CrossMark

# A social learning analytics approach to cognitive apprenticeship

Eman Abu Khousa, Yacine Atif[*] and Mohammad M. Masud

*Correspondence:
yacine.atif@uaeu.ac.ae
College of Information Technology,
UAE University, Al Ain, United Arab
Emirates

**Abstract**

The need for graduates who are immediately prepared for employment has been widely advocated over the last decade to narrow the notorious gap between industry and higher education. Current instructional methods in formal higher education claim to deliver career-ready graduates, yet industry managers argue their imminent workforce needs are not completely met. From the candidates view, formal academic path is well defined through standard curricula, but their career path and supporting professional competencies are not confidently asserted. In this paper, we adopt a data analytics approach combined with contemporary social computing techniques to measure, instil, and track the development of professional competences of learners in higher education. We propose to augment higher-education systems with a virtual learning environment made-up of three major successive layers: (1) career readiness, to assert general professional dispositions, (2) career prediction to identify and nurture confidence in a targeted domain of employment, and (3) a career development process to raise the skills that are relevant to the predicted profession. We analyze self-declared career readiness data as well as standard individual learner profiles which include career interests and domain-related qualifications. Using these combinations of data sources, we categorize learners into Communities of Practice (CoPs), within which learners thrive collaboratively to build further their career readiness and assert their professional confidence. Towards these perspectives, we use a judicious clustering algorithm that utilizes a fuzzy-logic objective function which addresses issues pertaining to overlapping domains of career interests. Our proposed Fuzzy Pairwise-constraints K-Means (FCKM) algorithm is validated empirically using a two-dimensional synthetic dataset. The experimental results show improved performance of our clustering approach compared to baseline methods.

**Keywords:** Learning analytics; Career readiness; Community of practice; Big data; Social networks; Computational science; Clustering; Fuzzy logic

## Introduction

Worldwide, 31 percent of employers are having difficulties filling available positions, not because there aren't enough workers, but because of "a talent mismatch between workers' qualifications and their specific skill sets, against combinations of skills employers want" (Group 2010; 2013). New educational approaches are needed to prepare graduates enter the workforce through improving their capacity to succeed in a knowledge economy (P21 2010). However, higher education systems do not sufficiently utilize

career-oriented data about current learners to improve the quality and the value of graduates in meeting market needs (Seely Brown 2008). Failure to exploit readily evident data and feedback on learning practices that match market needs, increases further the gap between education and industry and reduces intervention opportunities to prepare graduates for a successful career path with relevant professional performances. The pressure induced by education reforms and market needs require the integration of a new and smart learning environment in higher education to bridge diverse viewpoints and develop a common assertion of what it means to be career-ready. Developing this career-readiness capacity requires a sustained and progressive growth of professional habits and skills. Professional habits or dispositions could mature over time through a parallel path of professional development alongside the university's formal academic path. This path could further be extended to complement these habits with relevant skills. However, current methods of teaching and learning in higher education programs are not sufficient to facilitate the development of these career-readiness dimensions. To fill this gap, we propose a virtual structure named Community of Practice (CoP) as an alternative informal way to achieve this aim (Gannon-Leary and Fontainha 2007). CoP concept has actually gained momentum in different educational systems since the 1990s (Lave and Wenger 1991; Wenger 1999; Wenger et al. 2002). Many studies addressed the need to move towards CoP-based models of education to better serve the needs of 21st century students (Jakovljevic et al. 2013; Lea et al. 2005). This is mainly because sharing knowledge, especially tacit knowledge that is notoriously difficult to teach in traditional classroom configurations, has been accepted as a mean for innovation and competitive advantage.

In traditional higher education programs, students may spend years learning about a subject (learning about); only after amassing sufficient explicit knowledge, they are expected to start acquiring the (tacit) knowledge or exercise of how to be active practitioners/professionals in a targeted field (learning to be). But viewing learning as the process of joining a CoP fosters a new form of apprenticeship as students observe and emulate mentors, while engaging in a "learning to be" cycle to master the skill of a field. This involves acquiring the practices and the norms of established practitioners in the field through early and continuous cognitive and practical apprenticeship experiences. Under the guidance of established practitioners, students work together in a common (virtual) social space and participate in each other's learning process, while benefiting from mentor's feedback (Gannon-Leary and Fontainha 2007; Seely Brown 2008).

In our proposed approach, Social Networks (SNs) are employed to build online CoPs within higher education context (Gunawardena et al. 2009; Zhang et al. 2010) to influence learners following needed career prospects in the market. Besides their influential power, SNs have a substantial value in strengthening student-to-student interactions, enhancing student social engagements, and building campus communities toward improving student learning (Davis III et al. 2012). Facebook, one of the most powerful SN, is perceived to enhance the connectedness and sense of social learning in higher education settings (Baran 2010; Qureshi et al. 2015; Selwyn 2009); and to advance the practice from information-sharing to synergistic knowledge development and innovation (O'Brien and Glowatz 2013). Our approach builds a social structure that is centred around a business need and empowered with professional connectivity.

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 3 of 23

Towards that prospect, we devised a fuzzy clustering approach which predicts and sustains learner's career path along specific profeciencies. The clustering algorithm analyzes different categories of career readiness data to predict a hypothetical career practice and bring learners with similar career patterns together into the same cluster. This process leads to a social structure made up of CoPs, which are identified to specifically respond to imminent industrial needs. We consider personal specific preferences and predispositions of learners that do not disappear when they join CoPs to enrich learners' experience within CoPs as they contribute to their own growth and sustainability.

## Problem statement

Traditional higher education programs focus on instructing subjects with limited attention to actually prepare students for their future career and seizing current opportunities available in the job market. This creates the need to integrate career readiness into formal higher education to develop a new learning environment that bridges the gap between education and industry. The challenge of devising a smart learning environment that supplements formal education with career development pedagogies appears to be multifaceted. This complexity is due to the numerous factors induced when instilling professional habits and skills. Hence, the process requires to first synthesize professional habits into well-defined dimensions, and then to create a platform to nurture their development and evolution into professional practices. This is because industry-needs require both generic professional dispositions and specific domain knowledge, which are usually remote from the ones acquired in formal education. Hence, an educational environment that builds specific domain-related skills is expected to claim career-readiness upon graduation, in addition to general professional dispositions. One more challenge would be to devise the process to identify and bring individuals whose career prospects are deemed similar, into a common learning environment that is aligned with job market needs and opportunities, even before graduation. Formal predicticve analytics methods combined with contemporary social computing structures are discussed in this paper to address these issues.

## Research contributions

In our research work, we propose a new CoP model and SNs concepts to bridge the gap between higher education and industry by introducing an online social structure made up of interconnected CoPs. This structure extends the perspective of educational institutions and develops a joint effort with the industry to leverage education and workforce development. The proposed approach also provides indicators and means for institutions to intervene in order to positively affect career readiness. To enable this novel structure, we advocate three major modules: (1) career readiness, to assert professional dispositions, (2) career prediction to identify a domain of employment, and (3) career development that evolves into motivation and skills relevant to the predicted domain of practice. . In a previous research, we addressed the first module pertaining to career readiness that equip learners with generic professional habbit (AbuKhousa and Atif 2014). In this paper, we focus on career prediction, which derives career readiness data analytics out of an institution-wide portal that stores a data warehouse

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 4 of 23

about learners, along with individual learners' information which are structured into a career profile that includes attributes such as career interests and domain-related qualifications. We use these data insights to make informed decisions when categorizing learners into a prospective practice of employment. This step results into assigning learners to dedicated CoPs within which they thrive collaboratively refining their interests and practice-related skills. This career development process evolves into an online apprenticeship structure within CoPs, where social ties within and among CoPs realize our proposed (virtual) career development social structure, that unite likeminded learners with common career prospects and expert mentors. This social structure also maintains a potential influence from peers across CoPs to keep learners' horizons open in adjusting their career plan. Thus, the main contributions of this work are:

1. A CoP model in higher education to support career-prediction.
2. A portal structure to capture individual professional traits to support career-readiness.
3. A Career Profile data structure to record both individual professional traits and career aspirations.
4. A Fuzzy clustering algorithm to match similar career profile patterns and construct CoPs that are driven by current industrial needs.
5. A Social Learning Analytics (SLA) framework to track career development within CoPs.

### Running scenario

In a medical school, learners spend two years studying general medical knowledge (called Basic Sciences), and two years of Clinical Sciences were they get to spend time acquiring knowledge in different medical specialties. They learn about sub-specialties as well, but only after completing the required rotations across medical specialities to build background and interest into a potential medical career practice. The selected specialty results in a Residency program within the scope of the specialty, like family medicine, internal medicine, paediatrics, dermatology, surgery, etc.

In this scenario, our CoP concept is built around pediatrics professional practice, identified as underserved area with estimated deficit of 52 % in health care markets. Paediatricians follow the same medical training regime as other doctors that offer general medical knowledge with the opportunity to specialize in pediatrics. Subsequent pediatrics internships and residencies last several years to provide clinical rotations in general pediatrics, infancy care, and a chosen sub-specialization (such as paediatric cardiology, pediatrics pulmonology, or pediatrics emergency care). In 2010, only 33 % of general pediatrics residency graduates planned on sub specializing, yet health-care operators demand are growing for sub-disciplines. Our model fits in this scenario to drive medical study learners into pediatrics profession at early stage of their journey. As illustrated in Fig. 1, the model analyzes data from learners' individual profile as well as the business trends to support leaners in medical school in choosing their future practice specialty. We will refer further to this scenario throughout the different stages of our model in subsequent sections of this paper.

