment, we test one set of models using each individual link type (student, course, and temporal), and one set of models that uses *both* course-link and student-link features, following the use of multiple link types (Getoor, 2005; Lu & Getoor, 2003a), which led to more stable performance across datasets in the original work. See Table 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 student's neighbour's final grade at the time we want to predict that student's 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 neighbouring class labels as most link-based models do: the input for a feature aggregation function (neighbouring 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 that can serve as a strong *proxy label*: student 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 student prior cumulative GPA as a proxy for C in order to generate network features. The concept of proxy labelling has been used in other predictive tasks in education (Whitehill, Mohan, Seaton, Rosen, & Tingley, 2017), 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 labelling 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.

ISSN 1929-7750 (online).

Table 1. Feature aggregators. These represent the method for aggregating link attributes $LA(\mathbb{O}, \mathbb{L})$ over the neighbourhood (referred to as *link models* in prior work) (Lu & Getoor, 2003b, 2003a; Getoor, 2005; Getoor & Mihalkova, 2011). 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 neighbours 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 neighbourhood of target student.
Binary	Binary indicator for having neighbours in (4.0, ∞], (3.0, 4.0], (2.0, 3.0], (1.0, 2.0], (0.0, 1.0].
Proportion	Proportion of neighbours 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 neighbourhood 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 co-enrollment network.
Student	Generate links to any student co-enrolled with target student, even those in other courses. This generates identical feature values for each record for a given student, as it builds the co-enrollment network across all courses (these are the links shown in Fig. 1).
Temporal	Generate links to any student co-enrolled with the target student in any course in the preceding semester. This is the equivalent of using student-links from the previous semester.

4. The Co-enrollment 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 co-enrollment network by exploring network properties of degree distribution and assortativity to motivate our modelling approach.

4.0.1 Co-enrollment Networks Display Power-Law Degree Distribution

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 co-enrollment networks because 1) it provides a novel analysis of a previously unexplored property of co-enrollment networks on a full-scale network; 2) it provides evidence about whether the co-enrollment possesses properties similar to other general network types; 3) it can specifically confirm whether the co-enrollment 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 4) it can provide initial evidence of potential differences in network properties based on student performance.

We examined the cumulative degree distribution (proportion of nodes with degree $> n$) for the co-enrollment network for each of the 12 semesters evaluated in this analysis. A visualization of the cumulative degree distribution of a co-enrollment network is shown in Figure 2a and is compared to the cumulative degree distribution of a citation network from (M. Newman, 2003; Redner, 1998) in Figure 2b. The similarity suggests a power-law distribution, or *scale-free* network, which means that co-enrollment networks are similar in shape to many network types, including social networks, web-page networks, internet nodes, and document citation networks (M. Newman, 2003).

4.0.2 Exploratory Co-enrollment Network Analysis

Before providing an analysis of assortativity and the predictive capacity of network-based features, we offer exploratory results, both to motivate the models implemented in Section 5 and to provide an empirical basis for comparison in future work analyzing different co-enrollment networks. First, we evaluate the degree distribution across a coarse grouping of students' cumulative GPA. This analysis is shown in Figure 3, and demonstrates only minor differences in students' network degree by GPA. Kolmogorov-Smirnov tests rejected the null hypothesis that the degree distribution for each adjacent group was identical (e.g., rejected the null that the distribution of degree for students in *(3,4]* is identical to the distribution for students in *(2,3]*, with $p \leq 0.002$ for all comparisons). However, this visual display suggests that while these differences are statistically significant, if assortativity can be utilized in grade prediction, the patterns are likely much more complex than simply partitioning based on students' network degree. We will demonstrate an approach to utilizing network-based assortativity data in conjunction with node-based student data in Section 5.

Second, we also explored the identification of groupings in the network by applying t-Stochastic Neighbour Embedding (t-SNE) to the union of all network features. These results are shown in Figure 4, which demonstrates potential relationships between various student majors and the network features describing those nodes. We account for potential subject-based relationships in the network by creating separate models with fully independent parameters for each course subject area (further

ISSN 1929-7750 (online).

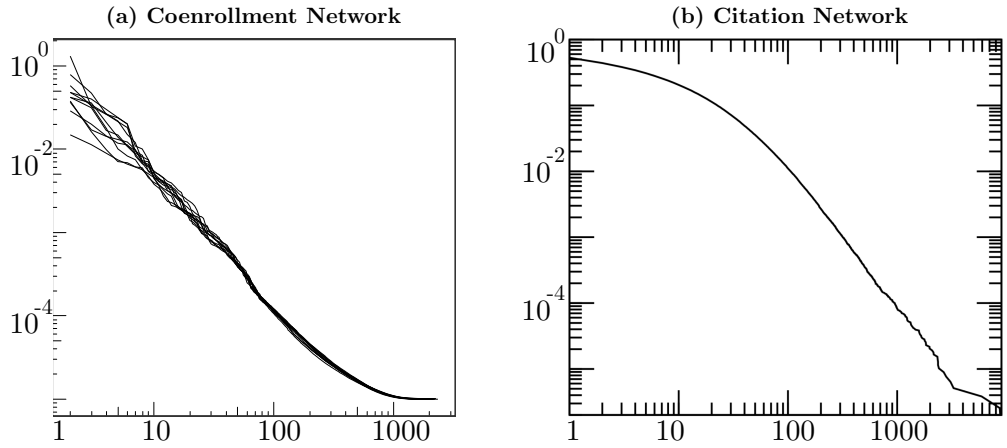

Figure 2. Cumulative degree distribution of (a) co-enrollment networks for the 11 semesters evaluated, representing approximately 22,000 nodes per network, compared to the (b) cumulative degree distribution of a citation network from (M. Newman, 2003; Redner, 1998). The similarity of shape on the log scale demonstrates that both networks conform to power-law degree distributions.

