Translating network position into performance: Importance of Centrality in Different Network Configurations

Srećko Joksimović
Moray House School of Education
University of Edinburgh
Edinburgh, UK
s.joksimovic@ed.ac.uk

Areti Manataki
School of Informatics
University of Edinburgh
Edinburgh, UK
a.manataki@ed.ac.uk

Dragan Gašević
Schools of Education and Informatics
University of Edinburgh
Edinburgh, UK
dragan.gasevic@ed.ac.uk

Shane Dawson
Learning and Teaching Unit
University of South Australia
Adelaide, Australia
shane.dawson@unisa.edu.au

Vitomir Kovanović
School of Informatics
University of Edinburgh
Edinburgh, UK
v.kovanovic@ed.ac.uk

Inés Friss de Kereki
Engineering School
Universidad ORT Uruguay
Montevideo, Uruguay
kereki_i@ort.edu.uy

ABSTRACT

As the field of learning analytics continues to mature, there is a corresponding evolution and sophistication of the associated analytical methods and techniques. In this regard social network analysis (SNA) has emerged as one of the cornerstones of learning analytics methodologies. However, despite the noted importance of social networks for facilitating the learning process, it remains unclear how and to what extent such network measures are associated with specific learning outcomes. Motivated by Simmel's theory of social interactions and building on the argument that social centrality does not always imply benefits, this study aimed to further contribute to the understanding of the association between students' social centrality and their academic performance. The study reveals that learning analytics research drawing on SNA should incorporate both – descriptive and statistical methods to provide a more comprehensive and holistic understanding of a students' network position. In so doing researchers can undertake more nuanced and contextually salient inferences about learning in network settings. Specifically, we show how differences in the factors framing students' interactions within two instances of a MOOC affect the association between the three social network centrality measures (i.e., degree, closeness, and betweenness) and the final course outcome.

Categories and Subject Descriptors
Education; K.3.1 [Computer Uses in Education] Distance learning

LAK '16, April 25-29, 2016, Edinburgh, United Kingdom

DOI: http://dx.doi.org/10.1145/2883851.2883928

General Terms
Social Processes, Learning

Keywords
Social network analysis, ERGM, MOOC, Academic achievement

1. INTRODUCTION

Social network analysis (SNA) has been one of the most commonly applied methods in learning analytics research [1, 2]. Network approaches can extend analyses beyond the individual level to focus on group dynamics. As such, SNA can provide insight into the quantity and types of interactions or relationships that occur between participants, groups and communities in conventional as well as online settings [1, 3, 4]. Recently, with the development of social networking sites that allow for a relatively straightforward extraction of social networks, the application of SNA in education has significantly increased [1, 5, 6]. However, despite the volume of SNA applied within education research, few studies have fully realized the potential of network analyses to provide new insights into our understanding of learning [3].

Although SNA provides a rich set of tools and methods that help improve the understanding of learning in social networks [3, 7], the majority of the studies utilizing SNA in education are primarily based on examining structural regularities underlying student interactions [4, 8]. Researchers mainly rely on network structural properties (e.g., centrality and density) [9, 10] or generative processes (e.g., triad closure), usually observed in isolation [8], to describe emerging patterns of students' engagement. For example, by examining measures of centrality, embeddedness or triadic closure in social networks, researchers can reveal who is interacting with whom and what is the strength of interactions, the actors occupying more central or peripheral positions in the network, and how such network engagement patterns can affect learning [3, 4, 10, 11]. Although with limited generalizability, such analyses are of great importance in uncovering weak and strong ties that bridge communities/groups of students, revealing the most influential actors or individuals that may have a more advantageous position [12, 13].

The major characteristic of the descriptive models used in the traditional application of SNA in (online) education has focused on *describing* relationships between observed variables, rather than explaining *why* such structure exists [8]. Although models for descriptive analysis help explain the association between network variables and identify potentially relevant processes in the network structure, they do not allow for the generalization of findings across the networks. The lack of inferential power that characterizes these mathematical, descriptive models (e.g., measuring centrality or density) is indirectly depicted through the interpretation of the association between learning outcome and measures of students' social centrality. Despite the prevailing, and largely unchallenged, understanding that occupying a higher social centrality leads to a higher academic performance [3, 9, 10], research findings are inconclusive about which centrality measure (or combination of measures) is the most significant predictor of academic achievement. Additionally, several recent studies have revealed somewhat contradictory results, indicating that the predictive power of social centrality measures highly depends on the context that frames students' interactions [11, 14].

A potential rationale for explaining the inconsistencies in the educational research may lie in the lack of accountability for the network context that frames social interactions [15, 16]. Research and practice in learning analytics commonly relies on general models (i.e., context independent) in order to inform learning and teaching processes, predict learning outcomes or provide appropriate scaffolds [15]. However, without considering specific learning settings, those models could lead to incomplete conclusions. Likewise, applying SNA without accounting for the processes that guide network formation and consideration of the quantity and quality of interactions could also result in a model that does not reliably capture the underlying social processes [8]. Thus, in order to provide for more valid inferences and identify the determinants that explain regularities of network formation, a sound theoretical approach driving the choice of the analytics methods is required. In so doing, the theory driven approach can help explain the underlying network structure and provide the context for the interpretation of revealed social processes.

1.1 SNA and MOOC research

The emergence of Massive Open Online Courses (MOOCs) has provided new opportunities for the application of SNA among researchers and practitioners interested in studying networked learning [17, 18]. Given the high numbers of students enrolling into MOOCs [19] and the immense amount of data related to students' participation and interaction collected by MOOC platforms, it has become even more challenging to understand patterns that drive learning in such networked settings. Therefore, studies investigating MOOCs have relied on SNA methods in order to visualize and examine regularities in interactions emerging from social learning activities that students and teachers engage with [20, 21], as well as to investigate the association between centrality in social networks and student performance [11, 14], to name a few. However, this research while valuable, still fails to adequately account for both context and the structural properties of the established networks.

To address this deficit the present study incorporates both theory related to the importance of "super-strong" ties [16, 22] in network development as well as the statistical methods for generalizing network inference, i.e., Exponential Random Graph Models (ERGMs) [23]. The study analyses two separate instances of the same MOOC offered in different languages during the same period of time. In so doing, the study aims to provide further evidence for the importance of accounting for the contextually salient determinants that define network formation when studying social networks. In the following, we compared two social networks, emerging from student discussions, with respect to the statistical properties that define underlying network structures [23]. We utilized statistical network analysis (i.e., ERGMs specifically), rather than mathematical (descriptive) methods, as it is a more comprehensive approach to explaining uncertainty inherent in the observed data and determining which of the network processes present significant factors that frame the network evolution [4, 8, 23]. Finally, following the differences in the regularities framing the social relations within the two networks analyzed, we examined the association between social centrality measures (i.e., degree, closeness, and betweenness) and the academic performance (i.e., obtained certificate – *none*, *normal*, *distinct*), within the different contexts.