**Fig. 1** CoP apprenticeship model: a healthcare scenario
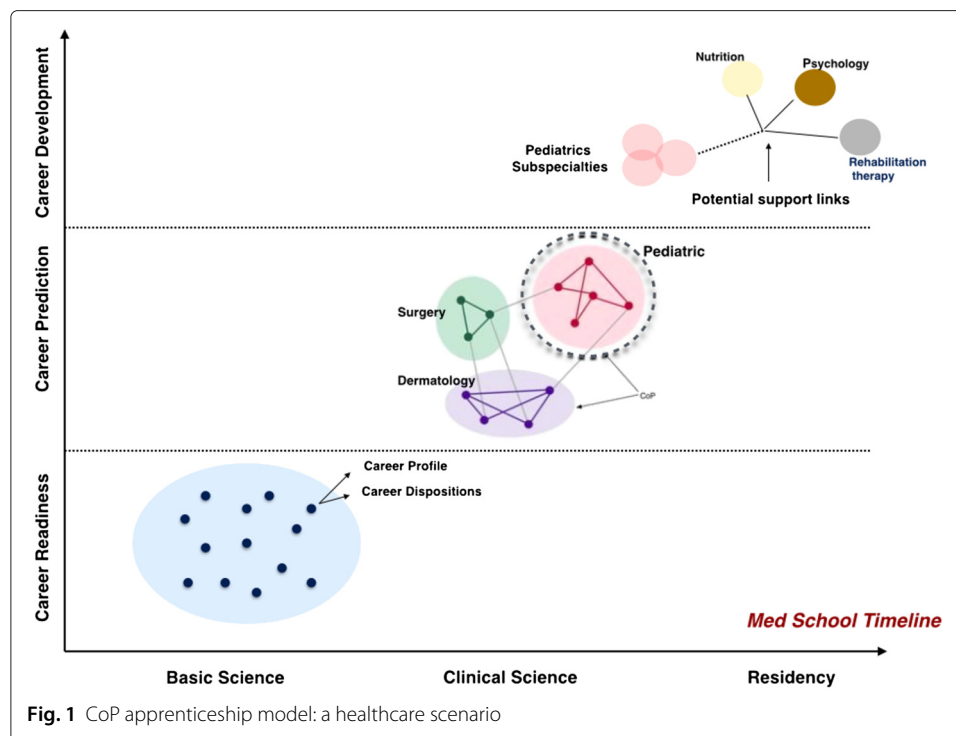
**Paper organization**

The remaining sections of this paper are organized as follows. Section 'Background and related works' provides some background and explores some related works. Section 'Community of practice apprenticeship model' presents the general framework of our proposed CoP apprenticeship model, while Section 'Fuzzy semi-supervised clustering algorithm' describes our proposed social learning analytics method for career prediction. Section 'Performance evaluation' reveals some results of our experimental analysis which demonstrate the advantages of our CoP clustering method over standard methods. Finally, Section 'Conclusion and future work' concludes the paper with a summary of our contributions and our future work.

**Background and related works**

**Social network for learning and professional development**

Social networks drive new forms of collaborations and contacts, and provide a fruitful platform for social learning as well. In social networks, people develop social relationships or ties, related to their domain of interest. These ties are leveraged for gaining access to new knowledge and learning opportunities (Haythornthwaite and De Laat 2010). The impact of online social networks on education has been addressed considerably in previous research works (Greenhow et al. 2009; Liccardi et al. 2007; Reich et al. 2012; Tian et al. 2011). For higher education in particular, online social networking with peers and faculty presents a dynamic platform for gaining information and knowledge which influences students' learning outcomes and academic achievements (Blankenship 2011; Hung and Yuen 2010; Yu et al. 2010). Some studies reported that students' social networking behavior is positively associated with their academic success and grade performance (Junco et al. 2011; Hwang et al. 2004). Furthermore, a link has been

revealed between social networking and college students' social well-being (Burke et al. 2010; DeAndrea et al. 2012; Helliwell et al. 2004; Steinfield et al. 2008). A comprehensive literature review and research directions pertaining to social networking in higher education has been presented in literature (Davis III et al. 2012)
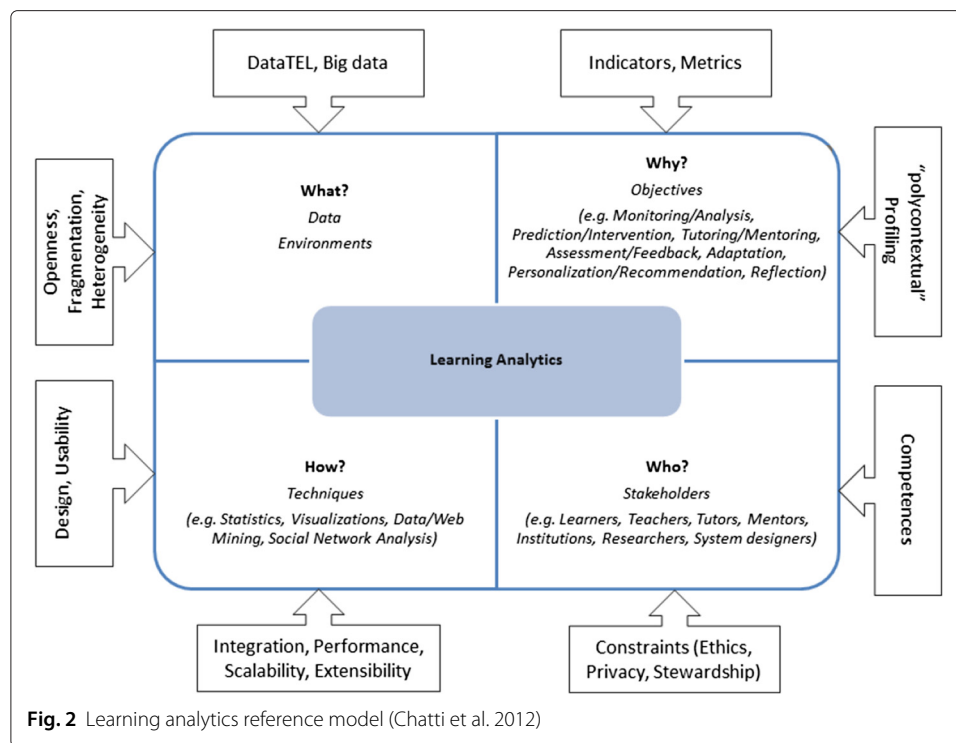
Moreover, social networks research has shown that having an extended social relationship is crucial for personal and professional development (Katz and Earl 2007; Ozgen and Baron 2007; Scott et al. 2011). Individuals could gain advantage from their personal social networks to enhance their opportunity to become entrepreneurs, to improve their job performance, to achieve higher mobility and to build career-related aspirations (Podolny and Baron 1997; Seibert and Kraimer 2001). In business, newcomers can benefit from social networks to learn organizational and tasks knowledge; and to enhance their social integration (Bauer et al. 2007; Morrison 2002).

Recent research works indicate that university students are active Facebook users to support their education experience (Hew 2011; Kabilan et al. 2010; Selwyn 2009). However, a study involving 1749 medical students who use Facebook for academic purposes, argued that they made no connections with professionally-oriented social networks that might be worthwhile for their future professional development, nor with other aspects of how social web technologies might support their professional practices (Gray et al. 2010). More importantly, most of the students indicated that Facebook didn't support their learning as they hoped, largely due to factors related to group organization and member self-discipline. Our research taps into the emergence of social structures to extend their education reach to professional and industry-related practices, in order to minimize the notorious gap between industry and education.

Our approach expands social network structures to professional career development, based on prescribed dispositions and involving the participation of expert mentors. Advances in Learning Analytics (LA) are employed to support the evolution of this extended social structure in order to match learners within higher education contexts and their predicted career orientation, while reinforcing joint social ties (with other similar learners) to support global intelligence about common practices of the predicted profession.

## Learning analytics

The widespread use of technology allows capturing unprecedented amounts of digital data about learners' interests and activities, as well as detailed sets of events and scenarios occurring in educational contexts. Learning Analytics (LA) is an emerging computational research discipline that focuses on developing methods to analyze and detect patterns to infer changes and improve learning outcomes (Ferguson 2012). As a concept, LA is drawn from data mining (DM) research applied to education (Romero and Ventura 2007). LA has a pedagogical orientation toward learners and teachers, emphasizing data in educational contexts then deriving new structural patterns from these data (Chatti et al. 2012; Pardo 2013; Siemens 2010). LA synthesizes several existing techniques such as information retrieval, machine learning and statistical algorithms to explore data and discover hidden patterns. This process aims to achieve objectives closely aligned with the learning experience ranging from simple feedbacks, to reflection and self-awareness in order to predict and recommend corrective personalized actions [Removed for blind review]. A typical LA model (Fig. 2) has four key components: data and environment (what kind

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 7 of 23



**Fig. 2** Learning analytics reference model (Chatti et al. 2012)

of data to collect and analyze), stakeholders (who is targeted by the analysis), objectives (why analyzing the collected data) and methods (how to perform the analysis) (Chatti et al. 2012; Greller and Drachsler 2012).

