

# USARE GEPHI PER VISUALIZZARE LA PARTECIPAZIONE NEI CORSI ONLINE: UN APPROCCIO DI SOCIAL LEARNING ANALYTICS

## USING GEPHI TO VISUALIZE ONLINE COURSE PARTICIPATION: A SOCIAL LEARNING ANALYTICS APPROACH

Ángel Hernández-García | Universidad Politécnica de Madrid. Departamento de Ingeniería de Organización, Administración de Empresas y Estadística, Escuela Técnica Superior de Ingenieros de Telecomunicación | Madrid, Spain |

✉ Universidad Politécnica de Madrid | Av. Complutense 30, 28040 Madrid, Spain | [angel.hernandez@upm.es](mailto:angel.hernandez@upm.es)

**Sommario** La formazione a distanza online pone grandi sfide, in quanto gli studenti devono adattarsi a processi di apprendimento auto-diretto, gli insegnanti devono adattarsi a un modo diverso di erogare contenuti formativi e conoscenza e i coordinatori di un corso devono far fronte a un maggior numero di studenti avendo meno tempo a disposizione per elaborare tutte le informazioni che vengono prodotte nell'ambiente di apprendimento.

Le tecniche di Social Learning Analytics forniscono strumenti e metodi che consentono di estrarre informazioni utili per migliorare il processo di apprendimento. Lo studio di caso presentato in questo articolo mostra come il software Gephi può essere utilizzato per fornire informazioni rilevanti che sono generalmente difficili da osservare da parte dei coordinatori dei corsi e dai docenti. Questo studio empirico utilizza i dati provenienti da un corso interdisciplinare con 656 studenti, dimostra l'utilità di Gephi per gli studi di Social Learning Analytics e illustra come Gephi può fornire indicatori significativi sull'attività degli studenti e sul loro impegno. Lo studio esamina anche le potenzialità del Social Learning Analytics per migliorare l'istruzione online attraverso la visualizzazione dei dati legati all'apprendimento.

**PAROLE CHIAVE** Learning analytics, Analisi delle reti sociali, Social learning analytics, Visualizzazione dei dati di apprendimento, Apprendimento a distanza, Apprendimento online, Gephi.

**Abstract** Online education poses big challenges, especially in distance education, where students must adapt to self-directed learning, teachers need to adopt a different way of delivering educational contents and developing knowledge, and course coordinators have to cope with high student numbers and strict time limits for processing all the information originated in the learning space.

Social learning analytics provides tools and methods for extracting information that is useful for improving the learning process. This case study shows how instructors and course coordinators can use the tool Gephi to generate relevant information that would otherwise be difficult to gain. Analysis of empirical data from a cross-curricular course with 656 students proves the usefulness of Gephi for social learning analytics studies and demonstrates how the tool can provide relevant indicators of student activity and engagement. The study also discusses the potential of social learning analytics for improving online instruction via learning data visualization.

**KEY-WORDS** Learning analytics, Social network analysis, Social learning analytics, Learning Data Visualization, Distance learning, Online learning, Gephi.

## INTRODUZIONE

Le aule virtuali pongono nuove sfide ai docenti e ai progettisti di corsi online, la più grande delle quali è il tenere traccia dei progressi degli studenti e delle attività del corso, e ciò per l'impossibilità di osservare direttamente le interazioni che avvengono in classe. Questo problema ha diverse implicazioni: in primo luogo, la mancanza di segnali fisici rende difficile, per i docenti, determinare il livello di impegno e di comprensione dei singoli studenti e dei gruppi di studenti (Reffay & Chanier, 2003); inoltre, poiché le strategie di apprendimento si spostano verso l'apprendimento auto-diretto, i docenti hanno la necessità di cambiare il loro ruolo da fornitori di contenuti didattici e conoscenza a facilitatori o guide (Dawson, 2008); infine, ai docenti mancano le informazioni per esaminare le dinamiche sociali della classe, che consentirebbero loro di agire sugli studenti più influenti o su quelli indifferenti<sup>1</sup>, in accordo alle esigenze didattiche. Inoltre, sebbene gli odierni sistemi online per l'apprendimento siano in grado di memorizzare grandi quantità di dati legati alle interazioni, questi dati sono generalmente separati dalle informazioni sui corsi e dai contenuti dei corsi; di conseguenza, in generale, questi dati non sono utilizzabili in quanto sarebbe necessario selezionare e presentare solo le informazioni rilevanti a fini formativi a ciascun tipo di utente (Macfadyen & Dawson, 2012).

Il Learning Analytics nasce come una soluzione per affrontare questo insieme di sfide. Mentre non vi è alcun accordo su una definizione standard di Learning Analytics, un'ampia concettualizzazione del termine definisce il Learning Analytics come "la misurazione, la raccolta, l'analisi e la comunicazione dei dati relativi agli studenti e ai loro contesti, al fine di comprendere e ottimizzare l'apprendimento e gli ambienti in cui si verifica" (Long & Siemens, 2011). Alcune tecniche di Learning Analytics a livello di corso possono essere così classificate sulla base delle loro finalità:

- Identificazione/categorizzazione dei dati (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014);
- Analisi al termine del corso<sup>2</sup>, per migliorarne la progettazione sulla base dei risultati ottenuti;
- Modellazione di sistemi di *alert* per individuare e correggere comportamenti anomali, che consentono di identificare studenti a rischio (Macfadyen & Dawson, 2010).

Queste tecniche di analisi, e in particolar modo le ultime due, si basano in genere su tecniche di visualizzazione.

Nonostante alcuni sforzi in questo campo - ad esempio, lo sviluppo dello strumento SNAPP<sup>3</sup> (Social Networks Adapting Pedagogical Practice) per

Moodle, Blackboard o Sakai -, la visualizzazione dei dati legati all'apprendimento non è standardizzata e quindi, quando viene eseguita, richiede l'adattamento dell'output a specifici requisiti. In questo studio viene mostrato il potenziale di Gephi, uno strumento di Social Network Analysis, Open Source e *general purpose*,<sup>4</sup> per l'analisi di dati legati all'apprendimento, fornendo indicazioni generali su come esplorare ulteriormente le sue capacità per un approccio di Social Learning Analytics.

Per fare ciò, questo articolo è strutturato come segue: la prossima sezione offre una motivazione alla necessità dell'analisi delle interazioni nella formazione a distanza online ed evidenzia come le tecniche di Social Network Analytics - e soprattutto le visualizzazioni che esse utilizzano - possano aiutare ad interpretare i dati legati all'apprendimento. Sarà poi presentato uno studio di caso con i dati dell'Università Aperta della Catalogna (Universitat Oberta de Catalunya - UOC) al fine di illustrare l'utilizzo di Gephi in un contesto reale. La sezione finale fornirà le conclusioni emerse dallo studio e descriverà le principali direzioni di ricerca in questo settore.