2. BACKGROUND

2.1 Social Network Analysis in Educational Research

The initial application of SNA dates back to the 1930s involving a Harvard study that analyzed interpersonal relations and the formation of cliques [24]. The concept of social centrality was first introduced in the 1940s, with a significant uptake noted in the 1950s and the 1960s [9, 24]. Nevertheless, from these early studies it appeared that while the researchers at the time agreed that *centrality is an important structural property of social networks*, there was a lack of consensus regarding what centrality means and how it should be measured [9]. In his seminal work, Freeman (1979) revisited the concept of centrality and identified three network structural properties that should be considered as a measure of centrality – degree, closeness, and betweenness. In formal online courses, SNA studies have aimed at revealing whether and how those structural properties, as defined by Freeman (1979) and others, are associated with learning. However, different studies have often produced contradicting results. For example, Russo and Koesten [25] showed that network *prestige* (in-degree) and *centrality* (out-degree) significantly predict cognitive learning outcomes. Cho and colleagues [26] also concluded that network centrality measures were significantly and positively associated with a students' final grade. However, results from Cho and colleagues [26] also revealed that only closeness centrality was a significant predictor of the course grade. The association between grades and the other two centrality measures – i.e., degree and betweenness centrality - was not statistically significant. Gašević and colleagues [27] also observed a significant association between grade point average (GPA) and two measures of network centrality (eccentricity and closeness centrality) in a fully online master of science in information systems program. However, similar to the Cho et al's [26] study, Gašević and colleagues [27] also failed to find a significant association between GPA and degree and betweenness centrality. Thus, without detailed contextual information it becomes challenging to conclude which of the centrality measures are considered important predictors of a student's overall academic achievement. More simply put, the absence of context limits our understanding of how network position influences student learning.

Research in MOOCs further argues for the necessity to account for various contextual factors when interpreting SNA in networked learning settings. Specifically, contemporary research shows that the association between student centrality in MOOC discussion forums and academic performance, depends on the context of the course [11, 14]. For example, Jiang and colleagues [14], analyzed the association between degree, betweenness and closeness centrality and student grades within two MOOCs in Algebra and Financial Planning. While the results indicated a significant and positive association between the final course grade and two centrality measures (degree and betweenness) for the Algebra MOOC, none of the measures were significantly correlated with the student grades for the Financial Planning MOOC. Further, the approach applied in the study by Dowell and colleagues [11] differs from the traditional application of SNA in MOOCs. More precisely, Dowell et al. [11] aimed at predicting two different achievement measures– final course grade and social centrality – using linguistic properties of student generated content. Results showed that the linguistic characteristics positively associated with social centrality were negatively associated with the final course grade, and vice versa. Although Dowell and colleagues [11] did not directly compare social centrality and course grades, their findings indicate that these two measures of learning tend to capture different achievement metrics, suggesting further that "the skills associated with these two learning-related outcomes differ" (p.7, *ibid.*).

This review of the existing literature, suggests that future research should provide additional insight into the contextual factors that may impact on the association between students' position in the network and their learning outcomes. Instead of focusing solely on the network structural properties to describe patterns of students' engagement within MOOC discussion forums, we aim to utilize statistical network analysis to provide contextual information about the processes that stimulate the underlying network formation. Particularly, we aim to reveal important regularities in interaction structure among the course participants that could provide a valid context for the interpretation of network structural properties. It should be noted that contextual factors are not necessarily related to the course design and instructional conditions. Here, we observe context in terms of the factors that frame individuals' social behavior. According to Simmel [28] the nature of interaction between the two individuals in a social network is derived from the collective behavior, which accounts for the general social situation that goes beyond the two focal parties.

2.2 Simmelian Ties Theory

In addition to the direct measures of the network structural properties, SNA research should also consider the contextual factors that influence the development of the network. The most influential research in SNA argues that those individuals who occupy more central roles (primarily focusing on betweenness centrality) will have higher potential to benefit from such positions and attain their goals [9, 13, 29]. Thus, in his seminal work, Granovetter [13] argued that weak ties are those that enable more straightforward access to information disseminated through a social network. Burt [12] goes even further arguing that the strength of ties is not as relevant as the fact that a given tie bridges otherwise distinct groups or cliques in the social network. As Burt noted "[p]eople whose networks bridge the structural holes between groups have an advantage in detecting and developing rewarding opportunities" [30, p. 354]. Both theories are in line with Freeman's [9] definition of centrality and assume that the more central persons in a social network occupy a more advantageous position. Nevertheless, Krackhardt [16] posits that centrality does not necessarily imply less constraints and more benefit. If a node is linked in what Krackhardt [16] calls a "Simmelian tie", such a position could impose additional limitations. In the context of the present study, this could suggest that while a student centrally positioned in the network has a high potential for control over the information flow, the actual realized gains for their learning may be diminished. Therefore, as Krackhardt [16] posits, traditional SNA analysis (in his case traditional role analysis) should be supported with Simmelian Ties analysis. In the present study, we argue that Simmelian Ties Theory [28] presents a sound theoretical framework in providing valid context for interpreting the importance of social centrality for the academic achievement.

Simmel's theory of social behavior focuses on studying relationships that occur between people in order to explain their actions [16, 28]. Simmel argued that context is the primary factor influencing what people do and why they behave in a particular manner. Context is determined "by the set of third others who also engage in various relationships with the two focal parties" [31, p. 16]. Thus, as Simmel argued, the establishment of such triadic nodes should be the fundamental unit of analysis in order to understand social behavior [16, 28]. Triads are considered to be qualitatively different from the dyadic relationships that Burt [12] and Granovetter [13], among others, focus on [16, 22]. This difference originates in the nature of the formed relationships. The two nodes forming a dyad are more independent and retain more individuality in their relationship [16, 22]. For instance, should disagreement occur in a dyad, both parties can choose to cease any further interaction. However, a triadic tie requires a higher level of negotiation. If a member of a group disagrees and ceases further interaction the group remains to exist and a connection remains. Thus, Krackhardt [22] described Simmelian ties as "super-strong" (p.24), ties that "qualitatively add durability and power" (p.24, *ibid.*), beyond the strong ties as previously defined by Granovetter [13] and Krackhardt [32].