The proposed LA tools in the literature use a combination of descriptive and predictive DM models (Ali et al. 2012; Hämäläinen 2006). For example, Xu et al. (2010) propose an analytical tool based on a clustering model that can be used to predict which kinds of teachers are more likely to adopt digital libraries. The proposed tool aims to help teachers become more effective digital library users. Zimmermann et al. (2011) construct a classification/regression model to predict graduate level of performance from undergraduate achievements in order to improve future graduate study admission procedures. Koprinska (2011) showed how correlation and regression in DM analysis can be used to gain a better understanding of the assessment results toward predicting final marks. This can be used to improve future offerings of courses and provide timely feedback to students during the semester. Another research work uses association rules to investigate student's patterns in using the Learning Management System (LMS) resources (Merceron 2011).

Most proposed tools in literature use data from adaptive learning systems/Intelligent Tutoring Systems (ITS), Web-based Courses and LMS to achieve adaptation of learning (Arnold and Pistilli 2012; Cabada et al. 2011); with classification and prediction as the most used LA techniques. Adaptive learning orchestrates the allocation of educational materials; and adapt their presentation according to the unique needs of each learner. LA achieves adaptation through guiding learners on what to do next by organizing instructional activities and learning resources according to their personal needs. The literature also shows that there are only few LA studies that target the learner and the teacher as key stakeholders (Chatti et al. 2012; Siemens and Long 2011; Siemens and d Baker 2012). Chatii et al. (2012) noted that this pattern should change in future as the focus of LA is

shifting toward more open, networked, personalized and lifelong learning environment. This is evidenced by the increasing use of Social Network Analysis (SNA) methods to build LA tools (De Liddo et al. 2011; Dawson et al. 2011; Leony et al. 2012; Pardo 2013; Rabbany et al. 2014), which are leading LA research to promote open learning environments (Colthorpe et al. 2015; Kitto et al. 2015; Martín et al. 2015; Segedy et al. 2015; Xing and Goggins 2015)

Our framework is positioned within this current trend aiming to apply LA techniques in providing a social environment that supports lifelong learning and professional development. Our model provides an environment that empowers learners to reflect and act upon feedback about their learning performance towards a career vision. Instructors or mentors are kept in this feedback loop to intervene at complementary levels during learning processes. Up to our knowledge, this approach is pioneering the integration of LA techniques for career success objectives by focusing on meta-learning dimensions that accompany formal education. In doing so, we use LA techniques to reveal hidden patterns of common traits among learners in higher education, which are viewed as future candidates for the job market. These patterns could evolve into communities of practice, bringing together learners with shared career interests to develop socially rather than individually their common career orientation.

Social Learning Analytics (SLA) is a distinctive subset of LA, which highlights the social perspective of learning. SLA draws on the significant educational research work evidencing that new skills and ideas are developed and passed on through interactions and collaboration; and that learning cannot be understood without reference to context. As a group of learners engaged in a joint activity, their success is related to a combination of individual knowledge and skills, environment, use of tools and ability to work together (Wells and Claxton 2008; Wertsch et al. 1995). SLA develop potentials to make use of data generated by learners' traces through their online activities in order to identify behaviors within learning environments that indicate their learning performance. A good discussion of different drivers behind the emergence of SLA is provided in (Shum and Ferguson 2012) concluding that LA in general must be reframed to place a special focus on online social interaction and social construction of knowledge. Our model uses SLA to synthesize a community of practice structure where learners thrive towards a prescribed career outcome. SLA techniques are also employed to drive the lifecycle of this community of practices based on the dynamics of learners such as individual dispositions, traces and ties in the social network. The literature identifies several SLA approaches as well as related tools and potentials in the context of innovative models of education (Shum and Ferguson 2012). Our work contributes to these innovative trends through computational techniques that cluster learners into social structures.

### Clustering algorithms

Clustering algorithms can be categorized into unsupervised and semi-supervised approaches depending on whether we have certain prior knowledge about the clusters. Unsupervised clustering assumes we do not have any knowledge about the clusters. Semi-supervised clustering, on the contrary, assumes that we know the labels of certain objects. These objects are usually used as "seeds" and the clustering process utilizes these seeds to improve the clustering performance. Constrained clustering is another method of semi-supervised clustering within which the final clusters need to satisfy certain constrains.

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 9 of 23

The most often used constrains are must-link and cannot-link. If two objects are connected by a must-link, they must be in the same cluster. If two objects are connected by a cannot-link, they must be in different clusters. In our work, we focus on the famous K-Means algorithm as a centroid based semi-supervised model. Our proposed algorithm is built on the baseline of two K-means candidate methods: Seeded K-Means (SKM) (Wang et al. 2011); and Pairwise-constraints K-Means algorithm (PKM) (Wagstaff and Cardie 2000; Davidson and Basu 2007).

SKM algorithm uses seed clustering to initialize the K-Means algorithm rather than random means. Given a dataset X, the goal is to split this dataset into K disjoint clusters $\{X_h\}_{h=1}^{k}$ such that the local objectives function is minimized. Let $S \in X$ be the subset of data objects, called the seed set. For each $x_i \in S$, the label $y_i = h$ of $x_i$ denotes the cluster $X_h$ which $x_i$ belongs to. The seed set S is partitioned into L disjoint sets $\{S_h\}_{h=1}^{L}$ where $L \leq K$. If $L = K$, the seed set is called complete. Otherwise, it is the case of an incomplete seeding. In SKM, each initial cluster center $\mu_h$ is computed as the mean of data objects with the label of h in the seed set.

PKM algorithm modifies K-Means algorithm to integrate domain knowledge based constraints that the search strategy is biased towards the solutions which respect these constraints as many as possible. These constraints are respected strictly or partially depending on the different clustering algorithms. The constraints are provided in the form of pair-wise constraints: must- link and cannot- link. A must-link $c = (x, y)$ or a cannot-link $c \neq (x, y)$ constraint between two objects x and y means that these two objects must or must not be in the same cluster, respectively. It is generally assumed that these constraints are provided by the domain expert or derived from domain ontology. Some PKM techniques force constraints satafication without violating the constraints (i.e COP-K-Means); while others allow constraints violation with certain penalties (i.e. CVQE).

## Community of practice apprenticeship model

Our solution aims at augmenting the formal curriculum instruction and physical classroom environments in higher education settings with a virtual "cognitive apprenticeship" environment synthesized by our CoP model. This social structure influences 21st Century education to narrow the industry gap by guiding learners towards a desired career path. Ancient apprenticeship methods helped earlier learners seeing parents or mentors plant or harvest corps with other partners, and piece together garments under the supervision of a more experienced tailor. We use this inspiration to augment formal schooling with the process of becoming a member of a mentored CoP that supports a successful career, immediately upon graduation. This process involves developing an identity as a member of a community. The process starts by joining the most suitable CoP based on initial career dispotion data and adverised career profile interests. CoP provides an apprenticeship model (Fig. 3) to promote learning environments which render key aspects of a discipline and make domain-specific practices visible to learners, while still enrolled in academia. CoP acts as a virtual classroom where social interactions and collective intelligence contribute to the development of individual career interests.

The proposed methodology to achieve these outcomes consists in first, defining and validating standard career disposition dimensions. These intangible disposition indicators are converted into numerical "raw scores" which are then stored in a data warehouse

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14
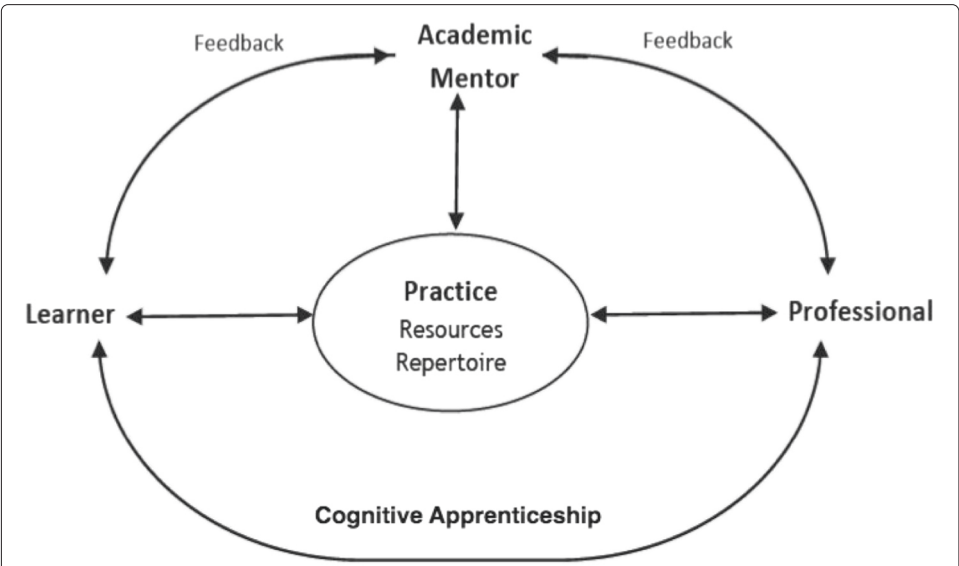
Page 10 of 23



**Fig. 3** CoP-Based Apprenticeship Learning

for further aggregation and analysis. This process creates the opportunity to systematically cluster individuals into similar career patterns to form CoPs. This new online social structure expands the perspective of educational institutions to provide a virtual platform that builds up learners' career readiness capacity along industry needs, and evaluate their professional development during the course of their academic study. Thus, we introduce a CoP-based instructional model illustrated in Fig. 4 that consists of three major modules: 1) career readiness, 2) career prediction, and 3) career development as a driver to improve career readiness and enhance professional success opportunities of learners in higher education institutions.