## INTERAZIONI NELLE AULE ONLINE E SOCIAL LEARNING ANALYTICS

La formazione online è caratterizzata dalla mancanza di contatto fisico tra docenti e studenti - e tra gli studenti - all'interno di una classe, nonché dal modo in cui i contenuti sono erogati - con gli studenti che assumono la responsabilità del proprio apprendimento. Per superare questo vincolo e facilitare la comunicazione nella classe, le piattaforme di e-learning sono dotate di funzionalità di comunicazione sincrone e asincrone.

Gli strumenti di comunicazione consentono agli insegnanti di osservare solo il lato più visibile delle dinamiche del corso: i messaggi scambiati dagli studenti che partecipano attivamente; da questi, gli insegnanti dovrebbero valutare se i diversi argomenti sono stati appresi e seguire l'evoluzione del discorso, decidendo quando intervenire per supportare gli studenti nel processo di apprendimento. Tuttavia, gran parte delle attività passive non possono essere osservate ed è estremamente difficile rilevare il coinvolgimento degli studenti che non partecipano attivamente. Inoltre, gli studenti che partecipano allo stesso corso possono avere stili di apprendimento molto diversi, e talvolta la mancanza di impegno attivo non corrisponde ad una mancanza di coinvolgimento o al fatto che l'apprendimento non stia avvenendo (Wise, Zhao, & Hausknecht, 2013).

<sup>1</sup> Letteralmente, l'autore usa il termine "disconnected" (n.d.t.).

<sup>2</sup> L'autore chiama questo tipo di analisi "post-mortem" (n.d.t.).

<sup>3</sup> <http://www.snappvis.org>

<sup>4</sup> Gephi è rilasciato con licenza GPL3 (GNU General Public License): <https://gephi.github.io/>

Per esempio, alcuni studenti possono migliorare la comprensione di un dato argomento attraverso la ricerca di risorse “aperte” e possono decidere di non condividerle con il resto della classe; altri studenti possono agire come “learning witnesses” o “studenti invisibili” (Beaudoin, 2002) e usare contenuti condivisi da altri. Pertanto, a volte è molto difficile per gli insegnanti determinare se l'apprendimento stia realmente avvenendo, sia per i singoli studenti che per la classe nel suo insieme. Inoltre, anche quando sono in possesso delle informazioni necessarie, i docenti possono non essere in grado di stabilire se le dinamiche sociali che si verificano in aula sono le più appropriate.

Le tecniche di *Social Learning Analytics* possono fornire queste informazioni supplementari necessarie per prendere decisioni consapevoli e migliorare il processo di apprendimento. Buckingham-Shum e Ferguson (2012) definiscono cinque categorie di *Social Analytics* distinguendo, in particolare, tra *Inherently Social Analytics* e *Socialized Analytics*. A loro volta, le tecniche di *Inherently Social Analytics* possono essere suddivise in *Social Network Analytics* - basate sull'analisi delle relazioni interpersonali - e *Discourse Analytics* - incentrate sulla costruzione della conoscenza basata sul linguaggio. Questo studio si concentrerà sulla *Social Network Analysis* (SNA). Negli esempi forniti, il software Gephi (Bastian, Heymann, & Jacomy 2009) verrà utilizzato per analizzare le diverse interazioni tra studenti e docenti e tra studenti, in un corso online.

La SNA considera ogni attore come un *nodo* della rete, mentre le differenti relazioni tra gli attori sono concepite come linee che collegano i nodi, chiamate *archi* o *legami*. Gli archi possono essere *non orientati*, se la relazione è simmetrica, ossia se l'esistenza di un arco ( $a, b$ ) implica quella di un arco ( $b, a$ ), con  $a$  e  $b$  nodi della rete. Gli archi si dicono invece *orientati*, o *diretti*, quando l'esistenza di un arco ( $a, b$ ) non implica l'esistenza del suo simmetrico. Gli archi possono anche avere pesi diversi a seconda, per esempio, dell'intensità o del numero di interazioni tra due nodi.

In generale, gli studiosi di *Social Learning Analytics* hanno focalizzato le loro ricerche sull'identificazione dei soggetti più rilevanti in una classe (Cambridge & Pérez-López, 2012), come gli studenti influenti, gli studenti a rischio o gli studenti *intermediari* (o *broker*), che hanno il ruolo di connettere diversi gruppi. Nurme-

tra gli altri, hanno validato l'utilità del *Social Learning Analytics* per i docenti.

I vantaggi delle tecniche di *Social Learning Analytics* non sono legati esclusivamente al calcolo dei classici parametri di rete, descritti nel riquadro “Metriche per la *Social Network Analysis*”, la cui interpretazione può non essere immediata per i docenti, ma anche al fatto che queste tecniche permettono di creare una rappresentazione grafica della rete che è molto più facile da interpretare.

In questo studio si valuteranno le potenzialità di Gephi, un software per “la visualizzazione interattiva e l'esplorazione di tutte le tipologie di reti e sistemi complessi, grafici dinamici e gerarchici”<sup>5</sup>, per il *Social Learning Analytics*. Gephi ha quattro caratteristiche interessanti per il *Social Learning Analytics*:

- calcola tutti i parametri o metriche rilevanti di SNA, per una determinata rete sociale;
- offre una visualizzazione in tempo reale della rete, consentendo agli utenti di interagire con essa;
- permette l'importazione di dati in diversi formati, come GML<sup>6</sup> utilizzato da altri strumenti di SNA, così come l'esportazione di immagini della rete in SVG<sup>7</sup> e PDF<sup>8</sup> e dei dati sulle metriche in formato CSV<sup>9</sup>;
- è facilmente espandibile e personalizzabile attraverso lo sviluppo di plug-in che lo rende una scelta ragionevole nel caso in cui è richiesto un ulteriore sviluppo.

## STUDIO DI CASO: UN CORSO DI FINANZA PRESSO L'UNIVERSITÀ APERTA DELLA CATALOGNA (UOC)

### Contesto e caratteristiche del corso

Al fine di illustrare l'uso di Gephi come strumento di *Social Learning Analytics*, sono stati usati i dati del corso semestrale di “Introduzione alle informazioni finanziarie” che è comune ai corsi di laurea in “Economia Aziendale”, “Marketing”, “Relazioni di lavoro” e “Turismo” dell'Università Aperta della Catalogna (UOC). Il corso si è svolto tra i mesi di settembre 2013 e gennaio 2014, e le ragioni principali di questa scelta sono state: la struttura del corso, il numero relativamente elevato di studenti e la loro eterogeneità (il corso è obbligatorio per la laurea in “Economia Aziendale” ed è facoltativo per gli altri tre corsi di laurea).