Simmelian ties theory differs from psychological theories, such as Heider's [33] balance theory, in explaining structural properties for the existence of symmetric and transitive triples, that are considered main processes in social networks [16]. According to Heider's [33] theory, people are motivated to establish and maintain relationships that would allow them to keep comfortable communicating with others. The Simmelian theory, on the other hand, assumes that once cliques are formed, they resist changing, becoming strong and stable, thus decreasing propensity to dissolve over time [28]. However, "there is no inherent motivation to form a clique" [31, p. 21], it is rather the social structure, or the context, that causes formation of certain network structures [28].

Building further on one of Krackhardt's [22] conclusions (i.e., that *traditional* SNA should be supported with Simmelian ties analysis), and given the theorized relationship between the social centrality and the expected benefits, it seems reasonable to analyze whether networks under study exhibit properties of Simmelian ties. In the educational context, such strong ties could indicate the existence of tightly connected groups, focused around common interests.

2.3 Exponential random graph models in Online Learning

A majority of studies applying SNA in online and distance education relies on mathematical models to describe relationships between observed variables [34]. Such studies are particularly useful in revealing important network characteristics or what processes should be observed within the social network [8]. For example, using descriptive models we would be able to determine whether Simmelian ties exist in a given network. However, in order to reveal whether these processes (i.e., propensity to form "super-strong" ties) occur more often than expected if ties were generated randomly, as well as what other micro-level processes (e.g., popularity, propensity for triad closure) determine social dynamics in a given network, we need to rely on statistical models [8]. The quadratic assignment procedure for analyzing dyadic data sets [35], Exponential Random Graph Models (ERGM) and stochastic blockmodels for the cross-sectional social network analysis and community detection [23, 36], as well as longitudinal models for studying evolution of networks and behavior [37] are some of the commonly proposed methods. ERGM specification allows us to model Simmelian statistics (i.e., a process of formation of "super-strong" ties). Hence, this approach is directly applicable for exploring hypothetical network processes that could explain the evolution of the observed cross-sectional network [8, 23].

As a generalization of *p1* models and Markov graphs [38], exponential random graph models for social networks, also known as *p** models, were introduced by Frank and Strauss [39] and Wasserman and Pattison [40]. ERGMs belong to the family of probability models for network analysis that allow for more generalizable inferences over the structural foundations of social behavioral patterns [23, 38]. Observing network ties as random variables, ERGMs allow for modeling overall network structure through a set of local network processes [38]. ERGMs assume that each tie within these local network processes (e.g., mutuality, transitivity or triad closure) is conditionally dependent, indicating further that "empirical network ties do not form at random, but that they *self-organize* into various patterns arising from underlying social processes" [41, p. 3]. Although ERGMs, and similar statistical methods (e.g., longitudinal probabilistic social network analysis – [4]), have been successfully applied in social sciences [42], medical research [43] and studying traditional education [8], their application in the context of online learning and MOOCs is rather sparse.

From the perspective of the analytical methods applied and the educational context analyzed, Kellogg et al.'s [5] study is perhaps the most relevant for our research. In their mixed methods study, Kellogg and colleagues [5] aimed at providing more comprehensive understanding of the dynamic processes that underlie peer support learning in MOOCs tailored towards educators in K-12 settings. The quantitative part of the study included application of SNA tools and techniques – descriptive network measures and ERGMs – in the analysis of the two interaction networks obtained from discussion forums. In order to examine mechanisms of peer support in the two MOOCs, Kellogg and colleagues [5] analyzed various patterns of selective mixing and network statistics: reciprocity, homophily by professional role (e.g., principal), gender, educational background, grade levels, differences in experience (i.e., heterophily), and three proximity mechanisms based on the state or country, geographical region, and group assignment. The results indicate a strong and significant reciprocity effect, suggesting that students are more likely to reply to a peer when there has been prior evidence of reciprocity. Nevertheless, homophily and heterophily effects, as well as proximity mechanisms differed across the networks analyzed.

2.4 Research questions

The education literature suggests that researchers predominantly rely on descriptive methods when applying SNA in online learning settings. There is far less evidence of the research accounting for network specific variables that could provide contextual background for the interpretation of the underlying processes. Given the inconsistencies in findings on the association between social centrality and learning outcome, we aimed at determining whether network social dynamics have an impact on the predictive power of network structural position. We were particularly interested to find out whether a network formed around an online course is characterized by the propensity to form Simmelian ties. We hypothesized that these "super-strong" relationships could influence the potential benefits students derive from occupying more central positions in the network. Thus, we defined the following two research questions:

RQ1. *Are there differences in the underlying processes that determine network formation within social networks formed in various online learning settings?*

RQ2. *Is the propensity for forming Simmelian ties significantly different than expected if ties were formed randomly?*

Eventual differences in the social dynamics that frame social interactions within the two networks analyzed would provide a valid context for the interpretation of the possible variances in the predictive power of the social centrality measures. Therefore, we defined our third research question as follows:

RQ3. *If there are differences in regularities that frame network structure among the course participants, how do these discrepancies affect the association between social centrality and academic performance?*

3. METHOD

3.1 Data

This study analyzed forum discussions within two instances of a single course that were delivered on the Coursera platform in Spring 2015. The two instances, Code Yourself![1] (CDY) and ¡A Programar![2] (APR), were designed to be identical with respect to the content and teaching methods, with the only difference being the delivery language, i.e., English in CDY and Spanish in APR.

The MOOC aimed to introduce young teenagers to computer programming, while covering the basic topics in computational thinking and software engineering. The content of this 5-week course consisted of lecture videos, quizzes and peer-assessed programming projects, which were translated and tailored for English and Spanish-speaking audiences. A common marking scheme was established, whereby students were deemed to have successfully completed the course (and obtained a certificate) when they had a score of at least 50% for the coursework. A distinction was awarded for students receiving a score of 75% or more. CDY and APR were designed to be identical not only in content, but also with respect to their simultaneous delivery with the MOOCs running from March-April 2015. This implies that all aspects of the MOOCs were equivalent including weekly course announcements and matching instructor-initiated prompts in the discussion forums, and adopting a common strategy for minimal instructor intervention in the forums.

[1] https://www.coursera.org/learn/codeyourself

[2] https://www.coursera.org/learn/a-programar

Despite the common approach for the two course instances, student engagement and performance was considerably different in CDY and APR. As shown in Table 1, almost 60,000 students enrolled in CDY and more than 25,000 in APR. However, almost the same number of students completed the two courses – 1,597 in CDY and 1,595 in APR. Moreover, regardless the smaller student cohort (in overall), higher number of students engaged with the forum discussions in the APR course, resulting in a more intensive forum activity produced (Table 1).

