Evaluating the Complementarity of Communication Tools for Learning Platforms

Leonardo Carvalho[1]**, Laura Assis**[2]**, Leonardo Lima**[3]**, Eduardo Bezerra**[4]**,**
Gustavo Guedes[5]**, Artur Ziviani**[6]**, Fabio Porto**[7]**, Rafael Barbastefano**[8]**, and**
Eduardo Ogasawara[9]

[1]**CEFET/RJ**
[2]**CEFET/RJ**
[3]**CEFET/RJ**
[4]**CEFET/RJ**
[5]**CEFET/RJ**
[6]**LNCC**
[7]**LNCC**
[8]**CEFET/RJ**
[9]**CEFET/RJ**

Corresponding author:
Eduardo Ogasawara[1]

Email address: eogasawara@ieee.org

ABSTRACT

Due to the constant innovations in communications tools, several educational institutions are continually evaluating the adoption of new communication tools (NCT) for their adopted learning platforms (LP). Notably, many educational institutions are interested in checking if NCT is bringing benefits in their teaching and learning process. We can state an important problem that tackles this interest as for how to identify when NCT is providing a significantly different complementary communication flow concerning the current communication tools (CCT) provided at LP. This paper presents the Mixed Graph Framework (MGF) to address the problem of measuring the complementarity of an NCT in the scenario where some CCT is already established. Since we are interested in the methodological process, we evaluated MGF using synthetic data. Our experiments observed that the MGF was able to identify whether an NCT produces significant changes in the overall communications of an LP according to some centrality measures.

1 INTRODUCTION

Communication tools are in constant evolution. They usually change the way people collaborate with each other. Not long ago, letters, telegrams, and other written communications on paper were the mainstream. However, since the beginning of the Internet, communication tools were extended through e-mail. The use of e-mail is widespread and almost ubiquitous in educational institutions, being responsible for the majority of the communication flow inside them Hansen et al. (2010).

Innovations in communication tools continue to occur, and several new features, such as instant messaging, blogs, and content management have been developed. Recently, new opportunities to empower communication among students-teachers have arisen with the advent of online learning platforms (LP) Dougiamas and Taylor (2003). There is some specialized LP, such as Moodle that can be customized and extended. Inside LP, instant messaging (IM), wikis, and other social applications can be better options for collaborative work and are, thus, gaining momentum in scholarly communications tasks Watson (2013).

As an emerging communication technology, online social networks provide a variety of communication services such as profiles, comments, private messaging, blogging, media file sharing, and instant messaging. Some of these communication tools provide their services through a mobile network Chai and

Kim (2012). These features are important as they help to break existing barriers to communication among members. They can stimulate interactions involving student-teacher.

Due to demands of privacy and other pedagogical decisions, educational institutions also may choose to establish private social networks that are integrated to LP and restricted to student-teacher of their main courses. These tools are commonly inspired on public social networks, such as Facebook, LinkedIn, including Web 2.0/3.0 collaboration tools. In particular, they have to focus on educational issues, such as improvement course interactions between teacher-student and better integration with other educational tools.

The choice for a particular LP and their customizations can be both time-consuming and expensive. Under this perspective and due to investments, educational institutions are concerned to measure the effective adoption of a new communication tool (NCT). They are thus searching for an efficient way to assess if an NCT is bringing benefits for their LP. We can state the problem as of how to identify when an NCT is providing a communication flow that is complementary to the current communication tool (CCT) being used in LP.

In this paper, we address the problem of measuring the complementarity of aan NCT in the scenario where some CCT is already an established in LP. In order toTo do that, we present the Mixed Graph Framework (MGF), which is designed to evaluate how complementarity the involved communication tools are by using a mixed graph modeling. The proposed MGF is based on the premise that the CCT can be considered as a baseline for evaluating any other tool to improve communication in educational institutions LP. It is important to use a commonstandard representation of the communication flow to enable comparison between them. In this work, the communication networks from the CCT and the NCT are modeled as graphs, named G_c and G_n, respectively. From these graphs, the MGF produces a mixed graph G_m to measure if aan NCT is acting as a complementary tool among students as compared with the CCT.

We have evaluated MGF using synthetic data that represents teacher-students communication flows. In our experiment, we assume the usage of a Moodle-like LP that is mainly adopting course messages as CCT and Social Network plugin as NCT. Based on the shared messages in both tools, we compute several metrics and conduct a statistical analysis on them to evaluate the complementarity of the NCT. Our experiments observed that the MGF was able to identify whether an NCT produces significant changes in the overall communication.

The remainder of the paper is organized as follows. Sections 2 and 3 present related work and the general background, respectively. The proposed MGF is described in Section 4. Section 5 presents our experimental evaluation. Finally, Section 6 concludes the paper.

2 RELATED WORK

The analysis of social networks is widely explored, and it has been studied for several years Ngai et al. (2015). Many of these studies focused on the information that can be extracted from these networks analyzing their dynamics and structure. When it comes to the impacts of communication tools adoption in an educational environment, the need for study in this area expanded in recent years Fidalgo (2012); Kellogg et al. (2014). These studies focus on the effects of the usage of social network tools in LP and the learning-teaching achievements Siribaddana (2014).

One of the main concerns about the adoption of communication tools is related to the notion of being social Wasko et al. (2009). We can find several publications on the use of open social networks, most of them representing the information flow as a graph. In such background, it is possible to extract metrics that enables data mining Nettleton (2013), such as groups identification (clusters or cliques) that are related to concentrations of communication flows inside the graph Prado and Baranauskas (2013). Many of these metrics, such as cohesion and average distance, are useful in network analysis, as they enable insights about the communication flows Newman (2003).

In some studies, the authors structure and compare social networks by analyzing the communication flow among students of courses available in a Distance Learning Scenario Hamulic and Bijedic (2009); Siribaddana (2014). Their research showed that the data from these social tools could be used to analyze the communication flow and draw conclusions to improve the available e-learning courses.

Some researchers have proposed frameworks for understanding social media Shriram and Kaur (2011); Chai and Kim (2012) that suggest a theoretical framework to understand social networking site users'

99 knowledge contribution behavior and inter-relationships among different research constructs adopted
100 Ngai et al. (2015).

101 Many works analyze social networks and study its behavior. Also, some papers propose frameworks
102 for these purposes. Nonetheless, as far as we know no other work suggests a Mixed Graph framework to
103 measure if an NCT is complementary to a CCT already in use in an LP.

104 3 BACKGROUND

105 This section presents the fundamental concepts for the paper and is organized into three main subsections.
106 Section presents general graph concepts. Section describes the major centrality-based measures that
107 are used as input for the performed statistical analysis. Section presents the general statistical tests for
108 non-parametric data sets.