Gli attori coinvolti nel corso sono: un professore di coordinamento, responsabile della progettazione del corso, del suo coordinamento, del monitoraggio degli studenti, dell'aggiornamento dei contenuti del corso e della gestione dei professori assistenti; dieci professori assistenti, assegnati ad una classe e responsabili dell'insegnamento; gli studenti, che vengono assegnati in modo casuale a una classe, con un massimo di 70 studenti per classe.

Per quanto riguarda il funzionamento del campus virtuale, ci sono tre diverse aree:

5 <http://gephi.org/about>

6 Graph Modeling Language

7 Scalable Vector Graphics

8 Portable Document Format

9 Comma-Separated Values

10 L'autore utilizza il termine Knowledge broker (NdT).

11 Vedi riquadro Metriche per la *Social Network Analysis* a pagina 156.

12 Pass indica il numero di studenti che hanno superato l'esame, Fail coloro che non lo hanno superato, mentre DNF indica il numero di studenti che non hanno terminato il corso.

la, Lehtinen e Palonen (1999), Dawson (2010), Sundararajan (2010), Rabbany, ElAtia, Samira, Takaffoli e Zaïane (2013), García-Saiz, Palazuelos e Zorrilla (2013), e Tobarra, Robles-Gómez, Ros, Hernández e Caminero (2014),

1. *pianificazione*: piano del corso, organizzazione, date importanti e altre informazioni pertinenti, quali compiti e risposte;
2. *comunicazione*, in cui avviene l'interazione tra gli studenti e i professori;
3. *valutazione*, una *mailbox* che consente agli studenti di inviare i compiti per una valutazione continua. Ci sono cinque attività di valutazione in itinere e gli studenti devono passarne almeno quattro per superare il corso.

L'area comunicazione è a sua volta divisa in tre sezioni:

1. *bacheca degli annunci*, dove solo i professori possono scrivere per fornire informazioni sul corso, come ad esempio risposte a domande di interesse generale, i requisiti per la presentazione e la valutazione dei compiti, le scadenze, ecc;
2. *forum di discussione*, dove ogni professore consulente può proporre agli studenti la discussione di studi di caso relativi al corso;
3. *forum generale*, una bacheca aperta per le domande degli studenti e per approfondimenti sugli argomenti di un corso - tipici messaggi riguardano domande, commenti, link di interesse, suggerimenti per migliorare i materiali del corso, ecc.

È opportuno sottolineare che la UOC è caratterizzata da un profilo atipico dello studente, in termini di età e di esperienza professionale, con un'età media di 32,4 anni nel 2012 e con circa il 90 per cento degli studenti che partecipano attivamente al mercato del lavoro.

### La raccolta dei dati

Tutte le interazioni tra studenti e docenti - e tra studenti - avvengono nell'area comunicazione. Pertanto, i dati di questo studio sono stati estratti dai log delle attività del sistema di apprendimento. Un totale di 114.756 record sono stati estratti, per 656 studenti (distribuiti in 10 classi), 10 professori assistenti e un professore di coordinamento. Inoltre, sono stati considerati i voti finali della valutazione in itinere di ogni studente.

### Gephi in azione

Tre diversi insiemi di dati sono stati generati a partire dai dati estratti dai log di sistema:

- Il *Dataset1* rappresenta ogni studente o docente come un nodo della rete, ed ogni arco orientato dal *nodo (utente) a* verso un *nodo b* ha un peso che rappresenta il numero di volte che il *nodo a* ha risposto ad un messaggio creato dal *nodo b*;
- Il *Dataset2* è simile al *Dataset1*, ma gli archi rappresentano il numero di volte che il *nodo a* legge un messaggio inviato dal *nodo b*; entrambi i dataset includono attributi del nodo per permettere l'applicazione di ulteriori filtri, come la classe a cui ogni studente è stato assegnato, il suo voto finale, il ruolo - studente o docente -, il numero to-

tale di messaggi letti, i nuovi messaggi creati e le risposte date da ogni utente.

- Il *Dataset3* rappresenta ogni messaggio come un nodo, e quindi gli archi orientati mostrano la relazione tra i messaggi - cioè, quale messaggio è stato scritto in risposta ad un altro messaggio; anche questo terzo dataset include attributi aggiuntivi per ogni nodo (messaggio), come il numero di volte che un messaggio è stato letto, la classe in cui il messaggio è stato inviato, il tipo di forum e il giorno del corso in cui è stato inviato. Gli obiettivi principali alla base di questa suddivisione dell'insieme originale dei dati è triplice, in relazione ai risultati attesi, alla loro analisi e visualizzazione:
  - Il *Dataset1* fornisce informazioni significative sulle interazioni di apprendimento *visibili* che avvengono in aula, cioè chi sta partecipando attivamente o meno nel corso, e visualizza la rete sociale delle interazioni attive; come diretta conseguenza di ciò, l'analisi di questo dataset facilita l'osservazione del ruolo svolto dai singoli studenti - i più influenti, i mediatori di conoscenza<sup>10</sup>, gli hub, le autorità<sup>11</sup> - e dagli insegnanti.
  - Il *Dataset2*, d'altra parte, fornisce informazioni significative sulle interazioni di apprendimento *invisibili* che si verificano in classe, e permette l'identificazione di "learning witnesses" e allievi passivi, così come gli utenti i cui messaggi sono stati più letti.
  - Il *Dataset3* offre informazioni importanti sulla rilevanza degli argomenti e delle discussioni, il loro valore potenziale in relazione all'impegno degli studenti e la loro distribuzione nel tempo.