Table 1. Descriptive statistics for the number of enrolled students, students engaged with the course content and discussion forum, as well as the obtained certificates

	CDY	APR
Enrolled	59,531	25,255
Engaged	**26,568**	13,808
Engaged with forum	1,430	**1,818**
Posted messages		
Threads	776 (1.69; 1.75)	1,081 (3.53; 5.12)
Posts	4,204 (3.13; 7.75)	5,940 (3.53; 5.12)
Comments	1,981 (3.42; 9.06)	2,686 (3.21; 6.75)
Total	**5,177**	**7,409**
Obtained certificate		
Normal	586	644
Distinct	1,011	951
Total	**1,597**	**1,595**

Note: Thread, Posts and Comments rows display counts in the following format – total (average; SD)

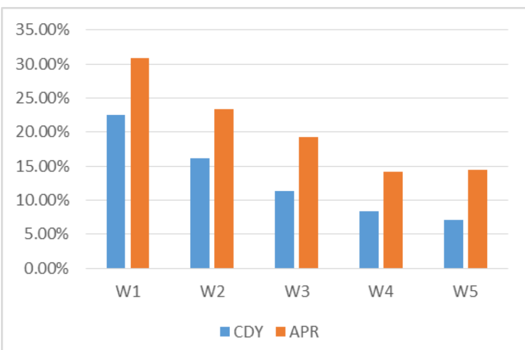

Figure 1. Proportion of students that watched a lecture each week

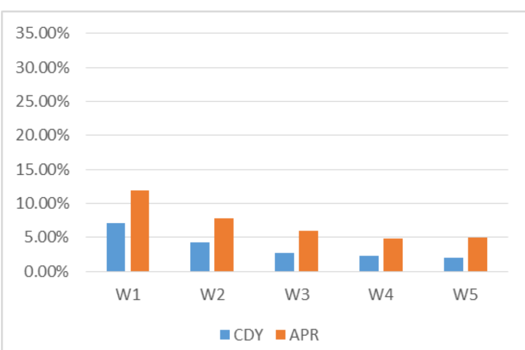

Figure 2. Proportion of students browsing forums each week

Large differences were also observed with respect to student engagement with the course materials. The proportion of students that visited the course, watched a lecture, submitted an exercise or browsed the forums each week in CDY was always smaller than the corresponding proportion for APR that week. As depicted in Figure 1 and Figure 2, in some cases this difference reached levels of about 8%. It is also worth mentioning that the weekly engagement steadily dropped in CDY during the 5-week duration. In contrast for APR there was a steady drop during the first 4 weeks, followed by an increase in engagement for the final week.

3.2 Analysis

3.2.1 Social Network Analysis

To address the first two research questions, we extracted two directed weighted graphs to represent interactions occurring within discussion forums for the two course instances (CDY and APR). Although several approaches have been proposed for extracting social networks from discussion forums, we relied on the most commonly applied approach that considers each message as being directed to the previous one [11, 44]. For example, if author A2 replied to a message posted by author A1, we would add a directed edge A2->A1. Further, if A3 posted a comment on A2's post, we would include A3->A2 edge as well. Finally, social graph included all the students who posted to the discussion forum.

Social network analysis was conducted through two complementary phases; statistical network analysis and structural (i.e., *traditional*) network analysis. The **statistical network analysis** was performed using ERGMs in order to reveal various networks statistics and examine processes that guided network formation for both of the courses instances. Relying on commonly used network statistics [4, 5, 8] we examined network formation mechanisms at the two levels; dyadic and triadic. At the ***dyadic level***, we aimed to investigate the effects of **selective mixing**, **reciprocity**, **popularity**, and **expansiveness**. *Selective mixing* reflects a students' propensity to interact with their peers based on the combination of their individual characteristics [8, 23]. Thus, we considered a homophily effect with respect to the following students' attributes:

- Achievement: none, normal, and distinct;
- Domestic: a student was from either the United Kingdom or Uruguay (as the course was offered by two universities from these two countries) or was from an alternate country;,
- Gender: male, female;
- Access group: student, instructor, or teaching staff.

Reciprocity, on the other hand, is a network statistic that models students' tendency to form mutual ties and cluster together [23]. In the case of our study, this property would allow for revealing whether students tend to continue interaction with their peers who replied to their posts. Finally, *popularity* and *expansiveness* tend to model processes that would indicate the existence of students who receive a significant number of replies to their posts or students who tend to reply more often to their peers' posts, respectively.

At the ***triadic level***, we examined effects of **triadic closure** and **Simmelian ties** formation. Existing research argues that cyclic and transitive triples are the common characteristics of networks emerging from social media [45]. However, with directed networks, these two statistics are captured within the *triangle* term [8, 23]. Nevertheless, models with *triangle* term are almost always degenerate [23], therefore, *geometrically weighted edgewise shared partner distribution (gwesp)* is used instead. We also modeled Simmelian ties [32] in order to examine whether the network(s) analyzed conform to the Simmelian ties theory. That is, whether the networks exhibit a formation of cliques of students that tend to interact with each other significantly more often than with the rest of their peers. Such a statistic could indicate that those students are primarily being focused on their field of interest and rarely interacting with other students.

The analysis of **network structural properties** relied on most commonly used SNA measures that capture various aspects of graph structural centrality – degree, closeness, and betweenness centrality [9, 10, 34]. Degree centrality is considered the most straightforward centrality measure, focusing on the local structure surrounding the node and indicating the number of connections (ties) a node has in the network [9]. It is commonly interpreted as a measure of *popularity* [34] or the extent to which observed node has a "potential for activity in communication" [9, p. 219]. Given that our focus was on the analysis of weighted networks, we relied on the weighted degree centrality, that accounts for the weight of edges a node has in the network [46]. Closeness centrality measures a distance of a given node to all other nodes in the network [9]. Closeness centrality measures nodes' potential to connect easily with other nodes. Finally, betweenness centrality is perhaps the most significant for the context of our study, given Krackhardt's [16] view on the association between the strength of the ties and expected benefits for the nodes that bridge two distinct parts of the network.

We consider three models, for each of the networks, based on the described set of statistics – a demographic attribute model (DM) that includes only processes based on students' characteristics; triadic closure and Simmelian ties model (TSM), including only *gwesp* and *simmelian* statistics; and a full model that combines the two (FM). Comparing likelihood-based measure of AICc, we further continued selecting the most parsimonious model, which would provide the best fit to our data. The social networks were analyzed using the *ergm 3.1.2* [47], an R package for statistical network analysis, and using igraph 0.7.1 [7], a comprehensive R software package for complex social network analysis research.