**Career readiness**

At the first stage of our scenario, learners fill out the Career Profile where they provide information about their competencies, qualifications, interests and skills. For example, going back to our scenario, students could list their medical career interests. Learners also complete a Career Readiness survey in order to measure their Career Dispositions. These are the generic skills that engenders the professional and deontological behaviors. In a previous work, we addressed this stage of career-readiness through the provision of an online instrument for collecting self-assessment data to produce willing, confident and creative lifelong learners (Atif et al. 2014). The provided instrument presents a storehouse view of career dispositionsthrough an integrated portal which captures self-stated learning experiences and converts them into analytical results. The outcome of this stage roots out deficiencies in dispositions for the targeted practice and prescribe improvement recommendations.
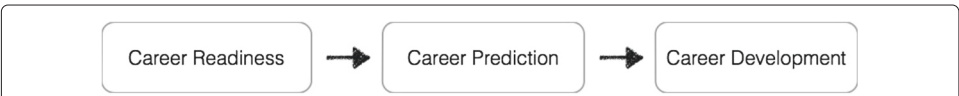


**Fig. 4** Career readiness framework

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 11 of 23

We define the concept of career dispositions that emerge as the joint set of attitudes and generic skills that dispose individuals to engage profitably with learning from new professional environment in order to be able to adapt to career changes and to manage their career growth. We model these dispositions as a 6-dimensional construct that comprises: Openness to challenge (OC), Critical Thinking (CT), Resilience (R), Learning Relationships (LR), Responsibility for Learning (RL), and Creativity (C). These dimensions in general describe the natural tendencies, mind state and preparations of each individual towards a professional practice. As implied by disposition label, high score learners in openness to challenge are those who are curious and open to new ideas and experiences. Critical thinkers are those who are evidence based decision makers. learners who score high in resilience dimension are those who are determined, competitive and achievement oriented. While social oriented learners score high on learning relationships dimension, dependable and motivated learners are most likely to score high in responsibility for learning dimension. Creative learners are those who are original, imaginative and adventuresome. We developed the Self-Reflective Career Dispositions Scale (SRCDS) metric that is a self-report instrument to quantify these dimensions and qualify learners to embrace professional practices. Career disposition indicators are converted into numerical "raw scores", which are then stored in a data warehouse for novel aggregation and analysis (Atif et al. 2014). This process provides the opportunity to create mentoring workflows to support a portfolio of assessments that gauge learners' progress across curricular instructions and their social and professional interactions in the industry-needs matching CoP they assigned to by career prediction process.

The developmental realization of a career is better achieved by uniting around a common goal to learn from each other and from expert domain-specific mentors. The collected learning data from this initial career association process is further analyzed against current industry trends to refine career-patterns that in-turn synthesize further industry-needs matching CoPs. Learning Analytics (LA) techniques identify indicators that bridge education with industry needs to leverage workforce developments. Model of ontologies will be used to describe industry needs and market trends; in order to be able to match them with the learners' domain of career interest (Maynard et al. 2005).

**Career prediction**

This paper scope falls within the Career Predication step through a model that allocates and connects learners who share common career interest to initiate a CoP experience. For example, medical practice students who share pediatrics interests could foster a comon CoP. Learners may actually be assigned to several CoPs according to their interests, which results into potential overlap between CoPs as learners interests may intiailly span multiple specialty prospects. At the hub of each CoP, there is a group of learners who displayed a high level of career dispositions (inferred from the portal analytics in the previous step). These seed learners support the elaboration of relationships with other medical practioners within selected disiplines labelling the CoP. Our model suggests to survey the current industry needs as part of CoP metadata. In the context of our scenario, the pediatric market demand analysis lists expert personnel deficiency in five sub-specialization for the next coming seven years (2016-2021) that are: Allergy/Immunlogy, Anesthesiology, Cardiology, Cardiothoracic Surgery, and Critical Care (Ministries 2015). Our dynamic CoP structure then evolves to transcend pediatrics medical practitioners into sub-disciplines,

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 12 of 23

forming new CoPs as illustrated further in Fig. 1. Each CoP is assigned an expert mentor to operate the community synergistic relationships. This includes sharing experiences and learning resources to sustain the development of interest and skills of community members in a collaborative effort. Our model suggests a new role provided by the industry which in this case is the medical sector to incorporate representative pediatrician with a pedagogical profile to mentor the community. CoP admits automatically all learners who pass the disposition threshold and meet the advertised discipline by the CoP.

Towards this end, the career prediction module analyzes data from learners' profile and career disposition values in order to predict a hypothetical career practice and bring learners with similar career patterns together into a common cluster. This process leads to a social structure made up of CoPs that are identified to specifically respond to imminent industrial needs. Learner's career profile construct (Fig. 5) is designed as a standard mean to collect and access information about learners while they are moving towards a predestined career path. Career profile augments an existing IEEE Learner Information Package (LIP) standard (LIP 2001) to capture learning data as well as career indicators. Our propose construct of career profile is structured into three main categories aimed at predicting and assisting learners with their career development throughout their formal education. We use LIP-defined Competency and Goal categories to specify domain-related qualifications, and long term career objectives of individual learners. We also introduce a new category labeled Professional as a slot for career dispositions ratings and other generic attributes pertaining to career readiness. As shown in Fig. 5. The multidimensional data attributes reflecting the professional aptitude, career prospects and dispositions of a learner are used to detect a CoP, where members share knowledge, experience and passion for a predicted practice to build capabilities and maintain momentum. The reliability of gathered data for the algorithm depends merely on the learners' awareness of the objective of the data use. Unlike the use of self-reported data in higher education for examination or evaluation purposes, learners are motivated to share their learning and behavioral data to improve their professional development and so to enhance their career advancement".

To solve the cold-start problem of CoP construction, we use the career readiness data warehouse discussed earlier as a source for initializing groups (or clusters) of learners and denote each such cluster as a CoP (Fig. 6). In order to conduct this initial grouping process, we apply a clustering technique that brings a seed set of learners into an initial set



**Fig. 5** Career profile

**Fig. 6** Career prediction & CoP construction
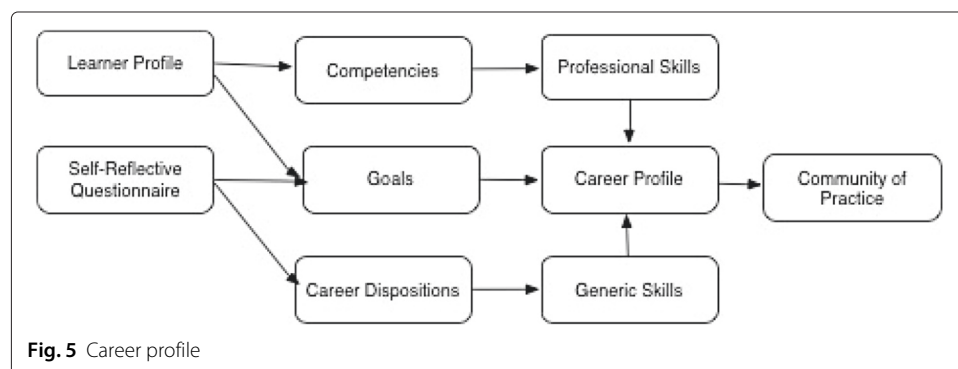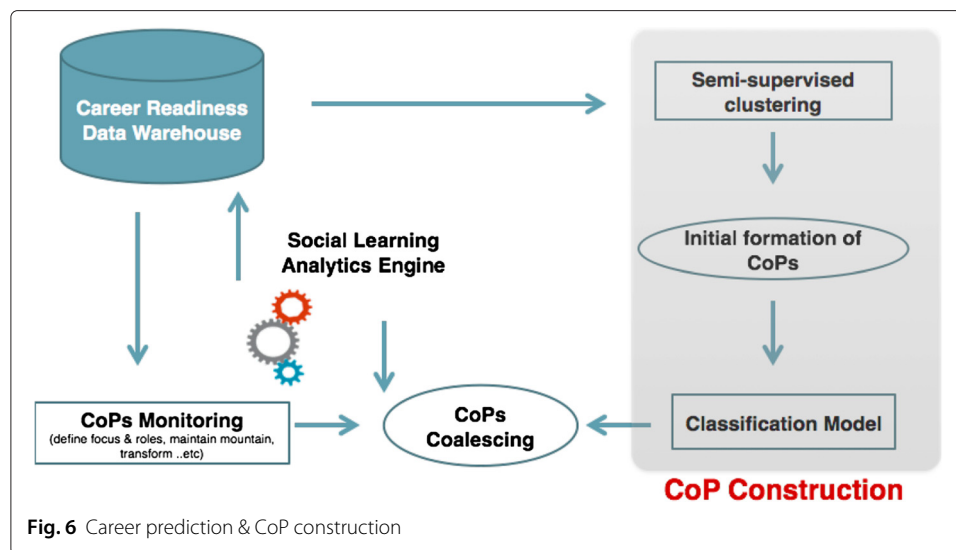
of CoPs. The seed set consists of learners who achieved high scores in career disposition values that are above a given parameter threshold. The collection of career disposition data through a portal structure is the subject of a previous work which we conducted (Atif et al. 2014). There is typically at least one seed member in each cluster (CoP) for which his/her career profile matches the definition suggested by the career ontology that yielded the CoP. The rationale of privileging highly ranked learners in their career disposition to create dedicated CoPs is driven by the prospects to sustain CoPs. From this initial stage, we infer the use of career disposition values only to provide seed set of new CoPs (including the initial ones). To this end, we developed a semi-supervised clustering algorithm detailed in Section Fuzzy semi-supervised clustering algorithm that is based on two of the most common partitioning methods: (1) Seeded K-Means algorithms that use labeled examples to initialize cluster centers (Wang et al. 2011); and (2) Constrained K-Means algorithms that enforce constraints to be satisfied during the clustering assignment; or penalize constraint violations using distance (Davidson and Basu 2007). Both methods are applied using the original unsupervised K-Means algorithm as elaborated further in the next section.