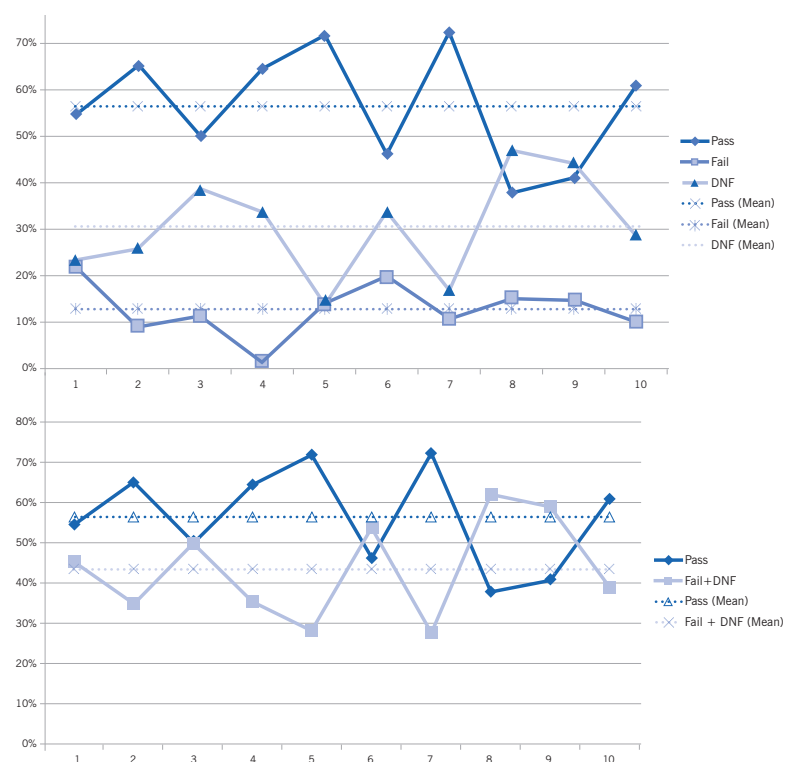


Figura 1. Tassi di successo, insuccesso e abbandono per ogni classe<sup>12</sup>.



### Analisi dei dati e risultati

In primo luogo è stata calcolata la distribuzione dei voti finali del corso, al fine di fornire ulteriori informazioni. La Figura 1 mostra quanti studenti hanno superato, non superato o abbandonato il corso in ogni aula.

Dalla Figura 1, si nota che il tasso di successo è di poco superiore al 50%, con tre corsi in cui sono stati registrati risultati inferiori (6, 8 e 9) e con due corsi che mostrano buoni risultati complessivi (5 e 7). I risultati del corso 3 sono stati particolari, con un solo studente che non è riuscito a superare la valutazione in itinere ma con un tasso di abbandono vicino al 40 per cento.

Una volta che un dataset viene caricato in Gephi, la rete sociale risultante è mostrata nella finestra denominata *Graph* nella vista *Overview* (vedi Figura 2). È poi possibile interagire con il grafico e calcolare le diverse metriche della rete sociale.

Le azioni che possono essere eseguite a questo punto sono le seguenti<sup>13</sup>:

- Calcolo dei parametri di rete (vedi riquadro Metriche per la Social Network Analysis). I risultati di questo calcolo vengono incorporati nella vista "Data laboratory".
- Filtraggio avanzato. Gephi permette di utilizzare molti tipi di filtri, come *filtri discreti*, *range* e *partition*, che possono essere applicati in maniera congiunta, fornendo agli utenti la flessibilità adeguata per raggiungere un elevato livello di raffinamento. I filtri definiti dall'utente possono essere salvati per essere riutilizzati in seguito.
- Manipolazione della rete. Come mostrato in Figura 2, in questa fase la rete sociale appare ancora un po' disordinata, e può essere molto difficile trarre conclusioni dalla sua osservazione. Sebbene a questo punto sia possibile selezionare e spostare singoli nodi o gruppi di nodi, ottenere informazioni rilevanti su ciascun nodo, effettuare operazioni come quella di zoom-in e zoom-out<sup>14</sup>, è consigliabile utilizzare i diversi layout preimpostati e la finestra chiamata "Ranking" per visualizzare la rete sociale.

- L'uso della finestra "Ranking" è piuttosto semplice ed estremamente utile per visualizzare le informazioni rilevanti. Le varie opzioni permettono di utilizzare un parametro di classificazione - comprese le metriche delle reti sociali una volta che queste sono state calcolate - e assegnare opportune dimensioni e colori ai nodi in base ai rispettivi valori. Tuttavia, e al fine di migliorare l'analisi, si suggerisce di applicare preliminarmente cambiamenti al layout.

- I cambiamenti del layout possono essere effettuati utilizzando le opzioni della finestra "Layout". Questo articolo si concentrerà sull'applicazione di tre layout: *ForceAtlas2*, *Fruchterman Reingold* e *Radial Axis*.

### LA VISUALIZZAZIONE DEI DATI LEGATI ALL'APPRENDIMENTO

Come mostrato in Figura 2, una volta caricati i dati, Gephi visualizza la rete corrispondente, disponendo gli elementi in modo casuale, per cui ricavare informazioni significative diventa un compito arduo. Questa sezione mostrerà come utilizzare i diversi tipi di visualizzazione per aiutare l'interpretazione dei dati legati all'apprendimento in una prospettiva di Social Learning Analytics.

Il primo layout, ForceAtlas2 (Jacomy, Heymann, Venturini, & Bastian, 2012), crea una rete basata sulle forze di attrazione e repulsione. In questo tipo di layout, i nodi si respingono tra loro in base al loro grado - il numero di collegamenti in entrata e in uscita - mentre gli archi attirano nodi connessi. Come risultato, i nodi con un alto grado tendono a separarsi dagli altri nodi con alto grado e attirano ad essi nodi (a cui sono collegati) che hanno un grado basso. Questa visualizzazione utilizza alcuni parametri per controllare la dispersione globale della rete, come la gravità - che controlla quanto le componenti sono disperse, separate e disconnesse -, la forza di repulsione tra i nodi, o il verificarsi di sovrapposizioni. Questo layout è estremamente appropriato per valutare le interazioni tra gli attori, come quelle relative alle interazioni attive nei forum

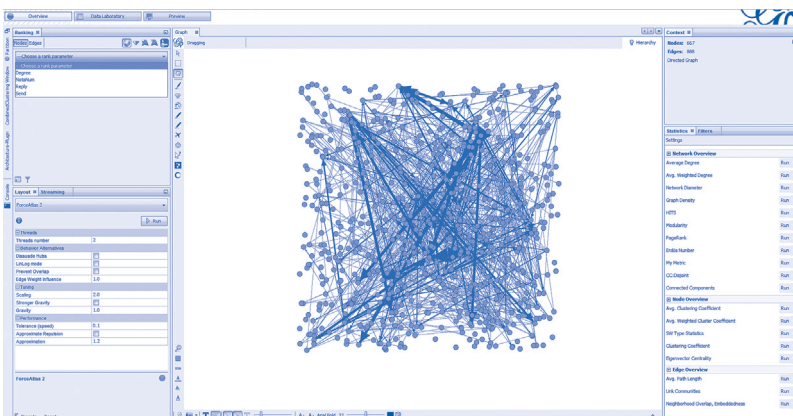


Figura 2. Vista Overview dopo l'importazione del Dataset1.