3.2.2 Regression Analysis

To examine the association between the dependent variable (i.e., obtained certificate), and the independent variables (i.e., three centrality measures), we adopted multinomial logistic regression (MLR) analysis [48], in order to answer our third research question. MLR is predictive analysis that is used to explain the association between a nominal dependent variable that has more than two levels (*none*, *normal*, and *distinct*), and one or more continuous independent variables [48]. It does not make any assumptions of normality, linearity and homogeneity of variance for the independent variables [48].

Aiming to observe the association between the three centrality measures – degree, closeness, and betweenness centrality – and the course outcome, we build three MLR models. Each model included one dependent (obtained certificate) and one independent variable (degree, closeness, or betweenness centrality). The analyses were performed using the *mlogit 0.2-4* package for R that enables estimation of multinomial logit models [49].

4. RESULTS AND DISCUSSION

4.1 Network Characteristics

Descriptive statistics (Table 2) indicate rather diverse processes within the two networks analyzed. Given the difference in the number of nodes (Table 2) it is expected that the APR network would have a considerably higher number of edges, and perhaps moderately higher weighted degree. However, higher modularity, average clustering coefficient and higher number of connected components, could indicate a less cohesive group of students within the CDY instance of the course [1]. Moreover, descriptive statistics also indicate a comparable number of reciprocal ties, whereas the number of "super-strong" ties is considerably higher in case of the English version of the course.

Table 2. Descriptive statistics for social networks extracted from CDY and APR discussion forums

Descriptives	**CDY**	**APR**
Edges	3,620.00	4,736.00
Avg. W. Degree	4.00	4.69
Density	0.002	0.001
Modularity	0.45	0.33
Conn. comp.	16.00	9.00
Avg. clust. coef.	0.12	0.09
Reciprocity	231.00	176.00
Simmelian	41.00	7.00
Simmelian ties	144.00	32.00
Popularity	758.55	839.00
Expansiveness	1373.42	1612.53

In case of both networks under the study, the full model provided the best fit, indicated by the lowest value for AICc (CDY: DM – 2,830,818.00, STM – 49,863.82, FM – 48,371.14, and APR: DM – 4,577,956.00, STM- 67,786.65, FM – 66,921.94). Estimated coefficients are presented in Table 3, whereas goodness-of-fit statistics indicate that models provide a satisfactory fit for the data. It is also important to note that we aimed at assessing homophily at the level of access groups (i.e., students, teachers, teaching staff) and triad closure (*gwesp*) (Section 3.2.1). However, those two statistics indicated an overall worse fit to our data than the selected (i.e., best fit) model; therefore, both statistics were excluded from the final models analyzed.

Table 3. Analysis of the estimates for the two ERG models – CDY FM and APR FM

	CDY		**APR**	
	Estimate	SE	Estimate	SE
Baseline (Edges)	**-5.45*****	**0.04**	**-5.81*****	**0.09**
Selective mixing				
Distinct	**0.98*****	**0.03**	**0.47*****	**0.12**
None	**0.15*****	**0.03**	**-0.20****	**0.08**
Normal	**0.60*****	**0.17**	**0.68****	**0.25**
Domestic	**-0.95*****	**0.03**	-0.09	0.07
Gender	0.02	0.03	-	-
Structural mechanisms				
Reciprocity	**3.81*****	**0.09**	**4.20*****	**0.55**
Simmelian	**4.89*****	**0.61**	-	-
Popularity	**-3.68*****	**0.10**	**-4.75*****	0.29
Expansiveness	-	-	-0.25	0.21

Note: * p < .05. ** p < .01. *** p < .001.

It is revealing that differential homophily for the final course outcome (i.e., obtained certificate) shows that both networks exhibited a higher likelihood of assortative mixing between the students who obtained the certificate. Similar to Kellogg and colleagues study [5], our results suggest that the more successful students tend to interact more often. However, the likelihood of interaction between the most successful students is higher in the CDY course. Whereas, the same effect holds between the students who did not obtain the certificate in case of the English instance of the course (although with less likelihood), the effect is negative in the Spanish version of the course. Students who did not obtain a certificate in the APR instance of the course were less likely to interact with each other.

Homophily for the students' country of residence, revealed a significant effect for the English instance of the course, whereas the effect was not significant in the Spanish version. Kellogg and colleagues [5] observed a similar effect - i.e., homophily by state or country) and found a significant positive increase in the likelihood that two students from the same state or country will create a tie. In our study, however, we examined selective mixing between *domestic* students. Given that two courses were particularly designed for two diverse groups of students, we aimed at investigating how that aspect would influence students' tendencies to connect with their peers. Our results revealed that students, who are considered "domestic" in the CDY course instance, were less likely to connect with their *domestic peers*. Observing students' demographic data, we could perhaps expect the same effect within both models, given that similar numbers of students (7% in CDY and 10% in APR) were considered domestic in both networks. However, the observed effect was not statistically significant for the Spanish version of the course.

The effect of reciprocity was significant for the models of both networks, indicating that students tended to continue interacting with peers who replied to their posts. Although the estimates seem rather high, those values are in line with results of Lusher, Koskinen, and Robins [50] and Kellogg et al. [5] studies, who also revealed a very strong effect of direct interaction between students. It appears that a strong effect of reciprocity could be seen as one of the defining characteristics of interaction in online social networks in general [50]. Moreover, Lusher and colleagues [50] further identified such networks as "self-disclosing" (p.249) and "bonding" (p.249), characterized by strong ties relations between the nodes. In such networks, students tend to self-disclose themselves, bonding with their peers, creating comfortable environment for knowledge sharing and learning [50]. However, given rather the low cohesion at the network level for both networks (i.e., low density – Table 2), it seems reasonable to conclude that students commonly interact within smaller groups of peer students [24].

The effect of Simmelian ties was not consistent across both the networks. While it was strong and significant for the CDY network, in the case of the APR course we were not able to fit the model with Simmelian statistics. Thus, although the strong effect for reciprocity could indicate existence of strong ties, it seems that the ties within the English version of the course *evolved* to "super-strong" ties, as defined by [16, 22]. The existence of Simmelian ties beyond the chance level is a significant defining characteristic of the social network emerging from the CDY discussion forum. These ties are structurally embedded within relatively small, highly connected and cohesive groups, commonly referred to as communities [45]. Interactions within those communities are more often and qualitatively different from interactions with other peer students. This finding could be further explained by a "rich-club phenomenon" (p.1), an analogy used by Vaquero and Cebrian [7] to explain "frequent and intense" (p.1, *ibid.*) interactions occurring within relatively small groups of students, where students benefit greatly from these structural arrangements.