109 This section presents the fundamental concepts for the paper and is organized into three main
110 subsections. Section 3.1 presents general graph concepts. Section 3.2 describes the major centrality-based
111 measures that are used as input for the performed statistical analysis. Section 3.3 presents the general
112 statistical tests for non-parametric data sets.

113 3.1 Graph Representation

114 Using graph theory terminology Ahuja et al. (1993), communication networks (such as LP and LP) can
115 be modeled as a weighted directed graph $G(V,E)$, where V is the set of $|V|$ nodes and E is the set of $|E|$
116 edges. A node $i \in V$ represents a member with a connection point. The arcs $(i,j) \in E$, $i \in V$ and $j \in V$
117 represent a communication link between two members.

118 A weight $w_{ij} > 0$ is assigned to each edge with ending nodes i and j and represents the amount of
119 communication flow between these two nodes. Since $G(V,E)$ is directed, it may be that $w_{ij} \neq w_{ji}$. The
120 adjacency matrix $A_{i,j} = a_{i,j}$ of the weighted graph G can be defined as:

$$a_{ij} = \begin{cases} w_{ij}, & \text{if there is an edge connecting the node } i \text{ to } j, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

121 3.2 Graph Centrality Measures

122 When some problem is modeled by a graph, many properties are associated with each node, such as
123 distance and centrality. These properties provide a summary of the graph.

124 A general centrality measure is the weighted closeness of a node v Opsahl et al. (2010). If a node v
125 represents a member in an educational network, the closeness of v measures how close a member is to
126 others. Collaborators that occupy central positions concerning closeness are important in communication
127 Wasserman and Faust (1994). The weighted closeness of a node v is computed by

$$C_c(v) = \frac{1}{\sum_{x \in V \setminus v} d(v,x)}, \tag{2}$$

128 where $d(v,x)$ is the weighted geodesic distance between the nodes v and x.

Another well-known measure is the weighted betweenness centrality of a node v Kolaczyk (2009).
It is a measure aimed at summarizing the extent to which a vertex is located 'between' other pairs of
vertices. Let us introduce some notation before formally define the betweenness centrality. Consider
arbitrary nodes $u,v \in V$. A path $P(u,v)$ which starts at u and finishes at v is an ordered sequence of nodes,
$P(u,v) = <u = v_1, v_2, \ldots, v_k = v>$, such that $e_i = (v_i, v_{i+1}) \in E$ for $i = 1, \ldots, k-1$. The length of the
path $P(u,v)$ is given as the sum of the edge weights of the path and the shortest path function $s_G(u,v)$
between nodes $u,v \in V$ is given by

$$s_G(u,v) = \min_{P(u,v)} \sum_{i=1}^{k-1} w_{i,i+1}.$$

The betweenness centrality for any given node $v \in V$ is then given by

$$C_b(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}, \tag{3}$$

129 where $\sigma(s,t)$ is the number of paths $P(s,t)$ of size $s_G(s,t)$ connecting s and t and $\sigma(s,t|v)$ is the number
130 of shortest paths passing through vertex v.

The third class of centrality measure is the *Kleinberg centrality* Kleinberg (1998). The main idea is
to identify nodes that correspond to *hubs* and *authorities*. A hub is a node that points to many relevant
nodes, and an authority node is the one that is pointed by many important nodes. Both are based on the
eigenvectors related to the highest eigenvalues of the matrices AA^T and A^TA. The hub centrality of the
node v_i, denoted here by $C_h(v_i)$, is the $i-$th entry of the vector x satisfying Equation (4), where $\lambda \in \Re$ is
the highest eigenvalue of AA^T.

$$AA^T x = \lambda x. \tag{4}$$

Similarly, the authority of a node v_i, denoted here by $C_a(v_i)$, is the $i-$th entry of the vector y satisfying
Equation (5), where $\beta \in \Re$ is the highest eigenvalue of A^TA.

$$A^T A y = \beta y. \tag{5}$$

3.3 Statistical Analysis

132 The need to compare two different datasets is widespread. Such comparison may vary according to the
133 objectives of the study. We can summarize two different statistical tests that are relevant to compare two
134 data sets: (i) distribution; and (ii) correlation Larsen and Marx (2005). For each one of these scenarios,
135 there is a set of statistical tests that can be used. They vary according to the distribution of the data
136 sets. Commonly, social medias are scale-free networks and follow a power-law distribution. In this case,
137 non-parametric tests are more adequate. For the sake of simplicity, we are going to present one statistical
138 test for comparing two data sets.

139 Mann-Whitney U test, also known as Wilcoxon rank sum test from the difference in medians, is
140 a distribution analysis test. The goal of this test is to measure the extent to which the medians of two
141 independent data sets are different from each other, *i.e.*, to check if the difference between the median of
142 these two data sets is significantly different from zero.

143 Spearman rank correlation test is a correlation analysis test, whose goal is to test if the rank correlation
144 coefficient between two variables is significantly different from zero. The null hypothesis establishes zero
145 correlation between two variables.

4 THE MIXED GRAPH FRAMEWORK (MGF)

147 In this section, we present a framework to evaluate the complementarity of communication tools using
148 a mixed graph modeling, called here Mixed Graph Framework (MGF). Algorithm 1 summarizes how
149 the MGF works. The first two lines (2-3) are related to modeling graphs for communication tools and
150 are described in further detail in Section 4.1. Line (4) is described in Section 4.2 and produces the
151 mixed graph. Line (5) is described in Section 4.3 and computes centrality measures to evaluate the
152 complementarity of the NCT concerning the CCT.

Algorithm 1 Main *MGF* Algorithm

1: **function** $MGF(D\ d_c,\ D\ d_n,\ ef_c,\ ef_n)$
2:      $G_c \leftarrow fExtract_c(d_c)$
3:      $G_n \leftarrow fExtract_n(d_n)$
4:      $G_m \leftarrow fMix(G_c,\ G_n)$
5:      return $fAnalyze(G_c,\ G_m)$
6: **end function**

1: **function** $fAnalyze(G_c,\ G_m)$
2:      $r_1 \leftarrow analyzeClosenessDist(G_c,\ G_m)$
3:      $r_2 \leftarrow analyzeClosenessCorr(G_c,\ G_m)$
4:      $r_3 \leftarrow analyzeBetweennessCorr(G_c,\ G_m)$
5:      $r_4 \leftarrow analyzeEigenTopK(G_c,\ G_m)$
6:      return $\{r_1, r_2, r_3, r_4\}$
7: **end function**

4.1 Extract Functions

The first two activities of Algorithm 1 encompass modeling graphs from the communication tools. Graphs $G_c = (V_c, E_c)$ and $G_n = (V_n, E_n)$ are, respectively, generated through the extraction Functions $fExtract_c$ and $fExtract_n$ that are applied over the CCT and NCT datasets.