### Career development and SLA

The members' constant interactions within CoP create a dynamic knowledge container and a repertoire of shared practices and experiences. As the community thrives, learners develop their domain pracrices, and may recognize and then reach out other potential members (away from pediatrics) to migrate to other CoPs e.g. nutritionist, psychologist, etc. This gateway accomodates possible changes on Career Profile. However, the evolution of CoPs is outside the scope of this paper as we focus essentially on iniitial career predicitons whereas the career development stage is part of our future work.In this section, we provide a brief description of how this module operates.

The proposed module supports long-term career development utilizing an SLA engine and a CoP management component. SLA engine aims to investigate networking process, roles, properties of ties, relationships and how learners develop and maintain these relationships to support their career development. Specifically, we are interested

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 14 of 23

in measuring user engagement and how they develop from a peripheral participation to centripetal participation in ongoing activities of the community. On other words, measure the interaction volume (e.g. login frequency, duration of login and number of connection) and the size of contribution to the practice resources (e.g. number of contributions, frequency of posts, and average length of posts). We expect learners to develop a changing understanding of practice over time by shifting from knowledge consumption only to knowledge creation through a social interaction process. Moreover, we propose to use an SLA engine to track the development of career dispositions in relation to the set of skills required by the industry for each designated career.

In order for the community to grow and have meaning, the individual members must be motivated to engage with it actively to create and maintain information flow. In this essence, we propose a CoP management system that has three main functions: (1) Define CoP focus and major roles; (2) measure the effectiveness of CoP; and (3) dynamic updates when changes occur in learner's profiles and/or industry needs. For measuring CoP effectiveness, we propose developing a comprehensive set of evaluation measures inspired by: (1) criteria to underpin the CoP of learners in the educational context (e.g. development of learners' reflective experience, encouragement of multidisciplinary knowledge sharing, and support learning through cognitive and practical apprenticeship (Jakovljevic et al. 2013); and (2) fundamental elements of successful online CoP (e.g. knowledge generating interactions, efficiency of involvement, connections to the world, and belonging and relationships) (Wenger et al. 2002).

## Fuzzy semi-supervised clustering algorithm

In a semi-supervised clustering setting, a small amount of labeled data is available to aid the unsupervised clustering process. For seeded clustering, we know the labels of certain objects. These objects are usually used as "seeds" and the clustering then utilizes these seeds to improve the clustering performance. For pairwise constrained clustering, we consider a framework that has pairwise must- link and cannot-link constraints (with an associated cost of violating each constraint) between instants in a dataset, in addition to having distances between the instants. In our proposed clustering algorithm, we assume the followings:

– We have seeds and each class will have at least one seed. The seed labels are always correct.
– We have pairwise constraints, must-links and cannot-links. These constraints could be wrong.
– We allow fuzzy labeling, namely each instance can be in more than one cluster.
– All labels are assigned to both seeds and constraints.

One challenging problem occurs when and whether a violation of the link constraint should be penalized. In traditional semi-supervised clustering algorithms, a violation of the link constraint is always penalized. Now, as we allow the instances to be associated with multiple labels, a constraint can be violated legitimately. For example, as shown in Fig. 7, the must-link between B and C is only within Cluster C2. If we use label C2 for B, and label C3 for C, the must-link can be violated legitimately. On the contrary, for the cannot-link between A and D, there is no way that it can be violated legitimately. Thus,

**Fig. 7** Example of the fuzzy semi-supervised learning
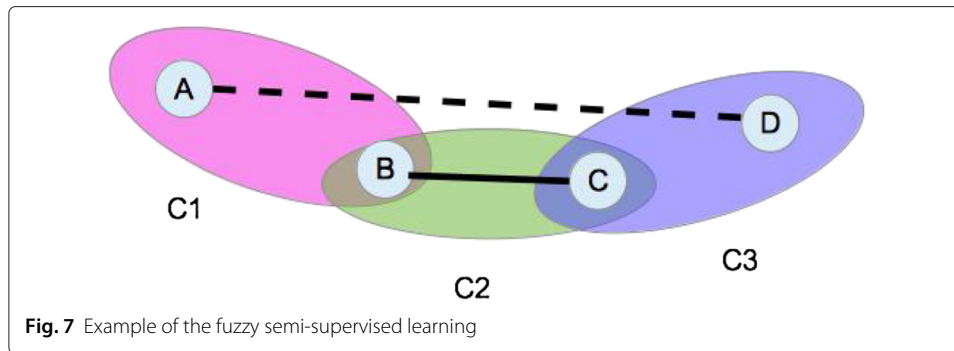
the penalty function needs to be re-designed to allow fuzzy labeling and to estimate if a constraint violation could be legitimate or not.

According to this logic, we developed the Fuzzy Pairwise-constraints K-Means (FCKM) algorithm that is presented in Algorithm 1; while notions and symbols are described in Table 1. The main steps of the FCKM algorithm are ad follows:

1.   Initialize the centroids of each cluster as the average of the seeds belonging to that cluster
2.   Assign instances to minimize the new objective function $O_{new1}$ shown in Eq. (1)
3.   Update the cluster centroids to minimize the objective function as shown in Eq. (2)
4.   Repeat until convergence

For each cluster $C$, we first identify all the seeds $S_{c_1}, S_{c_2}, \ldots, S_{c_t}$ belonging to the cluster. Then we initialize the centroids of each cluster as the average of the seeds belonging to that cluster $\mu_c = \sum_{i=1}^{t} \frac{S_{c_i}}{t}$. As we allow soft-constraints, namely the pairwise constraints could be wrong, we apply a penalty function on each constraint violation. As we showed in the above example, not every violation should receive a penalty. We need to determine when a violation should not receive a penalty. Assuming we are assigning the instance $x_a$, we develop the following new objective function (Eq. 1), which is an updated version of previous works (Davidson and Basu 2007):

**Table 1** Notions and symbols

| Symbol | Description |
| --- | --- |
| X | The input domain |
| C | Number of clusters |
| $\mu_c$ | Initial centeroids of cluster |
| i,j | Indices running over clusters |
| a,b | Indices running over instances or output clusters' labels |
| $x_a$ | Input data instance $x_a \in X$ |
| $y_a$ | Output cluster lable $y_a \in [C]$ |
| $D(x_a, \mu_j)$ | Distance between instance $x_a$ and center of cluster j |
| $C_=$ | Must-link constraints |
| $C_{\neq}$ | Cannot-link constraints |
| $h^* = argmin_h O_{new}$ | Instance assignment that minimally increases the error terms |

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 16 of 23

$$O_{new1} = \frac{1}{2} \sum_{x_a \in C_j} D(x_a, \mu_j)^2$$

$$+ I(label((x_a, x_b) \in C_=) \neq j) \frac{1}{2} \sum_{(x_a, x_b) \in C_=, y_a \neq y_b, y_a = j} D(x_b, \mu_j)^2$$

$$+ I(label((x_a, x_b) \in C_{\neq}) \neq j) \frac{1}{2} \sum_{(x_a, x_b) \in C_{\neq}, y_a = y_b, D(x_a, \mu_{y_a}) < D(x_b, \mu_{y_b}), j = h'(x_b)} D(x_b, \mu_j)^2$$

For instances that are not part of constraints, perform a nearest cluster centroid calculation. For pairs of instances in a constraint, for each possible combination of cluster assignments, the function is calculated and the instances are assigned to the clusters that minimally increases the error term $h^* = argmin_h O_{new}$. $I(A)$ is an indicator function defined as follows: $I(A) = 0$ if $A = True$ and $I(A) = 1$ if $A = False$, and $label(x_a, x_b)$ is the label of the constraint. Thus when a link is violated, we check if its associated label is different from the label that $x_a$ is assigned to. If yes, the violation is not penalized.