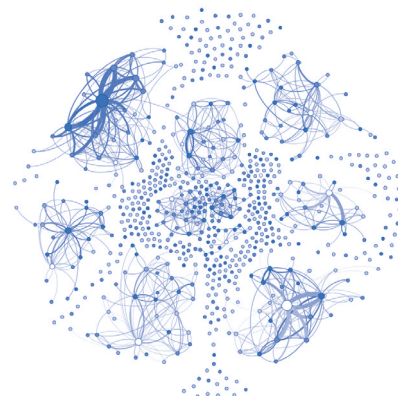


Figura 3. La visualizzazione con il layout ForceAtlas2 del Dataset1.

di discussioni (in questo studio, quelli del Dataset1), nonché per analizzare i comportamenti “invisibili”, come la lettura dei messaggi, e anche per scoprire contenuti utili e valutare interazioni di classe passive - ad esempio il Dataset2.

Come esempio, la Figura 3 mostra la rete ricavata dal Dataset1, dove il *Ranking* è stato impostato in modo da rappresentare le dimensioni di un nodo in base al numero di *in-degree*, cioè il numero totale di risposte ricevute, mentre i nodi più scuri rappresentano gli utenti con un alto numero di messaggi inviati. Inoltre, un terzo strato informativo può essere aggiunto per l'analisi visuale di alcuni degli attributi o metriche dei nodi (ad esempio, la classe di appartenenza o il voto); questo terzo strato è stato omissso dalla Figura 3 per motivi di chiarezza.

Il secondo layout, chiamato *Fruchterman Reingold* (Fruchterman & Reingold, 1991), è basato su un algoritmo *a forza-diretta* che accentua i nodi connessi posizionandoli vicini tra loro. L'algoritmo cerca anche di realizzare una distribuzione uniforme dei nodi nella rete, minimizzando gli incroci tra gli archi; il risultato è una visualizzazione come quella in Figura 4, dove è stato usato il Dataset2, con la dimensione del nodo messa in relazione con il valore di *in-degree* e i nodi più scuri che rappresentano un numero elevato di messaggi inviati da un utente. Come prevedibile, questa rete è più densa di quella del Dataset1, ed è facile identificare le 10 aule. Questa visualizzazione, come sarà mostrato in seguito, è utile per visualizzare i comportamenti attivi degli studenti e degli insegnanti classe per classe.

Infine, il layout *Radial Axis* è un tipo di layout circolare che permette una classificazione gerarchica secondaria dei nodi a seconda dei criteri selezionati. Il layout *Radial Axis* è installato in Gephi come plugin gratuito denominato *Circular Layout*<sup>15</sup>, l'idea principale è che i nodi sono uniformemente distribuiti attorno ad un cerchio secondo l'ordine determinato da un primo criterio di selezione (es. *in-degree* crescente, *out-degree* crescente, random, o altri), e radialmente distribuiti in base ad un secondo criterio (es. temporale). Tipi di visualizzazione simili

sono stati precedentemente utilizzati in contesti di Learning Analytics (Gómez-Aguilar, Theron, & García-Peñalvo, 2009) come mezzo per visualizzare informazioni temporali sulle interazioni che avvengono in aule online. Pertanto, questo layout può essere estremamente adatto per rappresentare dati temporali; ciò implica che i dati importati debbano includere queste informazioni. Tuttavia, l'analisi dei dati temporali non è semplice e quindi anche un attributo che indica il tempo trascorso in secondi, minuti, ore o giorni rispetto all'inizio del corso dovrebbe essere considerato come un'alternativa (in pratica, ciò richiederebbe la sottrazione del timestamp della data di partenza per tutti i record nel registro del sistema di apprendimento prima di procedere con l'importazione dei dati).

### Visualizzare i dataset: alcuni esempi

Data la vasta gamma di possibilità derivate dalle diverse opzioni disponibili per filtrare i dati in Gephi, questa sezione userà alcune configurazioni per fornire esempi di come Gephi possa essere utilizzato come strumento per estrarre informazioni significative dai dati dei log dei sistemi di apprendimento e contribuire a migliorare l'intero processo.

#### Il Dataset1: la rete dei messaggi inviati e ricevuti

Come spiegato in precedenza, il Dataset1 rappresenta una rete di docenti e studenti (nodi) e le loro relazioni basate sulle comunicazioni avvenute nelle bacheche (archi diretti e pesati, con il peso che rappresenta il numero di volte in cui l'utente *a* ha risposto all'utente *b*). Per l'analisi, ulteriori attributi dei nodi sono stati utilizzati: il voto finale di ogni studente, il numero totale di messaggi inviati e il numero totale delle risposte ricevute. L'analisi e la visualizzazione di questo tipo di reti con *ForceAtlas2* fornisce un'idea sulle interazioni tra gli attori che avvengono nella classe online e, in base ai criteri di selezione usati, questo layout può fornire indizi sulla relazione tra l'impegno degli studenti e il loro rendimento (Figura 5a), la coesione dei gruppi (Figura 5b), l'attività dello studente e l'identificazione dei soggetti rilevanti che possono agire come “hub della conoscenza”, autorità o mediatori (Figure 5c-5e).

Questa visualizzazione - come si evidenzia nelle Figure 5c-5e - fornisce informazioni immediate su quali studenti non partecipano per niente alle dinamiche del corso (studenti sconnessi), nonché il ruolo degli insegnanti in ogni classe. La Figura 5a conferma i risultati di alcuni studi (Alstete & Breautell, 2004; Davies & Graff, 2005; Hrastinski, 2008) secondo cui l'impegno attivo dello studente è positivamente correlato ai risultati accademici, e permette ai professori coordinatori di valutare il livello relativo di impegno in ogni classe,

- <sup>13</sup> Per un rapido tutorial vedere: [https://gephi.org/tutorials/gephi-tutorial-quick\\_start.pdf](https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf)  
<sup>14</sup> Vedere <https://gephi.org/tutorials/gephi-tutorial-visualization.pdf> per l'insieme completo delle azioni che possono essere eseguite nella finestra *Graph*.  
<sup>15</sup> <https://marketplace.gephi.org/plugin/circular-layout/>