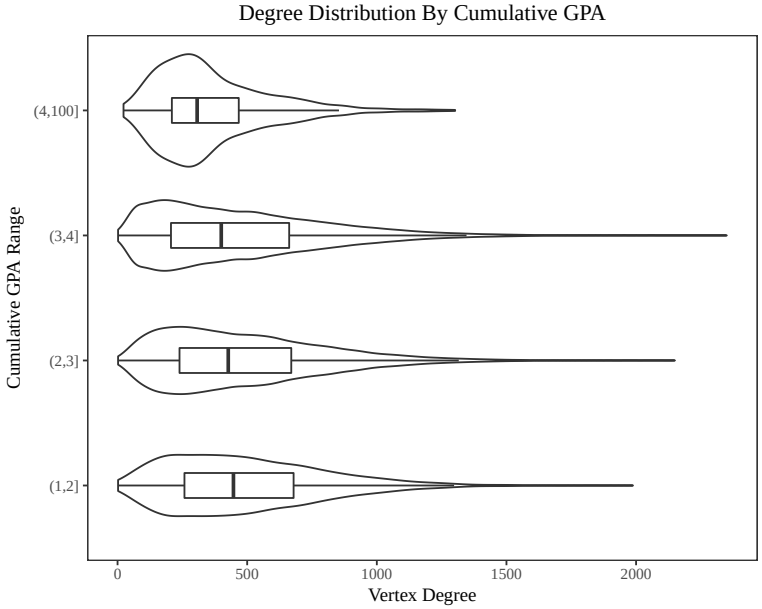

Figure 3. Distribution of students' network degree across various GPAs (using student links). While the plot shows statistically significant differences in the distribution of degree according to Kolmogorov-Smirnov testing, it also suggests that complex models may be required to fully capture relationships between performance and student position in the co-enrollment network.

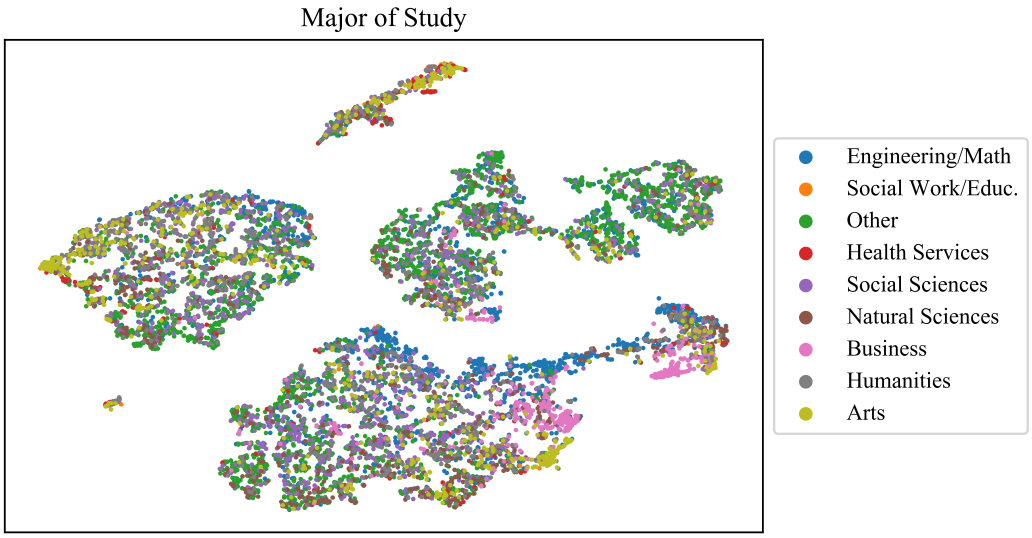

Figure 4. Two-dimensional projection of t-SNE performed on network-based features. These results show potential patterns in network-based features according to major of study (and potentially other aspects of network-based features), but also suggest that the relationships are complex and subtle.

detail in Section 5). Future work should explore more refined clustering within the co-enrollment network; such an analysis is beyond the scope of this paper.