The effect of *expansiveness* was not significant in the APR social networks. However, we were not able to fit the model to a satisfactory quality using this network statistics in case of the CDY network. On the other hand, the strong negative effect of *popularity* in the CDY network is also in line with Kellogg's [5] study. Kellogg et al. [5] and Lusher and colleagues [50] argue that such an effect could indicate that all the students have a similar level of popularity and that most likely networks were not "centralized on in-degree" [5, p. 275]. Considering the previous results (i.e., the strong effect of reciprocity) this result seems quite intuitive. Moreover, given the fact that we observed interactions within a discussion forum, this effect further contributes to the understanding that students in both networks tended to engage into further interaction with their peers, rather than simply posting a message without the intent to contribute the further discussion.

In addressing the **first and second research question**, we were able to conclude that the observed networks differ with respect to the determinants of network formation. The most notable difference is related to the structure of "super-strong" ties, where CDY network exhibit a formation of cliques formed around students who tend to interact within the strong and stable groups of peers, which "resist change" [31, p. 21]. Although the APR network showed the same regularities with respect to reciprocity of interaction and popularity, the effect of Simmelian ties was not present. Finally, the APR network also revealed higher tendency that students would interact more often with higher performing peers.

4.2 Social centrality and academic achievement

Analyzing the association between the students' centrality and the final learning outcome further revealed differences between the two networks. Specifically, in the case of the CDY course instance, only weighted degree centrality was significantly associated with the course outcome – $\chi^2(1)$ = 9.048, p=.011. However, multinomial regression analysis showed that an increase in weighted degree significantly increased the likelihood of obtaining certificate with distinction, compared to not completing the course successfully, whereas there was no significant difference between *normal* certificate and failing the course (Table 4). On the other hand, closeness and betweenness centrality were not significantly associated with the course outcomes.

Table 4. Results of the multinomial regression analysis of the association between social centrality and the final learning outcome (i.e., obtained certificate)

		Estimate	SE	t
Weighted Degree				
CDY	distinct	**0.008***	**0.004**	**2.720**
	normal	0.007	0.004	1.618
APR	distinct	**0.046*****	**0.006**	**7.318**
	normal	**0.046*****	**0.006**	**7.413**
Closeness				
CDY	distinct	0.002	0.038	0.046
	normal	0.062	0.066	0.934
APR	distinct	**-0.064***	**0.030**	**-2.113**
	normal	**-0.105****	**0.037**	**-2.816**
Betweenness				
CDY	distinct	0.000009	0.000005	1.621
	normal	-0.000003	0.00001	-0.185
APR	distinct	**0.0001*****	**0.00002**	**5.584**
	normal	**0.0001*****	**0.00002**	**5.562**

Note: * p < .05. ** p < .01. *** p < .001; Reference levels for each of the analysis was "none" – i.e., student did not obtain a certificate.

The APR social network revealed different patterns. All of the observed centrality measures were significantly related to the likelihood to obtain a certificate – weighted degree, $\chi^2(1)$ =

90.217, p<.001; closeness, $\chi^2(1) = 9.679$, p=.008, and betweenness, $\chi^2(1) = 59.832$, p<.001. Even more so, an increase in each of the centrality measures significantly increased the likelihood of both – obtaining a certificate with distinction, and a *normal* certificate (Table 4), compared to not completing the course. It should be noted that direction of closeness centrality is opposite to the betweenness and degree centrality – lower values indicate lower distance (i.e., higher closeness) of a given node to all other nodes in the network [10].

There are two important aspects of the findings presented in the previous section. First, we would argue that our results support [16, 22] understanding of the importance of social centrality in providing greater opportunity for well–positioned individuals. Although Krackhardt [16, 22] discusses the potential to bridge between two social groups (i.e., betweenness centrality), we would posit that the importance of the most commonly addressed centrality measures in educational research – degree (to a certain extent), closeness, and betweenness – should be interpreted with respect to the propensity to form Simmelian ties. Following Krackhardt's [16] argument that "occupying a bridging role can be *more* constraining" (p. 184, *ibid.*), our results show that depending on the given context, a higher social centrality does not necessarily imply a better academic performance. In that sense, we could conclude that those students who are occupying positions between strongly connected groups of students might not be able to benefit significantly from their position. Observed from the perspective of roles, as defined by Krackhardt [16], this finding could further indicate that students within the CDY course instance tended to primarily interact with peers who share the same interests, and perhaps have the same or similar level of knowledge. Nevertheless, further research is needed to address this assumption.

The second important finding of our results relates to the development of an interactive "rich-club" [7]. In their analysis of the relationship between the social structure and performance, Vaquero and Cebrian [7] concluded that students tend to interact within the groups of strongly connected peers. Vaquero and Cebrian [7] labeled those groups as a "rich-club", where students engage in interaction with their peers at the very beginning of the course, and tend to remain within the same cliques throughout the course. Vaquero and Cebrian [7] further showed that those persistent interactions are maintained between high performing students, whereas low performing students would usually attempt to join those groups later in the course. However, such attempts would usually fail to produce reciprocity in the interaction with high performing students. Thus, those "rich-clubs" or the groups of strongly connected students could be easily connected with Krackhardt's [16] cliques (i.e., groups of students connected with "super-strong", Simmelian ties).

From the analysis of the two social networks it would appear that interaction within the CDY discussion forum tended to follow the social structure as noted in Vaquero and Cebrian's [7] study. This could imply that students within the APR course instance were more socially inclusive, and supportive of their peers who may have joined late in the discussions. On the other hand, it could also mean that the majority of students in the APR course instance were simply engaged in the discussions from the very beginning of the course. Both of these possible interpretations require further research to more comprehensively explain the reasons for the observed differences in social interactions within two different networks of students (i.e., student in CDY and APR course). Nevertheless, it should be noted that we do not assume that those students who attained a more central position in a social graph are necessarily low performing students.

With respect to the **third research question**, our results support the assumption that social centrality in networks that are formed around strongly connected components (i.e., "rich-club" or Simmelian groups, as with the CDY network) is not associated with the final course outcome. Whereas, on the other hand, with *more relaxed* interactions (i.e., the APR network), however still assuming a high level of reciprocity in social ties, social centrality is significantly and positively associated with the course outcome (i.e., obtained certificate). Finally, it should be noted that weighted degree centrality diverges from this pattern to a certain extent (Table 4).