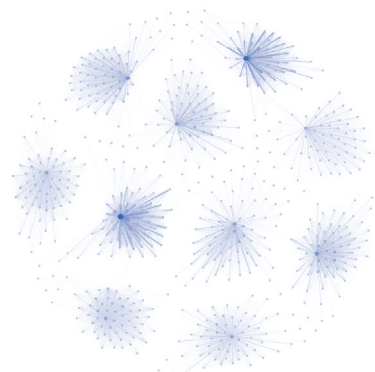
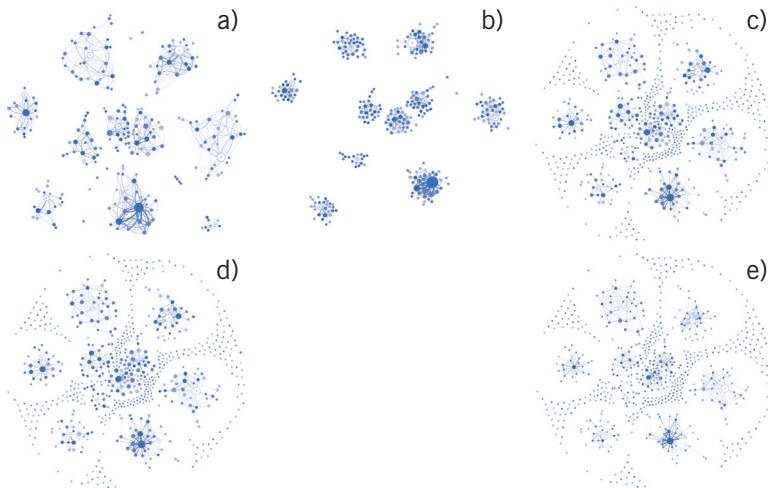


Figura 4. Visualizzazione del Dataset2 con il layout Fruchterman Reingold.



**Figura 5.** La visualizzazione del Dataset1 usando il layout ForceAtlas2. L'intensità del colore dei nodi rappresenta il voto finale e la dimensione del nodo rappresenta rispettivamente l'in-degree (a), i messaggi inviati (b), hub (c), autorità (d) e betweenness centrality (e) 16.

e ai professori assistenti di ricevere indicazioni per favorire la partecipazione in classe. Inoltre, questo layout aiuta a identificare le relazioni tra gli studenti, informazione che può essere utile per creare gruppi di studio. La Figura 5b utilizza il concetto di *componente fortemente connessa* come criterio di selezione, e facilita l'individuazione di studenti che sono debolmente collegati ai loro pari, cioè che potrebbero non essere pienamente integrati in aula. A loro volta, le Figure 5c-5e forniscono informazioni utili sugli attori rilevanti; in queste figure i nodi più grandi rappresentano gli studenti che agiscono rispettivamente come hub, autorità o mediatori. Si osservi che queste visualizzazioni dell'intera rete

possono essere particolarmente utili per il professore coordinatore. Tuttavia, dato che il livello di attività potrebbe variare notevolmente da una classe a un'altra a causa di molti fattori (ad esempio, gli stili di insegnamento), i valori estremi (outliers) in un gruppo potrebbero avere un notevole impatto sull'intera rete (per esempio, per quanto riguarda le dimensioni relative dei nodi quando si utilizza la funzione *Ranking*). Pertanto, dovrebbe essere effettuata un'ulteriore indagine sulle singole classi, utilizzando la funzione di filtro e ricalcolando i parametri di SNA in modo che il valore di Ranking possa essere ricalcolato. Queste differenze vengono mostrate nella sezione successiva con il Dataset2.

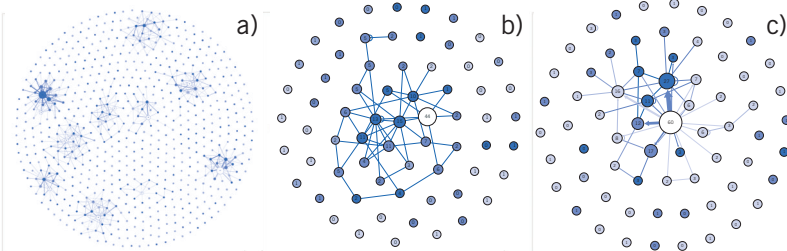
Anche il layout *Fruchterman Reingold* può essere utile per analizzare i dati provenienti dal Dataset1. Tuttavia, come mostrato nella Figura 6, le informazioni fornite da questa visualizzazione potrebbero essere più utili quando la rete è filtrata sulla singola classe. Come si vede nelle Figure 6b e 6c, possono essere osservati modelli comportamentali molto differenti, in particolare quelli relativi al ruolo dei docenti (in bianco). Anche se un'analisi approfondita di questi dati va oltre lo scopo di questo studio, i risultati hanno mostrato che le classi che hanno raggiunto risultati inferiori sono quelle che hanno avuto docenti che tendevano a intervenire troppo o troppo poco.

#### Il Dataset2: la rete dei messaggi letti

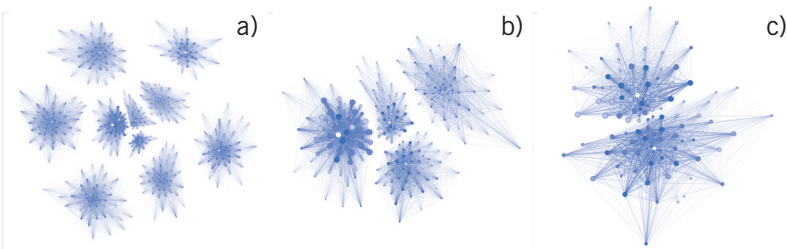
L'analisi di questo tipo di reti è analoga a quella descritta per il Dataset1 - in particolare, per i nodi è stato usato lo stesso insieme di dati usato nel Dataset1, mentre i dati relativi agli archi sono differenti. Tuttavia, poiché anche gli studenti meno attivi tendono a leggere i messaggi (almeno all'inizio del corso), le reti risultanti tendono ad essere molto più dense e più difficili da interpretare (vedi Figura 4). Una visualizzazione ForceAtlas2 fornirebbe informazioni sugli studenti che non partecipano attivamente. Come nello scenario precedente, anche in questo caso è consigliabile effettuare un filtro per singola classe al fine di evitare l'effetto dei valori estremi (outliers) (vedere Figure 7a-c).

#### Il Dataset3: La rete dei messaggi inviati

Nel Dataset3, ogni nodo rappresenta un messaggio inviato in qualunque forum di discussione. Pertanto, un layout ForceAtlas2 sarebbe utile, ma dal momento che le discussioni tendono ad aumentare nel tempo, il suo uso potrebbe non essere adeguato per corsi di lunga durata o che comprendono un numero elevato di studenti (vedi Figura 8). Invece, se si includono i dati temporali, un layout Radial Axis è fortemente consigliato (vedere Figura 9). L'uso della visualizzazione Radial Axis ha, però, un'avvertenza: il raggio utilizzato per la visualizzazione tende ad aumentare con il numero di conversazioni e di messaggi per conversazione, e quindi l'utilità di questo layout è determinata dal numero totale di

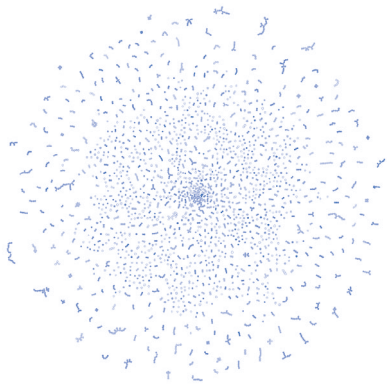


**Figura 6.** La visualizzazione del layout Fruchterman Reingold applicata rispettivamente al Dataset1 (a), alla classe 7 (b) e alla classe 8 (c). Le dimensioni dei nodi e il numero all'interno dei nodi rappresentano il numero di nuovi messaggi inviati, l'intensità del colore del nodo rappresenta il voto finale e lo spessore degli archi rappresenta il numero delle risposte.