4.0.3 Assortativity and Explanatory Power of Co-enrollment Network Features

This experiment is particularly concerned with uncovering and modelling relationships between the co-enrollment network and student performance. As such, we also examined the *assortativity* of each co-enrollment network. Assortativity measures the tendency for nodes in networks to be connected to other nodes that are like (or unlike) them in some way (M. E. J. Newman, 2003). We calculate the assortativity coefficient as

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

as described in (M. E. J. Newman, 2003) 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 a student's GPA and the GPA of students they are connected to in the co-enrollment network. Essentially, the assortativity coefficient measures whether "birds of a feather flock together," "opposites attract," or neither (with respect to academic performance) in the co-enrollment 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 co-enrollment network at a major U.S. university, revealing at scale a result that has previously been only suggested by smaller-scale co-enrollment network analysis (Gašević et al., 2013). Furthermore, this result suggests that there are performance-based network dynamics across the student co-enrollment network that an effective predictive model might capture and use for grade prediction.

Having demonstrated some graphical properties of the network that 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: 1) a model with only the network-based features described above; 2) a model with only previous cumulative GPA; 3) a model with both network features and GPA. Each model is a simple ordinary least squares linear

Table 3. Assortativity coefficients (M. E. J. Newman, 2003) 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 co-enrollment 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 demonstrate 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	0.073	0.224	0.269	185.677	*
W2006	0.089	0.221	0.277	239.578	*
W2007	0.087	0.232	0.274	181.471	*
W2008	0.087	0.218	0.27	235.175	*
W2009	0.079	0.228	0.268	181.421	*
W2010	0.09	0.255	0.285	138.792	*
W2011	0.091	0.251	0.283	155.24	*
W2012	0.074	0.209	0.254	212.028	*
W2013	0.078	0.212	0.257	219.821	*
W2014	0.074	0.204	0.258	261.856	*
ALL TRAIN	0.079	0.226	0.266	1839.833	*
W2015	0.074	0.203	0.252	242.673	*
ALL	**0.079**	**0.224**	**0.265**	**2035.109**	*

regression model with only 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 a student's co-enrolled peers, motivating the more complex modelling 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 (Getoor, 2005; Getoor & Mihalkova, 2011; Lu & Getoor, 2003a, 2003b). Then, we test alternative model specifications, including a single "flat" model that trains a single discriminative classifier on the union of all object and link features, and a multi-view "blended" ensemble.

ISSN 1929-7750 (online).

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 student-course-link model uses both types of links, following (Getoor & Diehl, 2005). Temporal-link and student-link models are identical in structure to the student-link model, but use LA_{temp}, LA_{cr} in place of LA_{st}, respectively.

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
$\hat{C}(X) = argmax_{c \in C} \frac{P(c\|OA(\mathbb{O})) \cdot \prod_{t \in cr,st} P(c\|LA_t(X))}{P(c)}$

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 *co-enrollment 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 co-enrollment 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 5 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: 1) random forests allowed us to consider the wide object attribute feature space without having to perform potentially expensive feature selection or manage multicollinearity; 2) random forests can capture complex interactions between variables, while in a logistic regression these interactions would need to be manually specified; 3) random forests still capture the benefits of discriminative classifiers (Getoor & Diehl, 2005); and 4) prior research suggested that random forests are effective for network-based modelling (Van Assche, Vens, Blockeel, & Dzeroski, 2004).

We then fit a network-based model, the co-enrollment model, to capture the dependence between nodes using the proxy labelling 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 neighbours, which is known at the time of prediction (this is the *proxy labelling*), not from the true labels (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 modelling approach thus uses two types of models – a student (flat) model, and a co-enrollment (network) model – to predict student grades. Note in Table 5 that the structural modelling approach with student-link features utilizes *both* course and student links separately, following previous formulations of these models (Getoor & Diehl, 2005). 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, and produced 221 unique subject-level models for each specification examined, which required training a total of more than 2,600 unique models.

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 model's predictions of each individual outcome class in order to produce a single, more

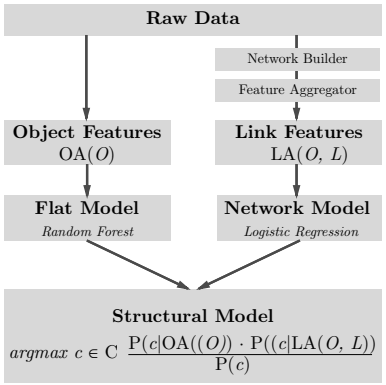

Figure 5. Structural model-building procedure. Separate flat (student) and network (co-enrollment) models are constructed from disjoint feature sets. The models are combined to generate a structural model using the predicted probabilities for each class label *C* and the class prior probability.

accurate final prediction (Dietterich, 2000; Sun, 2013). We implemented a model "blending" approach (Bennett, Lanning, & Others, 2007; Töscher, Jahrer, & Bell, 2009), which is an alternative to the more common stacked generalization technique (Wolpert, 1992). 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 12 network models – four network builders $\times$ three feature aggregators). For the ensembles, 75% of training data was used for training the base classifiers, and 25% was reserved for the probe set, which base models make predictions on (but are not trained on). We use eXtreme Gradient Boosting classifiers, or XGBoost (Chen & Guestrin, 2016), for the meta-learner classifier. XGboost is a common, highly flexible classifier that is often used for ensembling. While a neural network was used in the original blending experiment (Bennett et al., 2007; Töscher et al., 2009), initial experiments with this setup suggested that the neural network architecture failed to provide a substantial boost in performance (Gardner & Brooks, 2018), despite requiring extensive hyperparameter tuning and substantially more training time (approximately two orders of magnitude) for each iteration than the XGBoost model. The results of the ensemble model are shown in Table 6.