Once we assign an instance to a cluster $C_j$, we update the cluster centroid $\mu_j$ as follows (Eq. 2) (Davidson and Basu 2007):

$$\mu_j = \frac{\sum_{x_i \in C_j} x_i + sum_1 + sum_2}{|\mu_j| + total_1 + total_2}$$

$$sum_1 = I\left(label\left((x_a, x_b) \in C_=\right) \neq j\right) \sum_{(x_a, x_b) \in C_=, y_a \neq y_b, y_a = j} x_b$$

$$sum_2 = I\left(label\left((x_a, x_b) \in C_{\neq}\right) \neq j\right) \sum_{(x_a, x_b) \in C_{\neq}, y_a = y_b, D(x_a, \mu_{y_a}) < D(x_b, \mu_{y_b}), j = h'(x_b)} x_b$$

$$total_1 = I\left(label\left((x_a, x_b) \in C_=\right) \neq j\right) \sum_{(x_a, x_b) \in C_=, y_a \neq y_b, y_a = j} 1$$

$$total_2 = I\left(label\left((x_a, x_b) \in C_{\neq}\right) \neq j\right) \sum_{(x_a, x_b) \in C_{\neq}, y_a = y_b, D(x_a, \mu_{y_a}) < D(x_b, \mu_{y_b}), j = h'(x_b)} 1$$

The update rule applies that if a must-link constraint is violated, the cluster centroid is moved towards the other cluster containing the other instance. Similarly, the interpretation of the update rule for a cannot-link constraint violation is that cluster centroid containing both constrained instances should be moved to the nearest cluster centroid so that one of the instances eventually gets assigned to it, thereby satisfying the constraint. Our formal algorithm is formally depicted next.

---

**Algorithm 1** Fuzzy Pairwise-constraints K-Means (FCKM)

---

Input: A dataset $X = \{x_a...x_n\}$ to cluster, $C$ : the number of clusters, $S$ : set of seeds, set of $C_= \{(x_a, x_b)\}$, set of $C_{\neq} \{(x_a, x_b)\}$

Output: A partition of $X$ into $C$ clusters that is a local optima of the Eq. (1).

Method:

1. Initialize clusters: $\mu_c = \sum_{i=1}^{t} \frac{S_{c_i}}{t}$
2. Repeat until convergence:

    (a)  Assign each data point $x_a$ to the nearest cluster $h^* = argmin_h O_{new}$
    (b)  Update centroids $\mu_1...\mu_c$ according to Eq. 2

3. Return $C$ clusters.

---

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 17 of 23

## Performance evaluation

In this section, we show the performance of our algorithm based on simulated artificial data, and compare our results along two K-means candidate methods: (1) Seeded K-Means (SKM); and (2) Pairwise-constraints K-Means algorithm (PKM). In our experiment, we run the three algorithms to obtain a complete seeding set from a sample dataset. We specifically aim to test our algorithm's performance when the overlap degree increases as compared to baseline methods that do not support fuzzy assignments.

### Experiment setup

In order to simulate overlapped clusters, we used CircleCluster function that generates uniformly distributed data within a circle seen as a cluster, as follows:
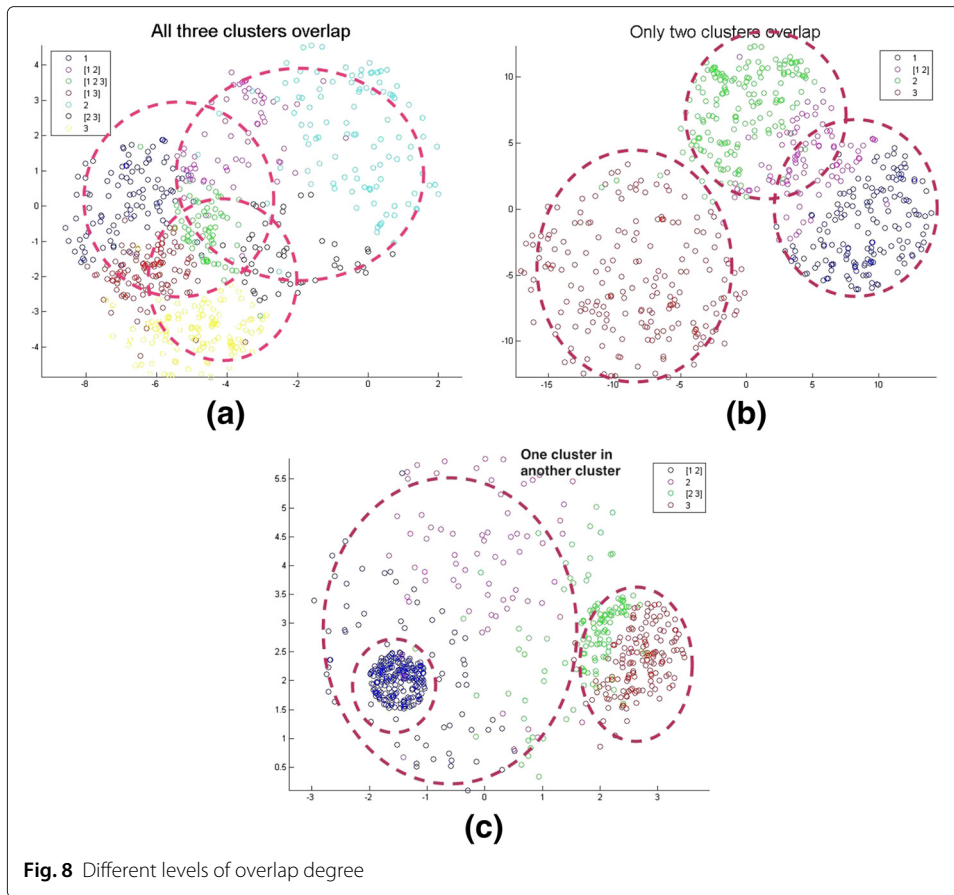
–   Randomly generate the center of the clusters. Then for each cluster, take a radius as input and randomly sample a given number of data points in the circle.
–   To determine if a data point belongs to multiple clusters, consider the distance of the data point to each cluster center. If the distance is no greater than the radius of the cluster, the point belongs to the cluster.

We then simulated a two-dimensional artificial data. The centers of clusters are generated randomly ($\mu = 0$; $\sigma = 1$) within the range, which is a circle with (0, 0) as the center and R = 15 as the radius of the circle in which cluster centers are generated. Then, for each cluster we consider its radius as input and then randomly sample a given number of data points within that circle (following a uniform distribution). To determine if a data point belongs to multiple clusters, we consider the distance of data points to each cluster center. If the distance is no greater than the radius of the cluster, then we consider that the point belongs to the cluster. The generated data set consists of three clusters (C = 3) with 200 samples in each cluster. The constraints used in our algorithm are generated as follows: for each constraint, we randomly pick two instances from the data (following a uniform distribution) and then we check their labels (which are made available for the evaluation purpose but not visible to the clustering algorithm). If they exhibit any common label, we generated a must-link constraint. Otherwise, we generate a cannot-link constraint.

In order to determine the effectiveness of the proposed algorithm and the reliability of the experiment results, we designed three data sets with three different levels of overlapping degree. Figure 8 shows an example of three instances of overlap situations for C = 3 with a) all three clusters overlap, b) only two clusters overlapping, and c) one cluster is entirely within another cluster. The overlap degree is controlled by the radius formula discussed earlier, and which also controls the number of instances within the overlap region from its minimum value in the first set to its maximum in the third set of experiments. For the same number of clusters and overlap degree, we generate different sets of seeds and constraints along the following ratios of the total number of nodes [1 %, 5 %, 10 %].

### Experiment metrics

To evaluate the performance of the clustering algorithms, we employ external metrics that utilize a priori knowledge of the classification information of the data set. External metrics rely on the true class memberships in the data set. For soft clustering, the most

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 18 of 23



**Fig. 8** Different levels of overlap degree

used external evaluation measure is the $F_{Score}$ metric. The $F_{Score}$ is a weighted combination of precision and recall to reflect the overall quality of the resulting clusters. For every resulting cluster *c*, the precision and recall are defined as follows:

- The precision is the ratio tp/(tp + fp) where tp is the number of true positives and fp the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- The recall is the ratio tp/(tp + fn) where tp is the number of true positives and fn the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.

*F_Score* is a combination (harmonic mean) of precision and recall to reflect the overall quality of the resulting clusters. The *F_Score* is defined as follows:

$$F\_Score = (1 + \alpha) \frac{precision \times recall}{(\alpha \times precision) + recall} \qquad (1)$$

Typically, precision and recall are given equal weight with $\alpha = 1$. Varying the coefficient $\alpha$ provides a means of biasing F- score towards precision or recall (e.g., $\alpha = 0.5$ biases it towards precision; $\alpha = 2.0$ biases it towards recall). The total *F_Score* is calculated as the average of the largest *F_Score* of each cluster.