A node $i \in V_c$ and $p \in V_n$ corresponds to members of their respective graphs G_c and G_n. An edge $e_{i,j} \in E_c$ represents a communication in CCT from member $i \in V_c$ to member $j \in V_c$ and the edge weight $w_c(i, j)$ corresponds to the number of messages exchanged from i to j. Similarly, an edge $e_{i,j} \in E_n$ represents a communication in the NCT from member i to member $j \in V_n$ and the edge weight $w_n(i, j)$ corresponds to the number of messages exchanged from i to j.

Both $fExtract_c$ and $fExtract_n$ are User Defined Functions (UDFs) that vary according to the adopted communication tools. For example, if CCT corresponds to course messages in an LP tool, the communication flow in the graph G_c between two members i and $j \in V$ are measured by the number of posts messages exchanged by them, as described in Equation (6). On the other hand, if the NCT is an LP tool, the communication flow is measured by the weighted average of comments and likes someone is interested in extracting from the LP, as described in Equation (7).

$$w_c(i, j) = |posts(i, j)| \tag{6}$$

$$w_n(i, j) = \frac{\beta|comments(i, j)| + \gamma|likes(i, j)|}{\beta + \gamma} \tag{7}$$

Figures 1(a) and 1(b) display illustrative examples of G_c and G_n, respectively. The graph G_c is obtained by applying $fExtract_c$ over the D_c dataset and the graph G_n is obtained by applying $fExtract_n$ over D_n dataset.
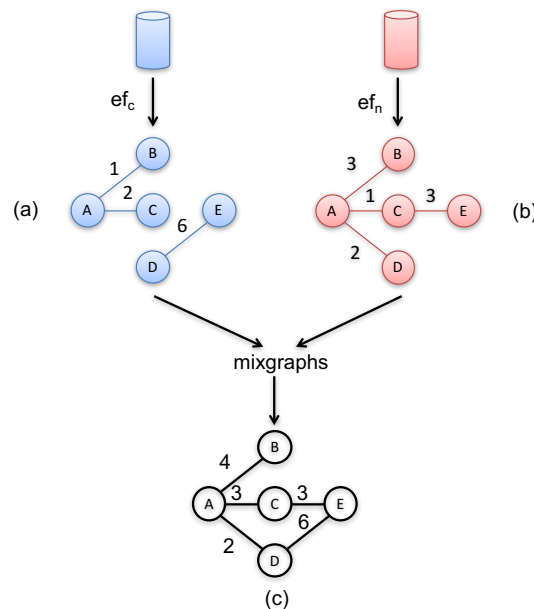

Figure 1. Communication flow: (a) G_c extracted from the CCT dataset; (b) G_n extracted from NCT dataset; (c) G_m produced by mixing G_c with G_n

4.2 Mixed Graphs

Let $G_m = (V_m, E_m)$ be the mixed graph with node set $V_m = V_c = V_n$ of order $|V_m|$ and edge set $E_m = E_c \cup E_n$. To each edge $e_{i,j} \in E_m$ a weight $w_m(i, j)$ is assigned as given by Equation (8). The mixed graph activity is described in Algorithm 2. It receives both G_c and G_n as an input and builds the mixed graph G_m of order $|V_m|$ with its edges weights given by the vector w_m.

$$w_m(i, j) = w_c(i, j) + w_n(i, j) \tag{8}$$

176   Note that the graph G_m represents the total flow of communication provided by the two communication
177 tools and can be used to identify whether the NCT is changing the communication flow or just mirroring
178 the communication flows between members in the CCT. An example of G_m can be observed in Figure
179 1(c) obtained from G_n and G_c.

Algorithm 2 Mixed Graphs

 1: **function** $fMix(G_c, nG)$
 2:      $V_m \leftarrow V_c \cup V_n$
 3:      $G_m \leftarrow EmptyGraph(|V_m|)$
 4:      **for** $i \leftarrow 1 \ to \ |V_m|$ **do**
 5:          **for** $j \leftarrow 1 \ to \ |V_m|$ **do**
 6:              **if** $i <> j$ **then**
 7:                  $w_m(i, j) \leftarrow w_c(i, j) + w_n(i, j)$
 8:              **end if**
 9:          **end for**
10:      **end for**
11:      return (mG, w_m)
12: **end function**

4.3 Complementarity Analysis

181 The complementarity analysis computes centrality measures of each vertex extracted from G_c and G_m.
182 These values are used to compute if such metrics from G_c are statistically significantly different from G_m.
183 In this case, it indicates that G_n is not simply an overlap of G_c, *i.e.*, actually bringing complementarity in
184 the overall communication. Such an activity is described in Algorithm 3.
185   It is worth mentioning that all centrality-based measures expect a weighted adjacency matrix as
186 an input. However, in all built graphs (G_c, G_n, and G_m), the weight of the edges corresponds to the
187 communication flow over a period. In this way, prior to any centrality computation, it is important to
188 convert flows to distances since more messages, e-mails, and post exchanges imply less distance between
189 two members. Such a transformation is described by Function *convertDist*(w) that applies Equation (9)
190 for all edges in Algorithm 3.

$$\overline{w}(i, j) = \frac{1}{w(i, j)} \tag{9}$$

191   Functions closeness, betweenness, and Eigen, respectively compute the weighted closeness, weighted
192 betweenness, and weighted Eigen vectors measures Opsahl et al. (2010) of G_c and G_m. The first line in
193 all functions described in Algorithm 3 is to convert the communication-based graph into a distance-based
194 graph according to Equation (9).
195   Function *analyzeClosenessDist* analyzes the closeness centrality distribution. The goal is to compute
196 if the difference in the median of the closeness of each graph is significantly different from zero. For that,
197 the nonparametric *Wilcoxon rank sum* is used Devore and Berk (2011). The intuition of this function is to
198 compute if the introduction of NCT changes the amount of communication flow significantly concerning
199 the CCT.
200   Functions *analyzeBetweennessCorr* and *analyzeClosenessCorr* correlate the betweenness and the
201 closeness centralities between G_c and G_m, respectively. For that, the nonparametric *Spearman correlation*
202 *test* is used Devore and Berk (2011). The intuition of these functions is to compute if the introduction of
203 the NCT changes significantly the way people interact concerning the CCT by analyzing the established
204 communication flows. This indicates if the NCT is not merely increasing the scale of messages among
205 persons, but if it is changing the communication flow structure. Such a test is complementary to
206 *analyzeClosenessDist*. We can have situations where *analyzeClosenessDist* may not differ but either
207 *analyzeClosenessCorr* or *analyzeBetweennessCorr* may present significant changes and vice-versa.

