Mining for Evidence of Collaborative Learning in Question & Answering Systems

Johan Loeckx
Artificial Intelligence Lab
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel
jloeckx@ai.vub.ac.be

ABSTRACT

Question and Answering systems and *crowd learning* are becoming an increasingly popular way of organising and exchanging expert knowledge in specific domains. Since they are expected to have a significant impact on online education [14], we will investigate to which degree the necessary conditions for collaborative learning emerge in open Q&A platforms like Stack Exchange, in which communities grow organically and learning is not guided by a central authority or curriculum, unlike MOOCs. Starting from a pedagogical perspective, this paper mines for circumstantial evidence to support or contradict the pedagogical criteria for collaborative learning. It is observed that although there are *technically no hindrances towards true collaborative learning*, the nature and dynamics of the communities are not favourable for collaborative learning.

The findings in this paper illustrate how the collaborative nature of feedback can be measured in online platforms, and how users can be identified that need to be encouraged to participate in collaborative activities. In this context, remarks and suggestions are formulated to pave the way for a more collaborative and pedagogically sound platform of knowledge sharing.

1. INTRODUCTION

Computer-assisted instruction (CAI) is one of the hottest topics in education research [9] and often claimed to revolutionise how we teach and learn [6]. Massive Open Online Courses or MOOCs are the newest manifestation of this phenomenon. However, while 2012 was being praised as "the year of the MOOC", more and more critical voices were heard during the last year and MOOCs are under increasing pressure to finally live up to their promise. Spoken in terms of of *Gartner's Hype Cycle* [8], we could say that we're either at the peak of inflated expectations, or already entering the *through of disillusionment* [3, 15, 10].

This however does not mean that online learning isn't advancing in many interesting directions: Kahn's academy emerged more or less organically when Salman Kahn started teaching his cousin mathematics using short videos. When Salman realized a lot more children could benefit from these lessons, he started distributing them on YouTube. Today, Kahn Academy reaches 10 million students per month, according to Wikipedia. Wikipedia itself has become an integral part of traditional education too. Some researchers expect that learning in general will evolve from an individual task centred around the teacher-student dichotomy, to a collaborative social activity, in which online knowledge bases like Wikipedia, forums, social networks and Question & Answering systems are playing an ever more important role [4]. In this paper, we will try to find evidence of the claimed collaborative properties of Q&A systems, more in particular the music forum site of Stack Exchange[1]. Though the analysis is based on text-based feedback, it is expected that the dynamics of feedback in collaborative activities also hold in multi-modal situations.

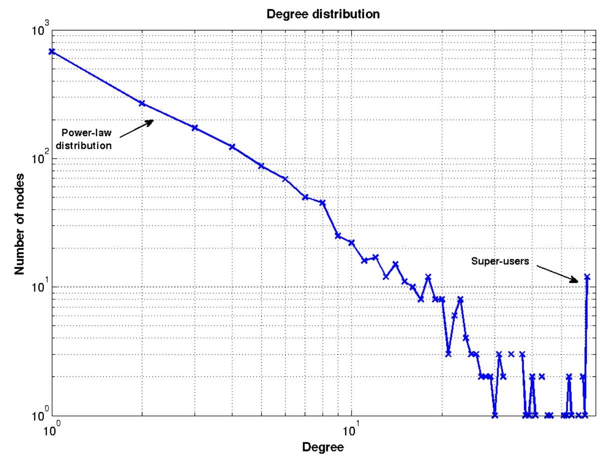

Figure 1: The degree distribution shows that the network of user-interaction is scale-free, which supports the hypothesis that there is no symmetry of knowledge.

This paper is structured as follows. First, the pedagogical background of collaborative learning is set out, based upon the work of Dillenbourg [7] and conditions for and indicators of collaborative learners are introduced. Next,

[1]http://music.stackexchange.com

educational data mining techniques are applied [12] to find evidence of collaborative learning in *crowd learning* systems, more specifically Question and Answering systems like Stack Exchange. Lastly, a critical discussion is performed and suggestions towards more collaborative Q&A systems are proposed, to end with conclusions.