Clustering accuracy is another evaluation measure that discovers the one-to-one relationship between real clusters and the ground-truth categories and measures the extent to

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 19 of 23

which each cluster contained the objects from the co-responding ground-truth category. It is defined as follows:

$$Accuracy = \frac{\left(\sum Truepositive + \sum Truenegative\right)}{\sum Totalpopulation}$$ (2)

### Experiment results

For the the Fuzzy Pairwise- constraints K-Means (FPKM) algorithm, the results showed it achieved higher accuracy than the baseline methods (see Fig 9). This is because the recall of FPKM is generally very high, much higher than those of the baseline algorithms, as the baseline algorithms do not consider overlaps and thus the assignment for the nodes in the overlapped region is relatively random. Many true positives are missed. The recall of the fuzzy algorithm is, however, affected by the degree of overlap: the more the clusters overlap, the lower the recall is. This is obvious because with more overlap, there are more true positives we need to capture and the more true positives the algorithm tends to miss. Thus the recall decreases, and so the overall accuracy, see Table 2.

Figure 10 shows *F_Score* curves when alpha is [0.5, 1, 2] for the three methods as the overlap increases. As the figure indicates, when degree of overlap increases for C = 3, the performance of the fuzzy algorithm becomes better than those of the baseline algorithms. It is noticed that when the overlap degree is low, the performance of our proposed method is less than the baseline methods. This is can be justified by the lower precision value achieved by the fuzzy algorithm. The denominator of the precision is the number of nodes assigned to the cluster. For the fuzzy algorithm, as it considers overlaps, it usually assigns more nodes to each cluster, which makes the denominator larger. However, when overlap degree increases, it is often the case that all three clusters overlap with each other - the baseline methods then tend to make many mistakes which makes the precision poor. We can see the precision of the baseline methods and so the F-score generally drops when overlap degree increases. On the contrary, the fuzzy algorithm returns better precision as overlap increase. This is because the fuzzy algorithm generally tends to assign more nodes to the overlapped region. When the clusters overlap more, more nodes assigned to the overlapped regions are correct, leading to higher precision.
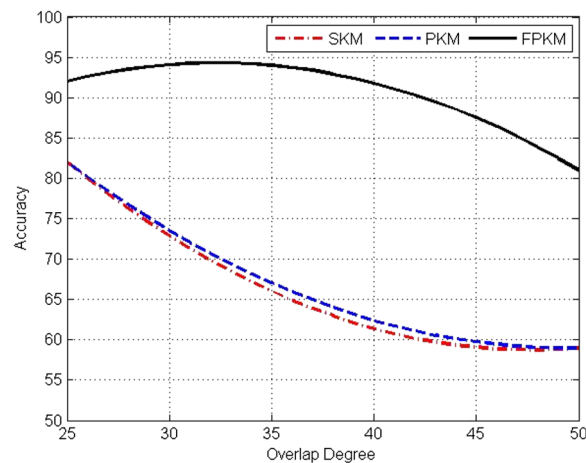


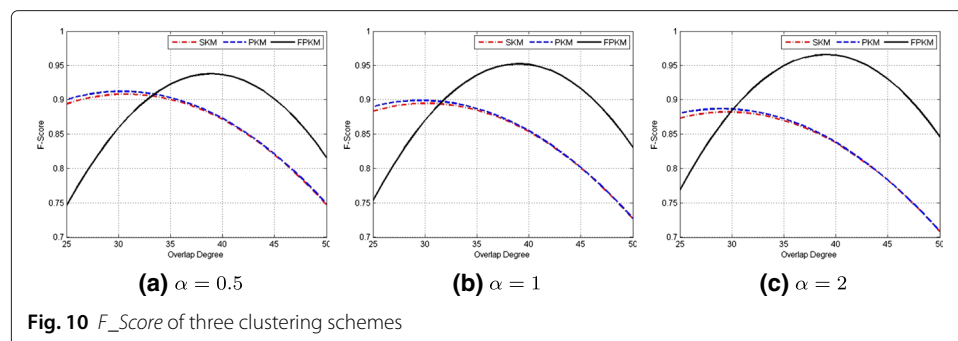**Fig. 9** Accuracy of three clustering schemes

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 20 of 23

**Table 2** Overlap degree vs accuracy of FPKM

| Index | No. of nodes in overlapped region | Accuracy |
| --- | --- | --- |
| 1 | 18 | 0.997778 |
| 2 | 19 | 0.996667 |
| 3 | 33 | 0.99625 |
| 4 | 83 | 0.994074 |
| 5 | 133 | 0.94037 |
| 6 | 180 | 0.752593 |
| 7 | 25 | 0.997778 |
| 8 | 14 | 0.996667 |
| 9 | 6 | 0.996111 |
| 10 | 103 | 0.902593 |
| 11 | 146 | 0.952963 |
| 12 | 173 | 0.914074 |
| 13 | 15 | 0.993889 |
| 14 | 16 | 0.995556 |
| 15 | 10 | 0.997222 |
| 16 | 128 | 0.858889 |
| 17 | 173 | 0.90963 |
| 18 | 188 | 0.767778 |

## Conclusion and future work

In response to the demands to bridge the growing gap between higher education and industry, we introduced a model to incorporate career readiness into formal education to form a new CoP-based learning model which utilizes learning analytics and social networks techniques. The proposed model consists of three major modules: career readiness, career prediction and career development. We first elaborated a learning analytics model to identify career indicators, as well as patterns that contribute to clustering learners into common virtual CoPs. The learners' relationships, engagement and interaction instances within CoPs are tracked using a social learning analytics framework to evaluate the development of domain-related skills under the guidance of an experienced mentor or an active member with superior career dispositions.

We further devised a semi-supervised clustering method to bring learners with similar professional traits that match a typical career pattern together into the same cluster. Our method aims to initially form a CoP with a seed set of learners who can drive the CoP activities and sustain its effectiveness. We emphasized the natural overlap nature of industrial needs and career paths by allowing each leaners to be in more than one



**(a)** $\alpha = 0.5$        **(b)** $\alpha = 1$        **(c)** $\alpha = 2$

**Fig. 10** *F_Score* of three clustering schemes

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 21 of 23

cluster. We experimentally show the improved performance of the proposed clustering approach when the overlap degree increases, in comparison with baseline line methods of seeded and pairwise-constraints K-means algorithm. Hence, our method has the potential to serve as a learning analytics tool to reveal hidden patterns of common traits among learners viewed as future candidates of the job market. These patterns could evolve into social communities of learners with shared career interests, that evolve socially rather than individually. A real data set that includes indicators captured by our career readiness module is expected to prove the concept proposed in this paper as part of our future work.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
All authors read and approved the final manuscript.

**References**
E AbuKhousa, Y Atif, in *Interactive Collaborative Learning (ICL)*, Big learning data analytics support for engineering career readiness (International Conference on, 2014), pp. 663–668. social learning analytics
L Ali, M Hatala, D Gašević, J Jovanović, A qualitative evaluation of evolution of a learning analytics tool. Comput. Educ. **58**(1), 470–489 (2012)
Y Atif, E Abu Khousa, SS Mathew, K Al Awar, N Al Sayari, in *Advanced Learning Technologies (ICALT)*, A portal support to cognitive apprenticeship (IEEE 14th International Conference on, 2014), pp. 449–453
KE Arnold, MD Pistilli, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Course signals at purdue: Using learning analytics to increase student success (ACM, 2012), pp. 267–270
B Baran, Facebook as a formal instructional environment. Br. J. Educ. Technol. **41**(6), 146–149 (2010)
T Bauer, TN Bodner, B Erdogan, DM Truxillo, JS Tucker, Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. J. Appl. Psychol. **92**(3), 707 (2007)
M Blankenship, How social media can and should impact higher education. Educ. Digest. **76**(7), 39–42 (2011)
M Burke, C Marlow, T Lento, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Social network activity and social well-being (ACM, 2010), pp. 1909–1912
RZ Cabada, MLB Estrada, CAR García, Educa: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a kohonen network. Expert Syst. Appl. **38**(8), 9522–9529 (2011)
MA Chatti, AL Dyckhoff, U Schroeder, H Thüs, A reference model for learning analytics. Int. J. Technol. Enhanced Learn. **4**(5), 318–331 (2012)
K Colthorpe, K Zimbardi, L Ainscough, S Anderson, Know thy student! combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated learning. J. Learn. Anal. **2**(1), 134–155 (2015)
S Dawson, JPL Tan, E McWilliam, Measuring creative potential: Using social network analysis to monitor a learners' creative capacity. Australas. J. Educ. Technol. **27**(6), 924–942 (2011)
I Davidson, S Basu, A survey of clustering with instance level constraints. ACM Trans. Knowl. Discov. Data. **1**, 1–41 (2007)
R Davis III, CH Deil-Amen, C Rios-Aguilar, MS Gonzalez Canche, Social media in higher education: A literature review and research directions (2012)
DC DeAndrea, NB Ellison, R LaRose, C Steinfield, A Fiore, Serious social media: On the use of social media for improving students' adjustment to college. Internet High. Educ. **15**(1), 15–23 (2012)
A De Liddo, SB Shum, I Quinto, M Bachler, L Cannavacciuolo, in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, Discourse-centric learning analytics (ACM, 2011), pp. 23–33
R Ferguson, The state of learning analytics in 2012: A review and future challenges. Knowl. Media Inst. Tech. Rep. KMI-2012. **1**, 2012 (2012)
P Gannon-Leary, E Fontainha, Communities of practice and virtual learning communities: benefits, barriers and success factors. Barriers and Success Factors. eLearning Papers. **5** (2007)
K Gray, L Annabell, G Kennedy, Medical students' use of facebook to support learning: Insights from four case studies. Med Teach. **32**(12), 971–976 (2010)
M Group, Talent Shortage Survey Results, Manpower 2010. Supply/Demand: 2010 Talent Shortage Survey Results (2010)
M Group, Supply/Demand: 2013 Talent Shortage Survey Results (2013)
C Greenhow, B Robelia, JE Hughes, Learning, teaching, and scholarship in a digital age web 2.0 and classroom research: what path should we take now? Educ. Res. **38**(4), 246–259 (2009)
W Greller, H Drachsler, Translating learning into numbers: A generic framework for learning analytics (2012)
CN Gunawardena, MB Hermans, D Sanchez, C Richmond, M Bohley, R Tuttle, A theoretical framework for building online communities of practice with social networking tools. Educ. Media Int. **46**(1), 3–16 (2009)
W Hämäläinen, Descriptive and predictive modelling techniques for educational technology. Licentiate thesis, Department of Computer Science, University of Joensuu (2006)
C Haythornthwaite, M De Laat, in *7th International Conference on Networked Learning*, Social networks and learning networks: Using social network perspectives to understand social learning (Aalborg, Denmark, 2010)