Algorithm 3 Analysis of Centrality (Closeness, betweenness, Eigen)

1: **function** *analyzeClosenessDist*(G_c, G_m)
2:    $vc_c \leftarrow closeness(convertDist(G_c))$
3:    $vc_m \leftarrow closeness(convertDist(G_m))$
4:    return $wilcox.test(vc_m, vc_m, conf.level = 0.95)$
5: **end function**

1: **function** *analyzeClosenessCorr*(G_c, G_m)
2:    $vc_c \leftarrow closeness(convertDist(G_c))$
3:    $vc_m \leftarrow closeness(convertDist(G_m))$
4:    return $spearman.cor.test(vc_m, vc_m, conf.level = 0.95)$
5: **end function**

1: **function** *analyzeBetweennessCorr*(G_c, G_m)
2:    $vb_c \leftarrow betweenness(convertDist(G_c))$
3:    $vb_m \leftarrow betweenness(convertDist(G_m))$
4:    return $spearman.cor.test(vb_c, vb_m, conf.level = 0.95)$
5: **end function**

1: **function** *analyzeHub*(G_c, G_m, k)
2:    $ve_c \leftarrow eigen(asHub(convertDist(G_c)))$
3:    $ve_m \leftarrow eigen(asHub(convertDist(G_m)))$
4:    $ratio \leftarrow overlap(top_k(ve_c), top_k(ve_m))$
5:    $sig \leftarrow hypergeo(ratio, m = k \cdot |ve_c|, n = (1-k) \cdot |ve_c|)$
6:    return $\{ratio, sig\}$
7: **end function**

---

²⁰⁸   Function *analyzeHub* is also a complementary analysis. It analyzes the influence of introducing new
²⁰⁹ edges in the communication flow. It starts by multiplying the adjacency matrix with its transpose targeting
²¹⁰ the main hubs in the communication flow. This is done in both graphs (G_c and G_m). Inside the function,
²¹¹ we calculate the top-k more central vertices in both graphs and the overlap between them (same central
²¹² vertices in both graphs). We also compute the probability using the hypergeometric distribution of such
²¹³ an occurrence.
²¹⁴   The MGF is implemented in R. Statistical tests, such as *Wilcoxon rank sum* and *Spearman correlation*
²¹⁵ tests, are available in many statistical packages, such as R Dalgaard (2008) and were included in *MGF*.

²¹⁶ 5 EXPERIMENTAL EVALUATION

²¹⁷ This section presents the evaluation of the proposed MGF in measuring if the NCT brings complementarity
²¹⁸ to the CCT inside a Learning Platform (LP). We used synthetic data to simulate both CCT and NCT
²¹⁹ usage to explore the MGF under different group configurations and educational scales. Both MGF and
²²⁰ experimental evaluation is made available at *https://github.com/eogasawara/mgf*.
²²¹   We have organized this section into three parts, as follows. Section 5.1 discusses synthetic data
²²² preparation that models LP Newman et al. (2002). In Section 5.2, we describe the general procedure of
²²³ growth network used in the experimental evaluation. In Section 5.3, we present a toy sample analysis to
²²⁴ illustrate the benefits of MGF. In Section 5.4, we conduct a sensitive analysis of MGF under different LP
²²⁵ scenarios.

²²⁶ 5.1 Synthetic data generation

²²⁷ Many networks can be framed in the definition of scale-free networks Barabási and Albert (1999). A net-
²²⁸ work is classified as scale-free if the degree distribution of its nodes follows the power law model Newman
²²⁹ et al. (2002). Scale-free networks have two general concepts: growth and preferential attachment. The
²³⁰ idea of growth points out to the constant increase of the number of nodes in the network. The preferential
²³¹ attachment means that the more connected is a node, the more likely is that it gets new links. The basic
²³² understanding for this second concept is that a new member on the network has a higher probability to
²³³ interact with a person who interacts with many people than with someone who is not so active in the
²³⁴ network.

The most notable feature of a scale-free network is the existence of nodes with degree much higher than the average degree in the network. The highest degree nodes are often called hubs and have specific meanings in each network. The presence of hubs is directly related to the robustness of the network. Most of the nodes are not hubs, and the probability of a significant impact on total flow with the departure of one of these low degree nodes is very low. On the other hand, the removal of a hub can cause a large impact on the communication flow or even a network partition.

In the experiments presented in our work, we generated G_c (simulating hierarchical teacher-students communication) and G_n (simulating a social network communication among all students) as scale-free networks. However, G_c follows the organizational structure formed by the traditional teacher-student relationship, whereas the G_n does not impose such a constraint. This assumption is reasonable since most LP are organized hierarchically (either teacher-students or tutors-students).

Algorithm 4 generates synthetic instances of CCT and NCT; and was implemented using *poweRlaw*, an R package to create scale-free graphs. Initially, the first three parameters k, v, e are related to generation of the subgraphs that will form CCT graph (i.e., G_c). It starts by creating k subgraphs in G_c. Each subgraph has v nodes with e edges. After that, the most central nodes in each subgraph, according to its closeness centrality, are connected to each other to establish a hierarchical communication in G_c. In the end of G_c build phase, this graph has $|V_c| = v_c = v \cdot k$ nodes and $|E_c| = e_c = (e \cdot k) + k$ edges. Then, the NCT graph G_n is generated with $v_n = |V_n|$ nodes and $e_n = |E_n|$ edges, such that $v_n = v_c$. By construction, G_n is strictly scale-free.

Table 1. Parameters used in the experimental evaluation

Parameter	Description
$v_c = v_n$	Number of nodes in both graphs, G_c and G_n
k	Number of groups in G_c
e_c	Number of edges (communication flows) in G_c
e_n	Number of edges (communication flows) in G_n

Algorithm 4 Synthetic dataset production

1: **function** *SyntheticDatasets*(k, v, e, e_n)
2:     **for all** $i \leftarrow 1$ *to* k **do**
3:         $G_c^i \leftarrow new\ ScaleFreeGraph(v, e)$
4:         $G_c \leftarrow G_c \cup G_c^i$
5:     **end for**
6:     **for** $i \leftarrow 1$ *to* $k_E - 1$ **do**
7:         **for** $j \leftarrow i + 1$ *to* k_E **do**
8:             $e_l \leftarrow connect(G_c^i, G_c^j)$
9:             $E_c \leftarrow E_c \cup e$
10:         **end for**
11:     **end for**
12:     $v_c \leftarrow v \cdot k$
13:     $v_n \leftarrow v_c$
14:     $G_n \leftarrow new\ ScaleFreeGraph(v_n, e_n)$
15:     return $(\{G_c, G_n\})$
16: **end function**

In Section 5.4, we explore three scenarios produced during synthetic data generation that correspond to representative contexts for LP, such as the number of vertices. A small course has the number of members greater than 10 and lower than 50, whereas in medium course the number of members is greater than or equal to 50 and lower than 250. Additionally, the number of messages and edges explored in our

258 study are in agreement with communications using both online social networks Benevenuto et al. (2009).
259 The scenarios adopted for LP are presented in Table 2.