**Figura 7.** La visualizzazione con il layout ForceAtlas2 del Dataset2 per l'intera rete (a), per le classi 4, 7, 8 e 10 (b) e le classi 7 e 8 (c). La dimensione dei nodi rappresenta il numero di nuovi messaggi inviati, l'intensità del colore rappresenta il voto finale e lo spessore degli archi rappresenta il numero totale di messaggi letti.





**Figura 8.** Visualizzazione mediante il layout ForcedAtlas2 del Dataset3 (intera rete).

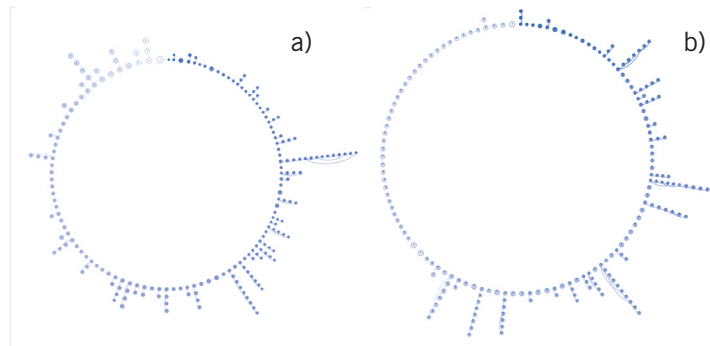
discussioni e di messaggi per discussione. Nell'esempio utilizzato nelle Figure 9a e 9b, la rete viene filtrata per classe, e un test sulle *componenti connesse* è stato eseguito utilizzando il calcolo dalla finestra *Statistica di Gephi*; successivamente, il layout Radial Axis è stato applicato, raggruppando i nodi (debolmente) connessi e ordinandoli nell'asse in base al loro ID<sup>17</sup> - in modo che i nodi più lontani dal centro rappresentassero i messaggi postati più in là nel tempo; poi, la dimensione del nodo è stata determinata dal numero di letture di ciascun messaggio - attributo della fase di pre-elaborazione - e colorato per data - dal blu più chiaro al blu più scuro - dall'inizio del corso per controllare l'evoluzione temporale.

Dalle rappresentazioni di Figura 9 è facile per i docenti vedere quali argomenti non sono stati ancora presi in considerazione e quali sono quelli più popolari o sono stati più coinvolgenti durante il corso. Questo potrebbe essere una fonte preziosa di informazioni anche per modificare le dinamiche del corso o per la pianificazione futura - per esempio, iniziando gli argomenti più rilevanti in anticipo.

## CONCLUSIONI

La formazione a distanza online deve ancora affrontare alcune sfide importanti per docenti e progettisti di corsi, soprattutto quando si tratta di tracciare le attività degli studenti. Negli ultimi anni, numerose ricerche nel campo emergente del Learning Analytics si sono concentrate su questo problema progettando sistemi in grado di estrarre, sintetizzare e visualizzare le informazioni in modo tale da guidare le azioni da intraprendere sulla base dei dati elaborati.

Il Social Learning Analytics si concentra sui dati di interazione tra i diversi attori che intervengono nel processo di apprendimento e utilizza modalità di visualizzazione dei dati per esplicitare le informazioni. Tuttavia, gli sforzi, ad oggi, si sono concentrati sull'uso di strumenti progettati "su misura" per specifici sistemi di apprendimento. Questo articolo ha dimostrato come un programma Open



**Figura 9.** La visualizzazione con layout Radial Axis del Dataset3 per le classi 1 (a) e 7 (b). La dimensione dei nodi rappresenta il numero di letture, l'intensità di colore di un nodo rappresenta la distanza del giorno dalla data di partenza del corso e l'etichetta identifica il tipo di utente che ha inviato il messaggio (studente o docente).

Source di SNA, ossia Gephi, possa essere utilizzato come strumento di Social Learning Analytics, utilizzando come esempio i dati di un corso online dell'UOC.

L'obiettivo principale di questo studio è quello di facilitare e incoraggiare i ricercatori e i professionisti ad esplorare ulteriormente le possibilità di questo tipo di approccio nell'analisi dell'apprendimento online, dimostrando come utilizzare strumenti esistenti per il Social Learning Analytics. Occorre considerare che, sebbene questo articolo sia diretto ad un pubblico generale, il gran numero di opzioni disponibili in Gephi può diventare una barriera per gli utenti non esperti; tuttavia, questa barriera può essere superata da una impostazione personalizzata delle opzioni del programma eseguita da utenti esperti e da una formazione specifica su come comprendere e manipolare meglio i risultati delle analisi provenienti da Gephi.

Si deve inoltre evidenziare che l'uso diffuso di comuni strumenti di SNA è in questo momento ostacolato dalla necessità di trasformare le informazioni dal formato utilizzato dal sistema di apprendimento al formato utilizzato dallo strumento di SNA. Tuttavia, la diversità dei formati supportati da Gephi e la relativa semplicità della struttura dei dati richiesta per l'analisi rendono relativamente facile sviluppare plug-in specifici per esportare i dati dai log dei sistemi di apprendimento. In questa logica, lo studio presentato va considerato anche come un invito per i programmatori a sviluppare queste interfacce.

Inoltre, questo studio si è concentrato sull'analisi delle interazioni sociali nell'apprendimento, ma l'uso di Gephi potrebbe estendersi ad altri tipi di Learning Analytics, come ad esempio l'analisi del discorso e le relative visualizzazioni.

Infine, va sottolineato che questo studio può essere visto come la "punta di un iceberg": il Social Learning Ana-

<sup>16</sup> Gli studenti che non sono coinvolti in alcuna conversazione sono stati filtrati nelle Figure 5a-5b. I nodi bianchi rappresentano i professori consulenti.

<sup>17</sup> È interessante notare che il layout Radial Axis sembra limitarsi a raggruppare un massimo di 127 nodi per asse secondario; pertanto, quando il numero di discussioni è più alto, dal 128esimo in poi, tutti i nodi vengono collegati lungo l'asse principale. Questa limitazione non sembra essere stata ancora affrontata nella documentazione di Gephi.