This specific ensembling approach, in which base models are fit using disjoint sets of conceptually grouped features and the results ensembled into a single model, is referred to as "multi-view learning" because the base 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 (Sun, 2013; C. Xu, Tao, & Xu, 2013). Multi-view learning has been used in learning analytics for MOOC dropout prediction (F. Jiang & Li, 2017; Li et al., 2016) as well as for other tasks, such as image recognition (where the term "view" applies more naturally). Other boosting algorithms (Adaboost) have been adapted elsewhere successfully for multi-view ensemble learning (Z. Xu & Sun, 2010). 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; we demonstrate such an analysis in Figure 6.

5.3 Results

The results of our prediction experiment are shown in Table 6. We find that the multi-view ensemble outperforms all other model specifications across all performance metrics considered (Multiclass AUC (Hand & Till, 2001), accuracy, Fleiss' Kappa and average multiclass sensitivity, specificity), demonstrating how sophisticated models may 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.

The more sophisticated structural models, which utilize both student- *and* course-network builders and replicates the original structural model (Getoor & Diehl, 2005), performs below the flat model, and even below a baseline of simply predicting grades using student GPA. This finding, that multiple link types do not improve performance when combined in a structural model as specified in Table 5, is counter to previous findings (Getoor & Diehl, 2005). 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 modelling originally observed in (Getoor & Diehl, 2005).

Such a conclusion is supported by the superior performance of the ensemble. The ensemble, which takes the predictions

ISSN 1929-7750 (online).

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 modelling 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 (Hand & Till, 2001) for details.

Model (Link type, feature type)	Multiclass AUC	Accuracy	Kappa	Sensitivity	Specificity
Multi-view Ensemble	0.729	0.657	0.339	NA	0.918
Structural (Temporal, Binary-Link)	0.713	0.642	0.325	0.398	0.833
Structural (Course, Mean-Link)	0.713	0.644	0.324	0.395	0.833
Structural (Temporal, Mean-Link)	0.712	0.646	0.324	0.393	0.832
Structural (Course, Binary-Link)	0.712	0.641	0.323	0.397	0.833
Structural (Temporal, Count-Link)	0.711	0.645	0.319	0.391	0.831
Structural (Temporal, Proportion-Link)	0.711	0.646	0.321	0.39	0.831
Structural (Student, Mean-Link)	0.71	0.646	0.319	0.388	0.831
Structural (Student, Binary-Link)	0.71	0.643	0.325	0.397	0.833
Structural (Course, Proportion-Link)	0.709	0.642	0.318	0.393	0.831
Structural (Student, Count-Link)	0.709	0.645	0.315	0.388	0.829
Structural (Student, Proportion-Link)	0.709	0.644	0.314	0.386	0.829
(Course, Count-Link)	0.707	0.639	0.313	0.39	0.83
Flat	0.695	0.651	0.304	0.358	0.825
Full	0.693	0.649	0.297	0.351	0.824
Predict GPA Baseline	0.684	0.523	0.212	0.355	0.812
Structural (Course + Student, Mean-Link)	0.672	0.64	0.284	0.342	0.821
Structural (Course + Student, Binary-Link)	0.671	0.638	0.291	0.351	0.824
Structural (Course + Student, Count-Link)	0.666	0.637	0.276	0.343	0.819
Structural (Course + Student, Proportion-Link)	0.664	0.637	0.274	0.339	0.818
Majority Class Baseline	0.5	0.571	0	0.25	0.75

of each (flat + twelve network models) for each potential outcome class 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 ensemble utilizes 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, or a slight decrease, relative to the flat model with only student features.

The results of the different network-builders and feature aggregators were relatively similar. This suggests that each network formulation, with any of the four methods for aggregating features across this network, capture relevant information about students' neighbours in the graph, and that the additional data provided by a model that combines multiple network-builders is overwhelmed by the additional assumption of independence between the two link types shown in the formulation in Table 5. In particular, it is relevant that mean-link models generally performed quite well, because the network component of these models contained only a single feature (the mean cumulative GPA of a given student's neighbours in the graph).

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 modelling 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.

As a note for practical implementation of the ensemble model, we observe that training this model requires considerably more computation than an individual structural model, as it requires first training 13 models – 12 structural models, plus a flat model – on the base classifier training set (75% of the full training data), and then training a meta-learner using the procedure described in Section 5.2. However, because the training of each of the base structural models is fully independent of each other base model, their training can be parallelized. Hence, the training time of the ensemble is reduced only to the time required to train the slowest of the 12 base structural models, plus the time required to train a meta-learner on the probe dataset (which is considerably faster, containing only 25% of the initial training data).

ISSN 1929-7750 (online).