Table 2. LP Scenarios

Scenario	Description
SE (G_n scale)	$v_c = 30,\ k_c = 3,\ e_c = 60$ $small: e_n = 25$ $medium: e_n = 45$ $large: e_n = 55$
SE (G_c groups)	$v_c = 30,\ e_c = 60,\ e_n = 45$ $low: k_c = 2$ $moderated: k_c = 3$ $high: k_c = 5$
ME (G_c groups)	$v_c = 150,\ e_c = 60$ $low: k_c = 10,\ e_n = 120$ $moderated: k_c = 15,\ e_n = 180$ $high: k_c = 25,\ e_n = 300$

5.2 Network Growth

261 Consider both G_c and G_n produced during the synthetic dataset production. We can apply the MGF
262 to compute metrics and check if G_n is complementary to G_c. However, to better explore MGF, in all
263 experimental evaluation we analyzed G_n using a network growth described in Algorithm 5. The goal is to
264 allow for the comprehension of the MGF behavior as we increase G_n from an empty graph until reaching
265 the entire G_n structure. According to Algorithm 5, the growth ratio δ filter both edge weights and the
266 number of edges in its entire structure according to its weight distribution. The edge weights for w_n are
267 all multiplied by $\frac{\delta}{100}$, to set the relative strength of usage in both networks. The lesser the value of δ, the
268 lesser is the communication flow inside the generated NCT. Additionally, only δ percentile of edges is
269 presented in $w_{n,\delta}$. This allows for simulating the increase of new relationships among members according
270 to time. Each combination of w_c, $w_{n,\delta}$ is used as input for *fAnalyze*. All metrics are collected and stored
271 in a result set *RS*. Once *RS* is complete, it is possible to plot charts, such as the ones presented in the
272 experimental evaluation.
273   Note that Algorithm 4 takes as input the growth ratio δ ($0 \leq \delta \leq 100$). Initially, the edge weights for
274 both G_c and G_n are randomly generated according to the same distribution. After that, Table 1 summarizes
275 parameters adopted in experimental evaluation.

Algorithm 5 Network Growth

1: **function** *NetGrowth*(w_c, w_n, r)
2:     $RS \leftarrow \{\}$
3:     **for all** $\delta \leftarrow 0$ **to** 100 **step** r **do**
4:         $w_{n,\delta} \leftarrow Filter(\delta, \frac{\delta}{100} \cdot w_n)$
5:         $w_{m,\delta} \leftarrow fMix(w_c,\ w_{n,\delta})$
6:         $RS \leftarrow RS \cup fAnalyze(w_c, w_{m,\delta})$
7:     **end for**
8:     *plotCharts*(RS)
9: **end function**

5.3 Toy Sample Analysis

277 To better understand the mechanics of the growth ratio, we present a toy graph that corresponds to one
278 of the smallest LP possible. It has ten vertices, two groups for G_c, and ten edges in both G_c and G_n
279 ($v_c = v_n = 10$, $k_c = 2$, $e_c = e_n = 10$). Figures 2(a) and 2(b) are respectively examples of the CCT
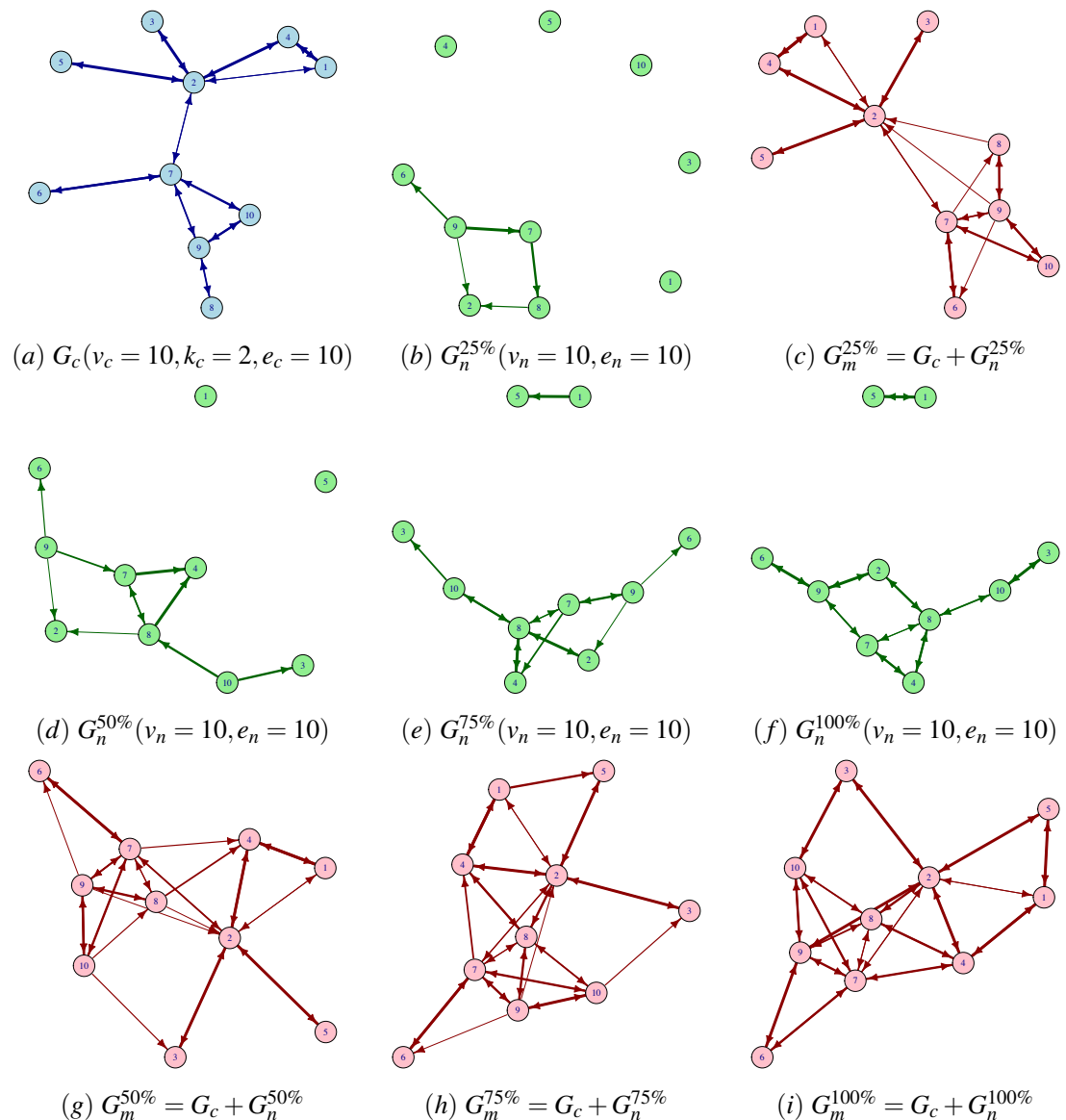
(a) $G_c(v_c = 10, k_c = 2, e_c = 10)$ (b) $G_n^{25\%}(v_n = 10, e_n = 10)$ (c) $G_m^{25\%} = G_c + G_n^{25\%}$