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 22 of 23

JF Helliwell, RD Putnam, *et al*, The social context of well-being. Phil. Trans. R. Soc London Series B Biol. Sci., 1435–1446 (2004)

KF Hew, Students' and teachers' use of facebook. Comput. Hum. Behav. **27**(2), 662–676 (2011)

H-T Hung, SC-Y Yuen, Educational use of social networking technology in higher education. Teach. Higher Educ. **15**(6), 703–714 (2010)

A Hwang, EH Kessler, AM Francesco, Student networking behavior, culture, and grade performance: an empirical study and pedagogical recommendations. Acad. Manag. Learn. Educ. **3**(2), 139–150 (2004)

M Jakovljevic, S Buckley, M Bushney, Forming communities of practice in higher education: a theoretical perspective (2013)

R Junco, G Heiberger, E Loken, The effect of twitter on college student engagement and grades. J. Comput. Assisted Learn. **27**(2), 119–132 (2011)

N Kabilan, MK Ahmad, MJZ Abidin, Facebook: An online environment for learning of english in institutions of higher education?. Internet Higher Educ. **13**(4), 179–187 (2010)

S Katz, L Earl, Creating new knowledge: Evaluating networked learning communities. Educ. Canada-Toronto. **47**(1), 34 (2007)

K Kitto, S Cross, Z Waters, M Lupton, in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, Learning analytics beyond the lms: the connected learning analytics toolkit (ACM, 2015), pp. 11–15

I Koprinska, in *EDM*, Mining assessment and teaching evaluation data of regular, advanced stream students (Citeseer, 2011), pp. 359–360

J Lave, E Wenger, *Situated Learning: Legitimate Peripheral Participation*. (Cambridge University Press, 1991)

M Lea, D Barton, K Tusting, Communities of practice in higher education. Beyond communities of practice: Language, power and social context (2005)

D Leony, A Pardo, L de la Fuente Valentín, DS de Castro, CD Kloos, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Glass: a learning analytics visualization tool (ACM, 2012), pp. 162–163

I Liccardi, A Ounnas, R Pau, E Massey, P Kinnunen, S Lewthwaite, M-A Midy, C Sarkar, in *ACM SIGCSE Bulletin*, The role of social networks in students' learning experiences, vol. 39 (ACM, 2007), pp. 224–237

I LIP, IMS Learner Information Package. Information Model, Best Practice and Implementation Guide, XML Binding, Schemas. Version (2001)

E Martín, M Gértrudix, J Urquiza-Fuentes, PA Haya, Student activity and profile datasets from an online video-based collaborative learning experience. Br. J. Educ. Technol (2015)

D Maynard, M Yankova, A Kourakis, A Kokossis, in *ESWC Workshop "End User Apects of the Semantic Web,"* Ontology-based information extraction for market monitoring and technology watch (Heraklion, Crete, 2005)

A Merceron, in *EDM*, Investigating usage of resources in lms with specific association rules, (2011), pp. 361–362

MH Ministries, Pediatric Subspecialty Physician Needs. Online Report (2015)

EW Morrison, Newcomers' relationships: The role of social network ties during socialization. Acad. Manag. J. **45**(6), 1149–1160 (2002)

O O'Brien, M Glowatz, Utilising a social networking site as a learning tool in an academic environment: Advancing practice from information-sharing to collaboration and innovation (ici). AISHE-J: All Ireland J. Teaching &amp; Learn. Higher Educ. **5**(3) (2013)

E Ozgen, RA Baron, Social sources of information in opportunity recognition: Effects of mentors, industry networks, and professional forums. J. Bus. Ventur. **22**(2), 174–192 (2007)

A Pardo, Social learning graphs: combining social network graphs and analytics to represent learning experiences. Int. J. Soc. Media Interactive Learn. Environ. **1**(1), 43–58 (2013)

JM Podolny, JN Baron, Resources and relationships: Social networks and mobility in the workplace. Am. Sociol. Rev., 673–693 (1997)

P21, Up to the Challenge: The Role of Career and Technical Education and 21st Century Skills in College and Career Readiness (2010). http://www.p21.org/storage/documents/CTE_Oct2010.pdf

H Qureshi, IA Raza, M Whitty, Facebook as e-learning tool for higher education institutes. Knowl Manag; E-Learning: Int J(KM&amp;EL). **6**(4), 440–448 (2015)

R Rabbany, S Elatia, M Takaffoli, OR Zaïane, in *Educational Data Mining*, Collaborative learning of students in online discussion forums: A social network analysis perspective (Springer, 2014), pp. 441–466

J Reich, R Murnane, J Willett, The state of wiki usage in us k–12 schools leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. Educ. Res. **41**(1), 7–15 (2012)

C Romero, S Ventura, Educational data mining: a survey from 1995 to 2005. Expert Syst. Appl. **33**(1), 135–146 (2007)

J Seely Brown, Open education, the long tail, and learning 2.0. Educ. Rev. **43**(1), 16–20 (2008)

N Selwyn, Faceworking: exploring students' education-related use of facebook. Learn. Media Technol. **34**(2), 157–174 (2009)

SE Seibert, ML Kraimer, The five-factor model of personality and career success. J. Vocat. Behav. **58**(1), 1–21 (2001)

JR Segedy, JS Kinnebrew, G Biswas, Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. J. Learn. Anal. **2**(1), 13–48 (2015)

A Scott, P Clarkson, A McDonough, Fostering professional learning communities beyond school boundaries. Australian J. Teacher Educ. **36**(6), 5 (2011)

G Siemens, What are learning analytics. Retrieved March. **10**, 2011 (2010)

G Siemens, P Long, Penetrating the fog: Analytics in learning and education. Educ. Rev. **46**(5), 30–32 (2011)

G Siemens, RS d Baker, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Learning analytics and educational data mining: towards communication and collaboration (ACM, 2012), pp. 252–254

R Shum, SB Ferguson, Social learning analytics. Educ. Technol. Soc. **15**(3), 3–26 (2012)

C Steinfield, NB Ellison, C Lampe, Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. J. Appl. Dev. Psychol. **29**(6), 434–445 (2008)

SW Tian, AY Yu, D Vogel, RC-W Kwok, The impact of online social networking on learning: a social integration perspective. Int. J. Netw. Virtual Organ. **8**(3), 264–280 (2011)

Abu Khousa *et al. Smart Learning Environments* (2015) 2:14

Page 23 of 23

K Wagstaff, C Cardie, Clustering with instance-level constraints. AAAI/IAAI. **1097** (2000)

X Wang, C Wang, J Shen, in *Web Information Systems and Mining*, Semi–supervised k-means clustering by optimizing initial cluster centers (Springer, 2011), pp. 178–187

G Wells, G Claxton, *Learning for Life in the 21st Century: Sociocultural Perspectives on the Future of Education*. (John Wiley & Sons, 2008)

E Wenger, *Communities of Practice: Learning, Meaning, and Identity*. (Cambridge University Press, 1999)

E Wenger, RA McDermott, W Snyder, *Cultivating Communities of Practice: A Guide to Managing Knowledge*. (Harvard Business Press, 2002)

P Wertsch, JV del Río, A Alvarez, *Sociocultural Studies of Mind*. (Cambridge University Press, 1995)

B Xu, M Recker, S Hsi, *The data deluge: Opportunities for research in educational digital libraries. Internet Issues: Blogging, the Digital Divide and Digital Libraries*. (Nova Science Pub Inc., New York, 2010)

W Xing, S Goggins, in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, Learning analytics in outer space: a hidden naïve bayes model for automatic student off-task behavior detection (ACM, 2015), pp. 176–183

AY Yu, SW Tian, D Vogel, RC-W Kwok, Can learning be virtually boosted? an investigation of online social networking impacts. Comput. Educ. **55**(4), 1494–1503 (2010)

S Zhang, C Flammer, X Yang, Uses, challenges, and potential of social media in higher education'. Cutting-edge Social Media Approaches to Business Education: Teaching with LinkedIn, Facebook, Twitter, Second Life, and Blogs, 217 (2010)

J Zimmermann, KH Brodersen, J-P Pellet, E August, JM Buhmann, in *EDM*, Predicting graduate-level performance from undergraduate achievements, (2011), pp. 357–358