5.4 Feature Analysis

While the predictive performance of models is important, particularly if these models are intended to be deployed as "early warning systems," interpretability of such models is equally valuable. A clear understanding of the most predictive features in each model not only stands to inform us about the relationships between various features and student academic performance. Feature analysis also provides another perspective on understanding the complex structure of the co-enrollment network, and how student attributes and the network structure are related to student performance. In this section, we present an initial feature analysis of four different models: The full model (which uses all *OA* and *LA* features), the flat model (which uses only *LA* features), the network model from the temporal binary-link model, and the multi-view ensemble.

The 20 highest-importance features for each model are shown in Figure 6; importance for the network model (which is a multinomial logistic regression) is measured using the absolute value of the Wald z-score of each coefficient in the model as recommended in (Hilbe, 2009); all other importances are measured by the average improvement in Gini impurity across all trees in the model. Note that high importance does not imply any specific direction of association between a given feature and student performance (high importance implies only that a feature is useful in discriminating between the various outcome classes, given the other variables in the model). Indeed, due to the flexible and complex models models used in this experiment, each feature likely has a nonlinear and complex relationship with the outcome of interest.

Several potentially interesting insights are demonstrated by the feature analysis above; we encourage the interested reader to conduct a detailed inspection, but provide some initial analysis here. First, the prior cumulative GPA, when available, is generally the strongest predictor of students' future performance. Second, network features, when available, are generally preferred over features besides cumulative GPA in the "full" model, which uses both flat and network features; many different network features (including student and temporal network builders, and mean-, proportion-, and count-link aggregators) show higher importance than the second-most important link attribute (high school GPA). Third, we see that the ensemble largely relies on the predictions of the "flat" model, assigning low, relatively well-dispersed importance to the predictions of the remaining models. This shows that while these structural models add useful information that improve the performance over a simple flat model, the yield is low and is generated by combining information across many models.

In interpreting the feature analysis results shown in Figure 6, we note that the reader should keep in mind that many of the student-level demographic attributes (e.g., gender, ethnicity, birth year) are based only on self-reports of these attributes, most often collected at the time of a student's application for admission to the university. Encodings of these attributes are also limited to only a single value from a predetermined list, determined by the institution. As such, these attributes can only be considered proxies for students' true identities along these various dimensions, which may be different, fluid, and more complex than the available data indicate. There are, of course, important ethical considerations that need to be taken into account when such features are used in practice or used for decision making. In particular, we note the need to ensure that models based on demographic attributes are fair, transparent, and beneficial to students.

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 co-enrollment networks, at scale, over 10 years of university records and over 1 million individual records. Our analysis demonstrates that university co-enrollment 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 co-enrollment graph. Additionally, we 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 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 co-enrollment networks constructed with student-level links. Structural models with student- and course-level links, which follow the original implementation (Getoor, 2005), perform worse than flat models, likely due to their additional assumptions of independence between the student- and course-level structural models. We build a multi-view blended ensemble of the full set of structural and flat models, which achieves 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. We also show several relevant results related to the feature importance within various models, including demonstrating that this ensemble largely relies on the "flat" model predictions.

These results suggest that network-based features can improve predictive models of student grades, but that structural models that make strong assumptions about independence between object and link attributes may not realize these performance gains.