(d) $G_n^{50\%}(v_n = 10, e_n = 10)$ (e) $G_n^{75\%}(v_n = 10, e_n = 10)$ (f) $G_n^{100\%}(v_n = 10, e_n = 10)$

(g) $G_m^{50\%} = G_c + G_n^{50\%}$ (h) $G_m^{75\%} = G_c + G_n^{75\%}$ (i) $G_m^{100\%} = G_c + G_n^{100\%}$

Figure 2. An example of current tool G_c (a) and new tool G_n (b) produced by Algorithm 4. The mixed graph G_m is produced by Algorithm 2 from both G_c and G_n. A network growth for new tool (G_n) with ration equals to 25% (b), 50% (d), 75% (e), and 100% (f); with their respectively effects in producing mixed graphs (G_m), $G_m^{25\%}$ (c), $G_m^{50\%}$ (g), $G_m^{75\%}$ (h), and $G_m^{100\%}$ (i). The width of edges are related to their weights

and the NCT graphs produced by Algorithm 4 according to this small setup. Figure 2(c) presents the produced mixed graph (G_m) from both G_c and G_n using Algorithm 2.

In the example, Figure 2(a) simulates communications that occur through CCT inside a small course. In this case, we assume that the course has two groups. It is possible to view some clusters of communication, which can be found among students who share a close relationship, such as work on related tasks, where the internal processes of the course make them to have a direct communication. Despite these clusters, it is possible to observe that the graph is connected. This means that with the mediation of one or more persons, the information can be disseminated through the network. In a small network like G_c, we can visually inspect the characteristics that are part of the goals of our analysis, such as connectivity, the presence of clusters, and center points connecting them which are the students identified as 2 and 7.
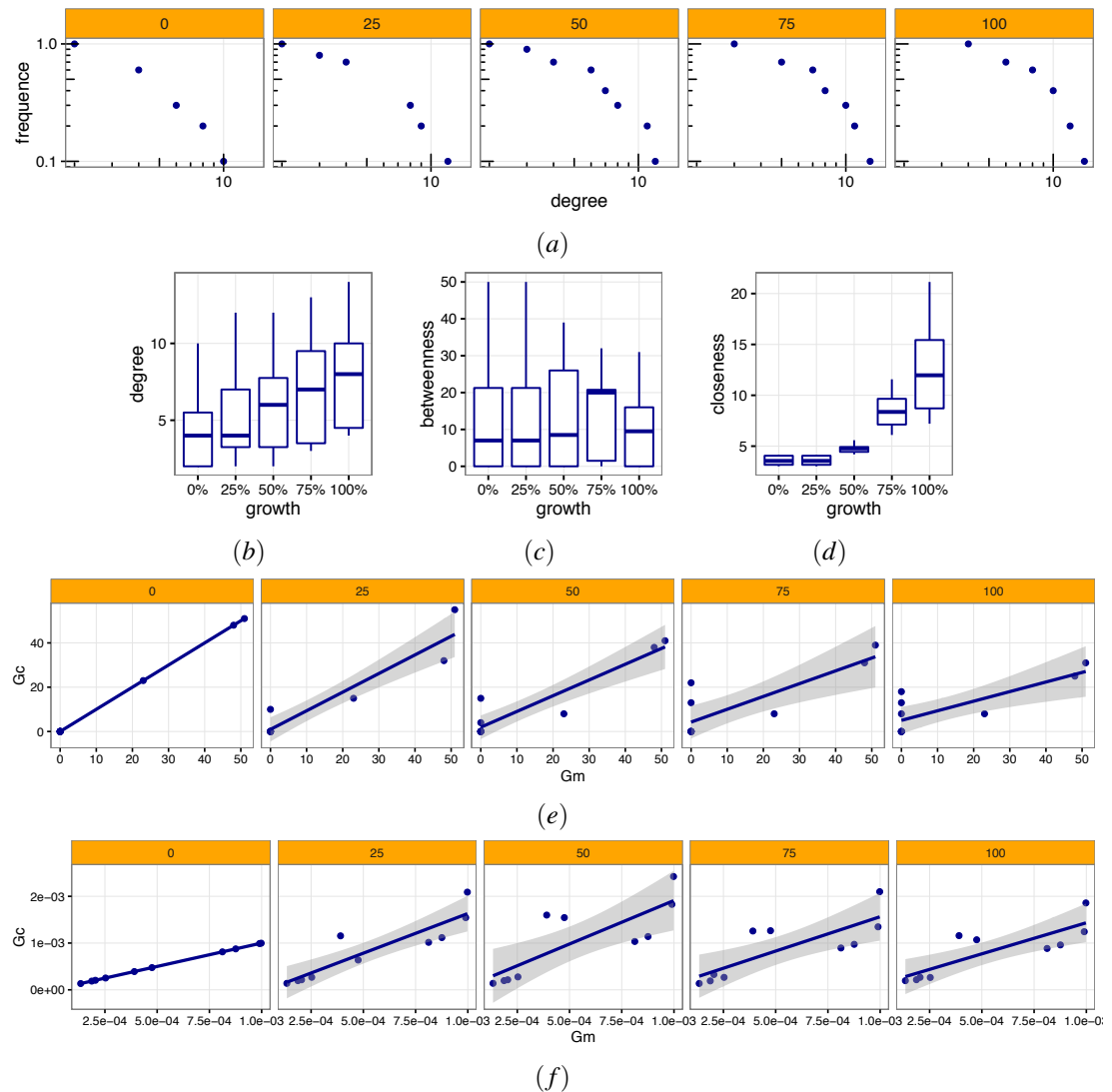
10/14

Figure 3. Descriptive statistics of G_m in the toy example grouped by growth ratio $\delta = \{0, 25, 50, 75, 100\}$. The degree distribution of G_m is in log x log scale (a). Box-plot of degree (b), closeness (c), and betweenness (d) distributions of G_m. Correlation plot of betweenness (G_c x G_m) (e). Correlation plot of closeness (G_c x G_m) (f)

Clusters communicate with each other through the central points. We applied a similar procedure to produce the graph associated to the NCT (G_n) depicted in Figure 2(b) and described in Algorithm 4.