5. CONCLUSIONS & IMPLICATIONS

This study investigated the importance of the context that defines students' social interactions for the association between structural centrality and learning outcome. Primarily, we grounded the theoretical framework in Simmel's theory of social interactions and Krackhardt's [16] argument that the "quality of tie itself interacts with the bridging role to produce more constraint on the unsuspecting actor" (p.184), to define network specific properties that would allow us to make more valid inferences. Finally, supplementing descriptive SNA with statistical network analysis and multinomial logistic regression, we were able to conclude that social centrality within the network characterized with "super-strong" ties, does not necessarily imply benefits. On the other hand, structural centrality in the network with reciprocal ties, where all participants have similar level of popularity, yet without a significant effect of "super-strong" ties, is positively associated with the likelihood of obtaining a certificate at the end of the course.

Analyzing roles in an organization, Krackhardt [16] concluded that "traditional role analysis on raw network relations" (p. 208), should be supplemented with the Simmelian ties analysis, arguing further that such an analysis provides "more insight into organizational phenomena" (p.208). Our study extends Krackhardt's [16] argument in two directions. Primarily, we argue that *any* traditional SNA (not just role analysis), should be supported with the Simmelian ties analysis, as those ties are qualitatively different from weak and strong ties as defined by Granovetter [13], and therefore provide a more comprehensive understanding of social interactions and the dynamics influencing the overall network. Moreover, as a consequence of this theoretical recommendation, it is reasonable to argue that traditional (primarily descriptive) approaches to the analysis of social interactions should be supported by statistical network analysis. Relying solely on mathematical approaches we are able to identify the most significant patterns in the established social interactions. However, in order to understand which of the identified patterns significantly determine network structure and occur beyond the chance, more profound (statistical) models are required [8, 23, 47].

Through the statistical network analysis methods, we were able to provide context to interpret an association between social centrality and academic achievement. Again we refer to the previous work by Krackhardt [16, 22, 31] to explain how Simmelian ties could affect one's position within an organization. Krackhardt [16] identified those "super-strong" ties as "more enduring, more visible, and more critical than sole-symmetric ties" (p.208), that is, ties that "constrain and influence" (*ibid.*).

One of the imposed connotations of our findings, for both research and practice domains, is the necessity to account for contextual information when interpreting the potential gains implied by the network structural properties. For example, revealing and visualizing network structure using deeply embedded relations (i.e., Simmelian backbones) [45] could significantly improve the quality of information presented in social learning analytics dashboards, such as the one presented in the work by Schreurs and colleagues [20]. Moreover, providing additional information about the social dynamics should supplement any feedback based on the measures of structural centrality. Likewise, research on predicting association between descriptive network measures and products of learning, in educational settings, should be constructed on valid theoretical assumptions that could support conclusions about inferred social dynamics.

Further research should also integrate temporal dynamics to investigate how certain network processes evolve over time. A promising approach in that direction would be application of Temporal ERGMs [51], or similar models, for studying time-evolving social networks. Moreover, as indicated by Edwards [42] and Kellogg and colleagues [5], as well as in our previous work [11], [52], SNA should be integrated with content analysis to account for the quality of students' contribution. Finally, it should be noted that 39% of CDY students who submitted the survey, stated that English was their first language. On the other hand, 97% of student who participated in APR course and submitted the survey chose Spanish as their first language. However, we were not able to include this information in the model, since majority of students who participated in the course did not submit the survey. This also reflected to the students who participated in the discussion forum. Nevertheless, investigating whether language, as a predominate medium for communication between students in a computer-mediated learning environment [52], influences development of the underlying social processes, presents a promising venue for future research.

Several limitations of our study need to be acknowledged. We analyzed students' interactions within discussion forum in two instances of a same MOOC. Although we relied on a most commonly accepted method for network construction, this approach tends to underestimate the intensity of all the interactions within the given settings. Moreover, analysis of interactions in a more informal settings, such as connectivist MOOC [53], would also contribute to the greater generalizability of our findings. Finally, data from different subject domains (e.g., social science) should be analyzed in order to account for diverse learning settings.

6. REFERENCES

[1] K. Cela, M. Sicilia, and S. Sánchez, "Social Network Analysis in E-Learning Environments: A Preliminary Systematic Review," *Educ. Psychol. Rev.*, vol. 27, no. 1, pp. 219–246, 2015.

[2] S. Dawson, D. Gašević, G. Siemens, and S. Joksimovic, "Current state and future trends: A citation network analysis of the learning analytics field," in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 2014, pp. 231–240.

[3] B. V. Carolan, *Social Network Analysis Education: Theory, Methods & Applications. Social Network Analysis Education: Theory, Methods & Applications. SAGE Publications, Inc.* SAGE Publications, Inc., 2014.

[4] K. Stepanyan, K. Borau, and C. Ullrich, "A Social Network Analysis Perspective on Student Interaction within the Twitter Microblogging Environment," in *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*, 2010, pp. 70–72.

[5] S. Kellogg, S. Booth, and K. Oliver, "A social network perspective on peer supported learning in MOOCs for educators," *Int. Rev. Res. Open Distrib. Learn.*, vol. 15, no. 5, 2014.

[6] D. McFarland, D. Diehl, and C. Rawlings, "Methodological Transactionalism and the Sociology of Education," in *Frontiers in Sociology of Education*, vol. 1, M. T. Hallinan, Ed. Springer Netherlands, 2011, pp. 87–109.

[7] L. M. Vaquero and M. Cebrian, "The rich club phenomenon in the classroom," *Sci Rep*, vol. 3, Jan. 2013.

[8] S. Goodreau, J. Kitts, and M. Morris, "Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks*," *Demography*, vol. 46, no. 1, pp. 103–125, 2009.

[9] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.

[10] S. Wasserman, *Social network analysis: Methods and applications*, vol. 8. Cambridge university press, 1994.

[11] N. Dowell, O. Skrypnyk, S. Joksimović, A. C. Graesser, S. Dawson, D. Gašević, P. de Vries, T. Hennis, and V. Kovanović, "Modeling Learners' Social Centrality and Performance through Language and Discourse," presented at the In Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain, 2015.

[12] R. S. Burt, *STRUCTURAL HOLES*. Harvard University Press, 1995.

[13] M. S. Granovetter, "The strength of weak ties," *Am. J. Sociol.*, pp. 1360–1380, 1973.

[14] S. Jiang, S. M. Fitzhugh, and M. Warschauer, "Social Positioning and Performance in MOOCs," in *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, 2014, vol. 1183, p. 14.