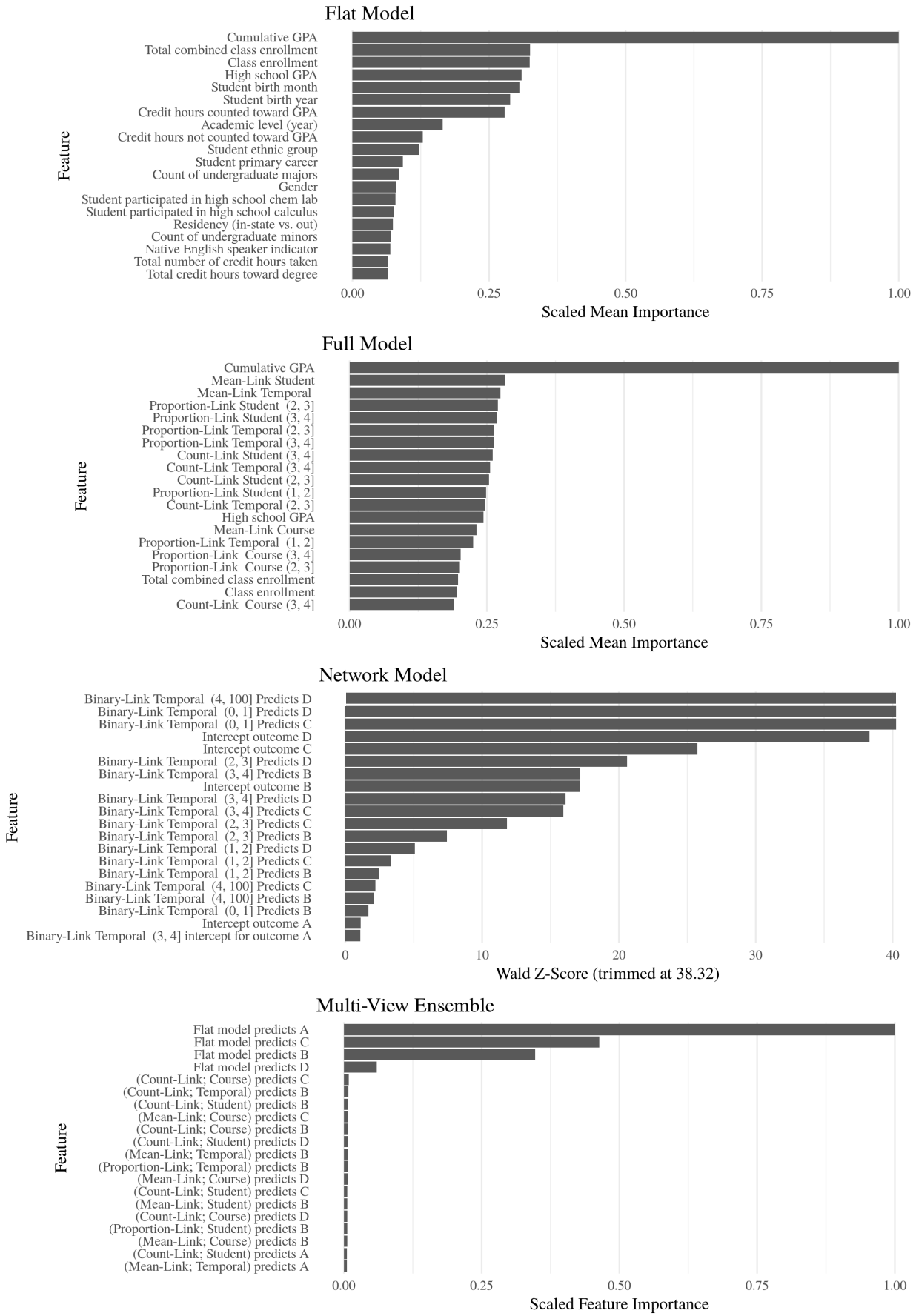

Figure 6. Feature importance metrics for various models evaluated. The network model shown is a model constructed using the binary-link network builder and a proportion-link feature aggregator. Note that the x-axis for the network model is trimmed at 38.32, the z-scores for the top three features exceeded 10^7.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial -NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

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 co-enrollment-based predictive models, as well as broader exploration of other modelling techniques. Modelling techniques of interest for future work include other discriminative classifiers (e.g., SVM) for creating the base student and co-enrollment models, and other ensembling techniques (e.g., blend optimization (Töscher et al., 2009)). Additionally, further research into the feature space used to build the co-enrollment/network model is required, and should explore different and novel methods for building link features: considering various sizes of neighbourhood with different link types; more or less granular bucketing for the count- and binary-link features; using different proxy labelling techniques; building temporal feature sets that extend back over multiple semesters or co-enrollment periods; considering prerequisite restrictions from a course graph (as not all students could be co-enrolled with other students); using the predicted grades of neighbours as "bootstrap" estimates and building a more traditional within-network model; and generating more "views" by further partitioning the original input data. 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 co-enrollment networks. Further investigation into the potential mechanisms through which co-enrollment influences student performance – if such an effect indeed exists – and how different types of co-enrollment relationships (course- and student-link) differentially impact performance will provide a stronger theoretical foundation for future predictive modelling efforts.

Further research evaluating similar models using other datasets, such as data collected from other universities with different student populations and compositions, would also assist in evaluating the generalizability of both the overall student modelling techniques proposed here, as well as the specific findings of our modelling and feature analysis evaluations.

Finally, future research should investigate and demonstrate the actionable insights supported by such models, and how they can support real-time decision making 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 (Lynch, Barnes, Xue, & Gitinabard, 2017) for their feedback on an initial version of this work.

8. Funding

This work was funded by the Michigan Institute for Data Science (MIDAS) under the Holistic Modeling of Education (HOME) grant.

9. Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Adams, P. (2006, October). Exploring social constructivism: Theories and practicalities. *Education 3-13*, *34*(3), 243–257. https://dx.doi.org/10.1080/03004270600898893

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

Baker, S., Mayer, A., & Puller, S. L. (2011). Do more diverse environments increase the diversity of subsequent interaction? evidence from random dorm assignment. *Econ. Lett.*, *110*(2), 110–112. https://dx.doi.org/10.1016/j.econlet.2010.09.010

Baldwin, T. T., Bedell, M. D., & Johnson, J. L. (1997). The social fabric of a team-based MBA program: Network effects on student satisfaction and performance. *Acad. Manage. J.*, *40*(6), 1369–1397. https://dx.doi.org/10.5465/257037

Becchetti, L., Castillo, C., Donato, D., & others. (2008). Link analysis for web spam detection. *ACM Transactions on*. https://dx.doi.org/10.1145/1326561.1326563

Bennett, J., Lanning, S., & Others. (2007). The netflix prize. In *Proceedings of KDD cup and workshop* (Vol. 2007, p. 35). brettb.net.

Bhagat, S., Rozenbaum, I., & Cormode, G. (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* (pp. 92–101). New York, NY, USA: ACM. https://dx.doi.org/10.1007/978-3-642-00528-2$_6$

Biancani, S., & McFarland, D. (2013). Social networks research in higher education. *Educ. Stud.*(4), 85–126.