Figure 2 explores different network growth (δ) of the new tool (G_n) using ratios such as 25%, 50%, 75%, and 100% in Algorithm 5. It is possible to observe that both the number of edges in G_n and their weights are explored in different growth ratios ($G_n^{25\%}$ (b), $G_n^{50\%}$ (d), $G_n^{75\%}$ (e), and $G_n^{100\%}$ (f)). This leads to different mixed graphs G_m: $G_m^{25\%}$ (c), $G_m^{50\%}$ (g), $G_m^{75\%}$ (h), and $G_m^{100\%}$ (i) by mixing G_c with G_n. By visually inspecting the instance of G_m presented in Figure 2, it seems that the hierarchical structure does not restrict the communication flow as the growth ratio of G_n increases.

To better comprehend the toy sample, Figure 3 presents descriptive statistics for G_m produced by mixing $G_c(v_c = 10, k_c = 2, e_c = 10)$ with $G_n(v_n = 10, e_n = 10)$. Figure 3(a) depicts the frequency of degree of G_m as G_n grows. The degree of vertices increases as G_n grows. The plots in log x log scale fits a power law distributions, *i.e.*, suggesting a scale-free graph. This behavior is also summarized in Figure 3(b). Additionally, Figures 3(c) and 3(d) describe the closeness and betweenness centrality distribution. In Figure 3(d), the box plot for growth ratios of 50%, does not present any intersection with

11/14

box plots of smaller growth ratios (0% and 25%). This indicates significant difference among them, *i.e.*, the median closeness of $G_m^{50\%}$ is higher than in G_c. Nevertheless, the betweenness described in Figure 3(c) does not present any significant difference among them.

Furthermore, Figures 3(e) and 3(f) present, respectively, a scatter plot for the closeness and betweenness correlation between G_c and G_m. The correlation is plotted with a confidence interval of 95%. It is possible to observe that both are correlated. This indicates, for example, that although Figure 3(c) indicates an increase in closeness introduced by G_n, such an increase does not change the topology of G_c, *i.e.*, it is not introducing a complementary behavior. It is actually just introducing an increase in the scale of G_m with respect to G_c.

However, analyzing these plots may not be applicable in general, especially for more extensive networks, such as in a distance learning education. To tackle this problem, the MGF uses statistical analysis to assess and monitor the complementarity of NCT. It applies the Wilcoxon rank sum test and the Spearman rank correlation test to both betweenness and closeness as described in our Main Analysis.

5.4 Sensitive Analysis

In this section, we evaluate the proposed MGF using synthetic data described in Section 5.1. It is worth mentioning that the objective of this section is not to assess the impacts of introducing a NCT. Instead, we intend to evaluate whether the MGF can distinguish G_c and G_m according to the influence of G_n. We have conducted a sensitivity analysis between networks. The goal is to identify if the NCT keeps the communication flow provided by CCT or if it introduces alternative and significant changes in the communication flows.
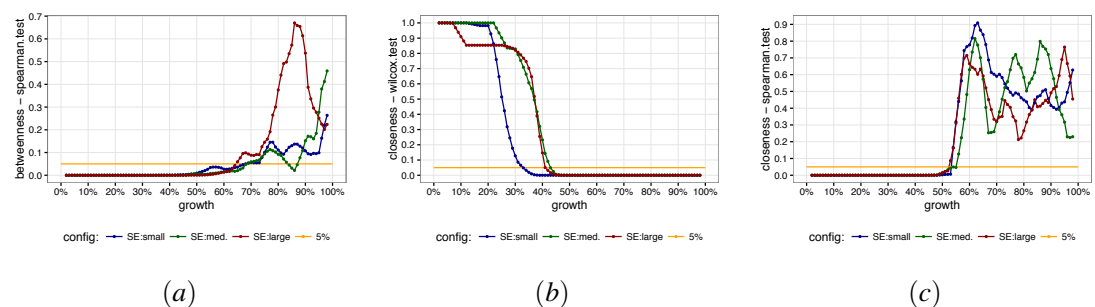

Figure 4. Scenario of Small Course - varying number of edges in G_n: betweenness correlation analysis (a), closeness median analysis (b), closeness correlation analysis (c)
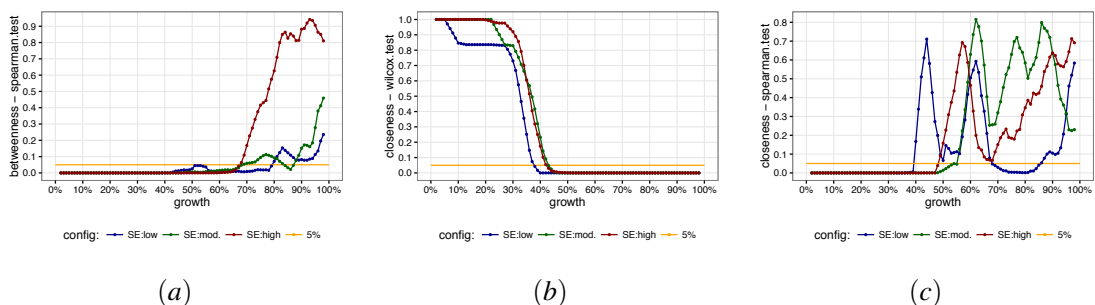

Figure 5. Scenario of Small Course - varying number of groups in G_c: betweenness correlation analysis (a), closeness median analysis (b), closeness correlation analysis (c)