## Metriche per la Social Network Analysis

Le tecniche di Social Network Analysis (SNA)<sup>1</sup> utilizzano diverse misure di centralità, impiegate anche per il social learning analytics, come la *degree centrality*, *betweenness centrality* (Freeman, 1979), *hub e autorità* (García-Saiz, Palazuelos & Zorrilla, 2013). Ne riportiamo qui una breve descrizione. Tutte queste metriche possono essere calcolate con Gephi.

La *degree centrality* rappresenta il numero di nodi a cui è connesso un determinato nodo, e generalmente rappresenta quanto un nodo sia influente all'interno della rete; valori elevati corrispondono a nodi della rete più influenti. È stato dimostrato che la *degree centrality* è un predittore del senso di comunità (Dawson, 2008) ed è legata al rendimento scolastico (García-Saiz et al., 2013). Alla *degree centrality* sono legate i concetti di hub e autorità, che forniscono indicatori sul livello di connettività di un nodo considerando, rispettivamente, i collegamenti in uscita e i collegamenti in entrata. Gli hub sono attori con numerose connessioni uscenti e hanno un ruolo rilevante per la loro capacità di collegamento con tutti gli altri attori della rete. Le autorità, invece, sono nodi con numerose connessioni entranti e rappresentano attori a cui gli altri attori della rete riconoscono un ruolo di rilievo.

La *betweenness centrality* evidenzia i nodi che interconnettono gruppi di nodi. Nel Social Learning Analytics, questa è la chiave per identificare quali attori - studenti e docenti - possono diffondere in modo più efficace e rapido le informazioni e le conoscenze nell'intera rete. Cho et al. (2007) associano la *betweenness* ai risultati dell'apprendimento, anche se questa relazione è più debole di quella trovata con la *degree centrality*. Nurmela et al. (1999) hanno dimostrato che tutor e docenti mostrano valori di *betweenness* più alti, risultato confermato anche da García-Saiz et al. (2013). Strettamente legato ai valori di *betweenness* vi è il concetto di broker. Di fatto un broker in una rete indica un attore che ha la capacità di mettere in relazione altri attori della rete che altrimenti non sarebbero connessi. I broker fungono da intermediari e hanno una posizione strategica all'interno della rete in quanto si trovano "in mezzo" nello scambio di informazioni tra gli altri attori della rete.

1 Analisi delle reti sociali (NdT).

lytics è un campo relativamente nuovo e, come tale, stiamo cominciando a raggiungere la consapevolezza delle sue possibilità e dei suoi ambiti applicativi negli scenari educativi, sia dal punto di vista della visualizzazione dei dati legati all'apprendimento, sia da quello dell'analisi statistica delle metriche di SNA. Pertanto, l'autore auspica che questo lavoro possa ispirare altri studiosi a promuovere la ricerca e l'utilizzo di questi approcci in ambito educativo, al fine di migliorare il processo di apprendimento nella formazione online.

### Ringraziamenti

L'autore desidera esprimere il suo ringraziamento al Dr. Inés González-González e al Dr. Ana Isabel Jiménez-Zarco della Universitat Oberta de Catalunya (UOC) per il loro aiuto nella raccolta dei dati e la descrizione del dataset originale utilizzato nello studio, nonché per le loro risposte a tutte le domande relative alla struttura e allo sviluppo del corso.

La traduzione dell'articolo è a cura di Giovanni Fulantelli e Davide Taibi.

## BIBLIOGRAFIA

- Agudo-Peregrina, Á.F., Iglesias-Pradas, S., Conde-González, M.Á., & Hernández-García, Á. (2013). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, 31, 542-550.
- Alstete, J., & Beutell, N. (2004). Performance indicators in online distance learning courses: a study of management education. *Quality Assurance in Education*, 12(1), 6-14.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proceedings of the Third International ICWSM Conference*, USA, 361-362.
- Beaudoin, M. F. (2002). Learning or lurking? Tracking the "invisible" online student. *The Internet and Higher Education*, 5(2), 147-155.
- Buckingham-Shum, S., & Ferguson, R. (2012). Social Learning Analytics. *Journal Of Educational Technology & Society*, 15(3), 3-26.
- Cambridge, D., & Perez-Lopez, K. (2012, April). First steps towards a social learning analytics for online communities of practice for educators. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 69-72). ACM.
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, 11(3), 224-238.
- Dawson, S., (2010). 'Seeing' the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5), 736-752.
- Davies, J., & Graff, M.G. (2005). Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, 36(4), 657-663.
- Fruchterman, T.M.J., & Reingold, E.M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129-1164.
- García-Saiz, D., Palazuelos, C., & Zorrilla, M. (2013). Data Mining and Social Network Analysis in the Educational Field: An Application for Non-expert Users. In *Educational Data Mining: Applications and Trends*. Berlin, Heidelberg, DE: Springer.
- Gómez-Aguilar, D.A., Therón, R., & García-Peñalvo, F. J. (2009). Semantic Spiral Timelines Used as Support for e-Learning. *Journal of Universal Computer Science*, 15(7), 1526-1545.
- Jacomy, M., Heymann, S., Venturini, T., & Bastian, M. (2012). ForceAtlas2, a graph layout algorithm for handy network visualization (forthcoming).
- Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, 51(4), 1755-1765.
- Long, P., & Siemens, G. Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 31-40.
- Macfadyen, L.P., Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.
- Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3), 149-163.
- Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL Log Files by Social Network Analysis. In C. M. Hoadley and J. Roschelle (Eds.), Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference (pp. 434-444). Palo Alto, CA: Stanford University.
- Rabbany, R., ElAtia, S., Takaffoli, M., & Zaiane, O. R. (2013). Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective. In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 1-30). Berlin, Heidelberg, DE: Springer.
- Reffay, C., & Chanier, T. (2003). How Social Network Analysis can help to Measure Cohesion in Collaborative Distance-Learning. *Designing for Change in Networked Learning Environments: Computer-Supported Collaborative Learning*, 2, 343-352.
- Sundararajan, B. (2010). Emergence of the Most Knowledgeable Other (MKO): Social Network Analysis of Chat and Bulletin Board Conversations in a CSCL System. *Electronic Journal of E-Learning*, 8(2), 191-208.
- Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. *Computers in Human Behavior*, 31, 659-669.
- Wise, A.F., & Hausknecht, S. N. (2013). Learning Analytics for Online Discussions: A Pedagogical Model for Intervention with Embedded and Extracted Analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 48-56), 8-12 April 2013. Leuven, BE/ New York, NY: ACM.