Brunello, G., de Paola, M., & Scoppa, V. (2010, July). PEER EFFECTS IN HIGHER EDUCATION: DOES THE FIELD OF STUDY MATTER? *Econ. Inq.*, *48*(3), 621–634. https://dx.doi.org/10.1111/j.1465-7295.2009.00235.x

Chakrabarti, S., Dom, B., & Indyk, P. (1998). Enhanced hypertext categorization using hyperlinks. In *Proceedings of the 1998 ACM SIGMOD international conference on management of data* (pp. 307–318). New York, NY, USA: ACM. https://dx.doi.org/10.1145/276305.276332

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). New York, NY, USA: ACM. https://dx.doi.org/10.1145/2939672.2939785

Cortes, C., Pregibon, D., & Volinsky, C. (2001, September). Communities of interest. In *Advances in intelligent data analysis* (pp. 105–114). Springer, Berlin, Heidelberg. https://dx.doi.org/10.1007/3-540-44816-0$_1$1

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer, Berlin, Heidelberg.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 57–66). New York, NY, USA: ACM. https://dx.doi.org/10.1145/502512.502525

Fawcett, T., & Provost, F. (1997, September). Adaptive fraud detection. *Data Min. Knowl. Discov.*, *1*(3), 291–316. https://dx.doi.org/10.1023/A:1009700419189

Foster, G. (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), 1455–1475. https://dx.doi.org/10.1016/j.jpubeco.2005.12.001

Friedman, N., Getoor, L., Koller, D., & Pfeffer, A. (1999). Learning probabilistic relational models. In *IJCAI* (Vol. 99, pp. 1300–1309). robotics.stanford.edu.

Gardner, J., & Brooks, C. (2018). Coenrollment networks and their relationship to grades in undergraduate education. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 295–304). New York, NY, USA: ACM. https://dx.doi.org/10.1145/3170358.3170373

Gašević, D., Zouaq, A., & Janzen, R. (2013). "choose your classmates, your GPA is at stake!" the association of cross-class social ties and academic performance. *Am. Behav. Sci.*, *57*(10), 1460–1479. https://dx.doi.org/10.1177/0002764213479362

Getoor, L. (2005). Link-based classification. In *Advanced methods for knowledge discovery from complex data* (pp. 189–207). Springer London.

Getoor, L., & Diehl, C. P. (2005, December). Link mining: A survey. *SIGKDD Explor. Newsl.*, *7*(2), 3–12.

Getoor, L., & Mihalkova, L. (2011). Learning statistical models from relational data. In *Proceedings of the 2011 ACM SIGMOD international conference on management of data* (pp. 1195–1198). New York, NY, USA: ACM. https://dx.doi.org/10.1145/1989323.1989451

Hadwin, A. F., & Järvelä, S. (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), 235–239.

Hand, D. J., & Till, R. J. (2001, November). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.*, *45*(2), 171–186. https://dx.doi.org/10.1023/A:1010920819831

Hilbe, J. M. (2009). *Logistic regression models*. CRC Press.

Hill, S., Provost, F., & Volinsky, C. (2006). Network-Based marketing: Identifying likely adopters via consumer networks. *Stat. Sci.*, *21*(2), 256–276. https://dx.doi.org/10.1214/088342306000000222

Hoel, J., Parker, J., & Rivenburg, J. (2005). Peer effects: do first-year classmates, roommates, and dormmates affect students' academic success. In *Higher education data sharing consortium winter conference, santa fe, NM*. Citeseer.

Jiang, F., & Li, W. (2017). Who will be the next to drop out? anticipating dropouts in MOOCs with Multi-View features. *International Journal of Performability Engineering*, *13*(2).

Jiang, Y., & Golab, L. (2016). On competition for undergraduate co-op placements: A graph mining approach. In *Proceedings of the 9th international conference on educational data mining* (pp. 394–399).

Johnson, D. W., & Johnson, R. T. (2009, June). An educational psychology success story: Social interdependence theory and cooperative learning. *Educ. Res.*, *38*(5), 365–379. https://dx.doi.org/10.3102/0013189X09339057

Koller, D. (1999, June). Probabilistic relational models. In *Inductive logic programming* (pp. 3–13). Springer, Berlin, Heidelberg. https://dx.doi.org/10.1007/3-540-48751-4$_1$

Kossinets, G., & Watts, D. J. (2006, January). Empirical analysis of an evolving social network. *Science*, *311*(5757), 88–90. https://dx.doi.org/10.1126/science.1116869

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th international conference on machine learning* (pp. 282–289).

Lee, C., Scherngell, T., & Barber, M. J. (2011). Investigating an online social network using spatial interaction models. *Soc. Networks*, *33*(2), 129–133. https://dx.doi.org/10.1016/j.socnet.2010.11.002

Li, W., Gao, M., Li, H., Xiong, Q., Wen, J., & Wu, Z. (2016, July). Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In *2016 international joint conference on neural networks (IJCNN)* (pp. 3130–3137). ieeexplore.ieee.org. https://dx.doi.org/10.1109/IJCNN.2016.7727598

Lu, Q., & Getoor, L. (2003a). Link-based classification. In *ICML* (Vol. 3, pp. 496–503). vvvvw.aaai.org.