In the first scenario described in Table 2 we explored the number of communication flows in the NCT of courses under a small, medium, and large scale. Regarding betweenness, Figure 4(a) indicates a significant difference for the correlation, when the growth ratio is greater than 60%. Additionally, in terms of closeness, both median (Figure 4(b)) and correlation (Figure 4(c)) presents a significant difference when the growth ratio are greater than 40% and 55%, respectively.
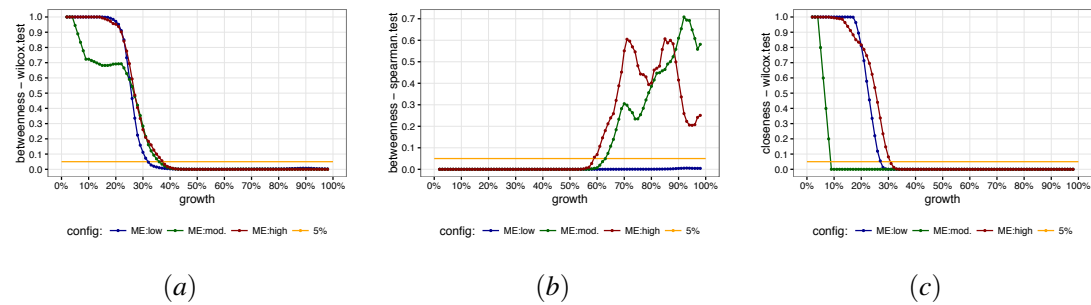
12/14

Figure 6. Scenario of Medium Course - varying both number of groups in G_c and number of edges in G_n: betweenness median analysis (a), betweenness correlation analysis (b), closeness median analysis (c)

We also explored the second scenario for courses, in which we vary the number of groups inside G_c. Figure 5(a) indicates a significant difference for the betweenness correlation when the growth increases. They were reached after a growth of 65%. In fact, the growth threshold for a significant difference occurs later when the group size is moderate or low. When it comes to closeness, both median analysis (Figure 5(b)) and correlation analysis (Figure 5(c)) present a significant differences when growth is greater than 40%. This is interesting as it indicates an increase in the number of messages in G_m and a difference in the network communication topology, as well.

In our third evaluation scenario, we explored the number of communication flows in the NCT and the number of groups inside G_c of a Medium Course under small, medium, and large scale. Regarding betweenness, as depicted in Figures 6(a) and 6(b), we observe a significant difference for the median and correlation as the growth ratio reaches 35% and 60%, respectively. A similar behavior occurs with closeness. Figure 6(c) indicates a significant difference for the closeness median when the growth is greater than 35%. In fact, for the medium size case, only when reaching an increase higher than 70% we found a clear significant difference. Before this value, we observed an oscillatory behavior around the significance threshold.

6 CONCLUSION

This paper proposes a Mixed Graph Framework (MGF), which aims at providing a set of quantitative approaches to analyze the complementary of a new communication tool (NCT) with relation to a current communication tool (CCT) in a learning platform (LP). This is done by measuring when the NCT brings significant differences in the overall educational communication flow concerning the usage of the CCT. We model CCT and NCT communication interactions as the weighted graphs G_c and G_n, respectively. From these graphs, the MGF computes a mixed graph (G_m) that combines both G_c and G_n considering their usage. Our approach is then able to identify changes in overall communication within the educational.

We also evaluated the proposed MGF using synthetic data from which we have conducted a sensitive analysis. The sensitivity analysis is used to compare the weighted closeness and betweenness of both G_c and G_m. Our approach can identify whether G_n is providing any changes in the entire communication flow. It is worth mentioning that our method does not propose adopting the NCT as a replacement for the CCT to promote communication empowerment. Instead, the goal of MGF is to aid managers in a decision-making process, giving them elements to conduct what-if analysis while deploying NCTs and measuring its influence in the entire set of communication solutions adopted in the educational.

We considered the evolution of a single network over time including time notion in the proposed framework as a promising future research. As well as performing case studies with networks of different sizes, which is a useful analysis for educational institutions with scenarios of reorganizations, mergers and divisions of courses.

ACKNOWLEDGMENTS

13/14

REFERENCES

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, New Jersey.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. (2009). Characterizing User Behavior in Online Social Networks. IMC '09, pages 49–62, New York, NY, USA. ACM.

Chai, S. and Kim, M. (2012). A socio-technical approach to knowledge contribution behavior: An empirical investigation of social networking sites users. *International Journal of Information Management*, 32(2):118–126.

Dalgaard, P. (2008). *Introductory Statistics with R*. Springer, New York, 2nd edition edition.

Devore, J. L. and Berk, K. N. (2011). *Modern Mathematical Statistics with Applications*. Springer, New York ; London, 2nd ed. 2012 edition edition.

Dougiamas, M. and Taylor, P. (2003). Moodle: Using Learning Communities to Create an Open Source Course Management System. pages 171–178, Chesapeake, VA, USA.

Fidalgo, P. S. P. (2012). Learning networks and moodle use in online courses: a social network analysis study.

Hamulic, I. and Bijedic, N. (2009). Social network analysis in virtual learning community at faculty of information technologies (fit), Mostar. *Procedia - Social and Behavioral Sciences*, 1(1):2269–2273.

Hansen, D., Shneiderman, B., and Smith, M. A. (2010). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann, Amsterdam, 1 edition edition.

Kellogg, S., Booth, S., and Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *The International Review of Research in Open and Distributed Learning*, 15(5).

Kleinberg, J. M. (1998). Authoritative Sources in a Hyperlinked Environment. SODA '98, pages 668–677, Philadelphia, PA, USA.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009 edition edition.

Larsen, R. J. and Marx, M. L. (2005). *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, Upper Saddle River, N.J, 4 edition edition.

Nettleton, D. F. (2013). Data mining of social networks represented as graphs. *Computer Science Review*, 7:1–34.

Newman, M. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256.

Newman, M. E. J., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572.

Ngai, E., Tao, S., and Moon, K. (2015). Social media research: Theories, constructs, and conceptual frameworks. *International Journal of Information Management*, 35(1):33–44.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251.

Prado, A. B. and Baranauskas, M. C. C. (2013). Addressing structural and dynamic features of scientific social networks through the lens of Actor-Network Theory. *Social Network Analysis and Mining*, 3(4):1263–1276.

Shriram, R. and Kaur, A. (2011). Assessment of online interaction pattern using the Q-4r framework. pages 544–557, Seri Pacific Hotel, Kuala Lumpur.

Siribaddana, P. A. (2014). Making Distance Learning an Effective Health Information Systems Training Strategy: A Combined Social Network Analysis and Content Analysis Perspective. *EJISDC: The Electronic Journal on Information Systems in Developing Countries*, (61):4.

Wasko, M. M., Teigland, R., and Faraj, S. (2009). The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, 47(3):254–265.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1 edition edition.

Watson, J. (2013). Combining international student social and academic transition online. *ISEJ International Student Experience Journal*, 1(2):31–36.