2. COLLABORATIVE LEARNING

2.1 Pedagogical approach

Existing definitions of collaborative learning in the academic fields of psychology, education and computer science, differ significantly and are often vague or subject to interpretation. We thus needed a theory that unified the different theories and was applicable to the online, computerised world as well. Not the least, it had to be easily operationalisable. A review of the literature brought us to the work done by Pierre Dillenbourg [7] that perfectly suited our requirements. Dillenbourg takes a broad view on the subject and argues that collaborative learning is a *situation* in which two or more people *learn* through *interactions*.

This means that *collaborative learning can not be reduced to one single mechanism*: just like people do not learn because they are individual but rather because the activities they perform trigger learning mechanisms, people don't learn collaboratively because they are together. Rather, the interactions between the peers create activities (explanation, mutual regulation,...) that trigger cognitive learning mechanisms (elicitation, internalisation, ...) [7].

For these processes to be effective, some requirements need to be fulfilled. A subset was extracted that could be measured numerically, albeit indirectly, using the information available in our data set (summarized in Table 1). In the next section we will have a closer look at these indicators.

2.2 Indicators

Dillenbourg discriminates three important aspects for collaborative learning to be effective and characterises situations, interactions and processes as *collaborative* if they fulfil the following criteria:

- Peers are more or less at the *same level*, have a *common goal* and *work together*;
- Peers *communicate interactively*, in a synchronous and *negotiable manner*;
- Peers apply mechanisms like *internalisation, appropriation and mutual modelling.*

These high-level criteria have been refined by Dillenbourg into more detailed conditions for collaborative learning, of which a subset has been summarised in Table 1. Each corresponding indicator provides indirect circumstantial evidence for each criterion, as our analysis was limited by the data available in the Stack Exchange. Nevertheless, as we will see, they give useful insight in the formation and dynamics of open online collaborative communities for learning.

The research in this paper can be seen as an extension of previous research in Educational Data Mining, that measured participation and interaction between students [11] and the successful formation of learner's communities [1, 13].

3. QUANTITATIVE ANALYSIS

Stack Exchange can be considered as a distant-learning autodidact platform in which communities are formed organically and learning is not guided by a curriculum or some central authority, but exclusively by the members of the community, in contrast with MOOCs. This paper aims at answering the question whether the necessary conditions for collaborative learning emerge spontaneously in these platforms. As the work is done in the context of the PRAISE project[2], a social media platform for music learning, the Music Stack Exchange data set was chosen.

Stack Exchange provides an open API, from which all data can be exported. The data set consisted of 2400 questions, 1500 active members and 1.7 million page views The platform is basically a forum in which anyone can ask and reply to questions. As a means of quality control, users can give up- and down votes to questions, and answers. People can also comment on questions and answers which is actually some kind of meta-discussion in which feedback on relevance, terminology, etc... is given. In the following paragraphs, the criteria listed in Table 1 will be studied in more detail.

3.1 Symmetry of action

Symmetry of action expresses the extent to which the same range of actions is allowed by the different users. Stack Exchange employs a system of so-called *privileges*, attributed according to your reputation[3]. These privileges are generally connected to *moderation rights*, rather than with the actions of asking and replying to questions – unless you have a negative reputation. The fact that users can exert the same actions, does not imply that this also actually the case. An analysis of the distribution of the ratio of answers over the number of questions, reveals that we can roughly discriminate *three kinds of users*, based upon their activity profile:

- *Silent users (62% of the registered users)* that never answer, e.g. users that don't register or register but do not ask questions nor reply to them;
- *Regular users (37% of registered users)* that give roughly as much as answers as they ask questions, that is, two on average;
- *Super-users (<1% of the registered users)*, these are 'hubs' that give at least 40x more answers than they ask questions.

The largest part (96%) of *regular users*, ask less than five questions, and 76% even asks only one question: there are *no 'parasite' users between the regular users that ask question but do not answer*. From the other side, only 8 'expert' super-users (0.5% of the community) were responsible for answering 25% of the questions. Above findings indicate that **the symmetry in action is highly skewed because of a small group of 'super-users' and a large group of 'silent users'.**

[2]http://www.iiia.csic.es/praise/

[3]http://stackoverflow.com/help/privileges

Aspect	Criterion	Indicator
Situation	Symmetry of action	Ratio of answers and questions per user
	Symmetry of knowledge	Scale-freeness of the user interaction graph
	Symmetry of status	Distribution of reputation within the community
Interactions	Synchronous	Response times of answering to questions
	Division of labour	Distribution of questions and answers in the community

Table 1: Criteria of collaborative learning according to Dillenbourg, with corresponding indicators. The indirect nature of the indicators stems from the fact that only meta data was available from the Stack Exchange data set, and that the criteria in general are very hard to measure quantitatively.