Lu, Q., & Getoor, L. (2003b). 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.*

Lyle, D. S. (2007, April). Estimating and interpreting peer and role model effects from randomly assigned social groups at west point. *Rev. Econ. Stat.*, *89*(2), 289–299. https://dx.doi.org/10.1162/rest.89.2.289

Lynch, C. F., Barnes, T., Xue, L., & Gitinabard, N. (2017). Graph-based educational data mining (G-EDM 2017). In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th international conference on educational data mining* (pp. 472–473).

McEwan, P. J., & Soderberg, K. A. (2006, May). Roommate effects on grades: Evidence from First-Year housing assignments. *Res. High. Educ.*, *47*(3), 347–370. https://dx.doi.org/10.1007/s11162-005-9392-2

Newman, M. (2003, January). The structure and function of complex networks. *SIAM Rev.*, *45*(2), 167–256. https://dx.doi.org/10.1137/S003614450342480

Newman, M. E. (2001, January). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U. S. A.*, *98*(2), 404–409. https://dx.doi.org/10.1073/pnas.98.2.404

Newman, M. E. J. (2003, February). Mixing patterns in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, *67*(2 Pt 2), 026126. https://dx.doi.org/10.1103/PhysRevE.67.026126

Oh, H.-J., Myaeng, S. H., & Lee, M.-H. (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* (pp. 264–271). New York, NY, USA: ACM. https://dx.doi.org/10.1145/345508.345594

Pandit, S., Chau, D. H., Wang, S., & Faloutsos, C. (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* (pp. 201–210). New York, NY, USA: ACM. https://dx.doi.org/10.1145/1242572.1242600

Provost, F., Dalessandro, B., Hook, R., Zhang, X., & Murray, A. (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* (pp. 707–716). New York, NY, USA: ACM.

Redner, S. (1998, July). How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B*, *4*(2), 131–134. https://dx.doi.org/10.1007/s100510050359

Sacerdote, B. (2000, January). *Peer effects with random assignment: Results for dartmouth roommates* (No. 7469). https://dx.doi.org/10.1162/00335530151144131

Schrire, S. (2006, January). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Comput. Educ.*, *46*(1), 49–70. https://dx.doi.org/10.1016/j.compedu.2005.04.006

Siegfried, J. J., & Gleason, M. A. (2006). Academic roommate peer effects. *Unpublished manuscript, Vanderbilt Univ. , Nashville*.

Stinebrickner, R., & Stinebrickner, T. R. (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), 1435–1454. https://dx.doi.org/10.1016/j.jpubeco.2006.03.002

Sun, S. (2013, December). A survey of multi-view machine learning. *Neural Comput. Appl.*, *23*(7-8), 2031–2038. https://dx.doi.org/10.1007/s00521-013-1362-6

Tang, L., & Liu, H. (2011, November). Leveraging social media networks for classification. *Data Min. Knowl. Discov.*, *23*(3), 447–478. https://dx.doi.org/10.1007/s10618-010-0210-x

Taskar, B., Abbeel, P., & Koller, D. (2002). Discriminative probabilistic models for relational data. In *Proceedings of the eighteenth conference on uncertainty in artificial intelligence* (pp. 485–492). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Töscher, A., Jahrer, M., & Bell, R. M. (2009). The bigchaos solution to the netflix grand prize. *Netflix prize documentation*, 1–52.

Traud, A., Kelsic, E., Mucha, P., & Porter, M. (2011, January). Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.*, *53*(3), 526–543. https://dx.doi.org/10.1137/080734315

Van Assche, A., Vens, C., Blockeel, H., & Dzeroski, S. (2004). A random forest approach to relational learning. In *Workshop on statistical relational learning.*

Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017, February). Delving deeper into MOOC student dropout prediction.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on facebook. *Am. J. Sociol.*, *116*(2), 583–642. https://dx.doi.org/10.1086/653658

Winston, G., & Zimmerman, D. (2004). Peer effects in higher education. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 395–424). University of Chicago Press.

Wolpert, D. H. (1992). Stacked generalization. *Neural Netw.*, *5*(2), 241–259. https://dx.doi.org/10.1016/S0893-6080(05)80023-1

Xu, C., Tao, D., & Xu, C. (2013, April). A survey on multi-view learning.

Xu, Z., & Sun, S. (2010, November). An algorithm on Multi-View adaboost. In *Neural information processing. theory and algorithms* (pp. 355–362). Springer, Berlin, Heidelberg. https://dx.doi.org/10.1007/978-3-642-17537-4$_4$4

Zhang, T., Popescul, A., & Dom, B. (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* (pp. 821–826). New York, NY, USA: ACM.

Zimmerman, D. J. (2003, February). Peer effects in academic outcomes: Evidence from a natural experiment. *Rev. Econ. Stat.*, *85*(1), 9–23. https://dx.doi.org/10.1162/003465303762687677

ISSN 1929-7750 (online).