[15] D. Gašević, S. Dawson, T. Rogers, and D. Gašević, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet High. Educ.*, vol. 28, pp. 68 – 84, 2016.

[16] D. Krackhardt, "The Ties that Torture: Simmelian Tie Analysis in Organizations," *Res. Sociol. Organ.*, vol. 16, pp. 183–210, 1999.

[17] D. Gašević, V. Kovanović, S. Joksimović, and G. Siemens, "Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative," *Int. Rev. Res. Open Distrib. Learn.*, vol. 15, no. 5, 2014.

[18] M. De Laat and F. Prinsen, "Social learning analytics: Navigating the changing settings of higher education.," *J. Res. Pract. Assess.*, vol. 9, no. 4, pp. 51–60, 2014.

[19] K. Jordan, "Synthesising MOOC completion rates | MoocMoocher," 2015. [Online]. Available: https://moocmoocher.wordpress.com/2013/02/13/synthesising-mooc-completion-rates/. [Accessed: 23-Aug-2015].

[20] B. Schreurs, C. Teplovs, R. Ferguson, M. de Laat, and S. Buckingham Shum, "Visualizing Social Learning Ties by Type and Topic: Rationale and Concept Demonstrator," in *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, New York, NY, USA, 2013, pp. 33–37.

[21] O. Skrypnyk, S. Joksimović, V. Kovanović, D. Gašević, and S. Dawson, "Roles of course facilitators, learners, and technology in the flow of information of a cMOOC," *Int. Rev. Res. Online Distance Learn.*, vol. (in press), 2015.

[22] D. Krackhardt, "Super Strong and Sticky," *Power Influ. Organ.*, p. 21, 1998.

[23] M. Morris, M. S. Handcock, and D. R. Hunter, "Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects," *J. Stat. Softw.*, vol. 24, no. 4, 2008.

[24] J. Scott, *Social Network Analysis*. SAGE Publications, 2012.

[25] T. C. Russo and J. Koesten, "Prestige, centrality, and learning: A social network analysis of an online class," *Commun. Educ.*, vol. 54, no. 3, pp. 254–261, 2005.

[26] H. Cho, G. Gay, B. Davidson, and A. Ingraffea, "Social networks, communication styles, and learning performance in a CSCL community," *Comput. Educ.*, vol. 49, no. 2, pp. 309–329, Sep. 2007.

[27] D. Gašević, A. Zouaq, and R. Janzen, "'Choose Your Classmates, Your GPA Is at Stake!': The Association of Cross-Class Social Ties and Academic Performance," *Am. Behav. Sci.*, 2013.

[28] G. Simmel, *The Sociology of Georg Simmel*. Simon and Schuster, 1950, 1950.

[29] D. J. Brass, "Being in the right place: A structural analysis of individual influence in an organization," *Adm. Sci. Q.*, pp. 518–539, 1984.

[30] R. S. Burt, "Structural Holes and Good Ideas," *Am. J. Sociol.*, vol. 110, no. 2, pp. 349–399, 2004.

[31] D. Krackhardt and M. Handcock, "Heider vs Simmel: Emergent Features in Dynamic Structures," in *Statistical Network Analysis: Models, Issues, and New Directions*, vol. 4503, E. Airoldi, D. Blei, S. Fienberg, A. Goldenberg, E. Xing, and A. Zheng, Eds. Springer Berlin Heidelberg, 2007, pp. 14–27.

[32] D. Krackhardt, "The strength of strong ties: The importance of philos in organizations," *Netw. Organ. Struct. Form Action*, vol. 216, p. 239, 1992.

[33] F. Heider, *The Psychology of Interpersonal Relations*. Taylor & Francis, 1958.

[34] P. J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*, vol. 28. Cambridge university press, 2005.

[35] W. Simpson and others, "The quadratic assignment procedure (QAP)," in *North American Stata Users' Group Meetings 2001*, 2001.

[36] C. DuBois, C. Butts, and P. Smyth, "Stochastic blockmodeling of relational event dynamics," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 2013, pp. 238–246.

[37] T. A. Snijders, "Models for longitudinal network data," *Models Methods Soc. Netw. Anal.*, vol. 1, pp. 215–247, 2005.

[38] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Soc. Netw.*, vol. 29, no. 2, pp. 173 – 191, 2007.

[39] O. Frank and D. Strauss, "Markov graphs," *J. Am. Stat. Assoc.*, vol. 81, no. 395, pp. 832–842, 1986.

[40] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[41] P. Wang, G. Robins, P. Pattison, and E. Lazega, "Exponential random graph models for multilevel networks," *Soc. Netw.*, vol. 35, no. 1, pp. 96–115, 2013.

[42] G. Edwards, "Mixed-method approaches to social network analysis," *Natl. Cent. Res. Methods*, p. 30, 2010.

[43] S. L. Simpson, S. Hayasaka, and P. J. Laurienti, "Exponential Random Graph Modeling for Complex Brain Networks," *PLoS ONE*, vol. 6, no. 5, p. e20039, 2011.

[44] M. de Laat, V. Lally, L. Lipponen, and R.-J. Simons, "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis," *Int. J. Comput.-Support. Collab. Learn.*, vol. 2, no. 1, pp. 87–103, 2007.

[45] B. Nick, C.-K. Lee, P. Cunningham, and U. Brandes, "Simmelian backbones: amplifying hidden homophily in facebook networks," in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, 2013, pp. 525–532.

[46] M. E. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E*, vol. 64, no. 1, p. 016132, 2001.

[47] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, Morris, and Martina, "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *J. Stat. Softw.*, vol. 24, no. 3, pp. 1–29, 2008.

[48] J. S. Cramer, "The standard multinomial logit model," in *Logit Models from Economics and Other Fields*, Cambridge University Press, 2003, pp. 104–125.

[49] Y. Croissant, *mlogit: multinomial logit model*. 2013.

[50] D. Lusher, J. Koskinen, and G. Robins, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press, 2012.

[51] S. Hanneke, W. Fu, and E. P. Xing, "Discrete temporal models of social networks," *Electron. J. Stat.*, vol. 4, pp. 585–605, 2010.

[52] S. Joksimović, N. Dowell, O. Skrypnyk, V. Kovanović, D. Gašević, S. Dawson, and A. C. Graesser, "How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs," presented at the The 5th International Learning Analytics and Knowledge (LAK) Conference (Accepted), 2015.

[53] G. Siemens, "Connectivism: A learning theory for the digital age," *Int. J. Instr. Technol. Distance Learn.*, vol. 2, no. 1, pp. 3–10, 2005.