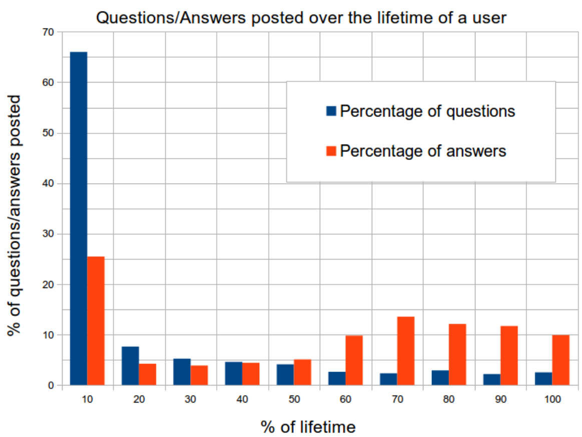

Figure 2: Users tend to ask more questions in the beginning when signing up, and start answering as they have been around some time.

3.2 Symmetry of status

Stack Exchange employs a *reputation system* by which members get rewarded or punished if a peer up- or down votes your answer or question, when your answer gets 'accepted', etc...

We would expect a "healthy" collaborative community to have a strong correlation between reputation and the time a user has been around on the platform: as users spend more time on the platform, their reputation builds up. An inquiry into the Stack Exchange music data set, however, reveals only a correlation of 0.23 between reputation and "time around". We could thus conclude that there is some **odd kind of symmetry, in the sense that no one really builds up reputation**.

3.3 Symmetry of knowledge

Traditionally, these reputation systems are believed to make a good indicator for the knowledge a user possesses. However, there are some problems with this reasoning:

- Knowledge is not a uni-dimensional measure, but is connected to a (sub) domain of expertise;
- Someone's reputation keeps on increasing, even without activity: there is a bias towards old posts and members;
- There is a bias towards "easy answerable questions".

Figuring the knowledge of the members directly is quite an impossible task to perform, especially in a broad and open-ended domain like music. To assess symmetry of knowledge, however, one could argue that *if* everyone in the Stack Exchange music learner's community has more or less the same expertise, *then*, on average, anyone would answer questions asked by anyone.

In other words, there would be no particular hierarchy in answering, rather the network of interaction would be "random" and *not scale-free*. Another way to put this, is to state that *no hubs of people would exist that answer significantly more questions than others*. A network is called *scale-free* if the degree distribution follows a power law[2]:

$$P(k) \sim k^{-\gamma} \tag{1}$$

with $P(k)$ being the fraction of nodes that have a degree k, and γ a constant typically between 2 and 3. Figure 1 reveals a power-law relationship, with exception this special group of "super-users". Above findings therefore suggest that **symmetry of knowledge is not observed.**

3.4 Division of labour

As pointed out before, a small group of super users answer vastly more questions than they ask: a group of 21 users answered half the questions. This is clearly not a balanced situation in which the total labour of answering questions, is equally distributed. Figure 2 shows the relative timing of when users ask and respond to questions over their lifetime.

Users tend to ask questions in the beginning (a visit to the site probably triggered by an urgent need to get a question resolved), but start answering more uniformly after a while. The graph also indicates that engagement is largest in the beginning. This information is relevant when developing platforms with a pedagogical purposes: **users probably need to be "bootstrapped", allowing them to give lesser answers and ask more questions in the beginning, so they get "locked into" the platform**.

Note that a relative plot was preferred, in which the x-axis indicates the % of the lifetime, 0% being the moment of signing up, and 100% the date the data set was obtained. It allowed us to grasp the details of both users that had just signed up, as well as users that have been active for a long time (especially as the rate of signing up is probably not constant but increases with time).

3.5 Synchronous feedback

To keep people engaged in an activity, according to the "theory of flow" [5], immediate feedback is necessary. In the case

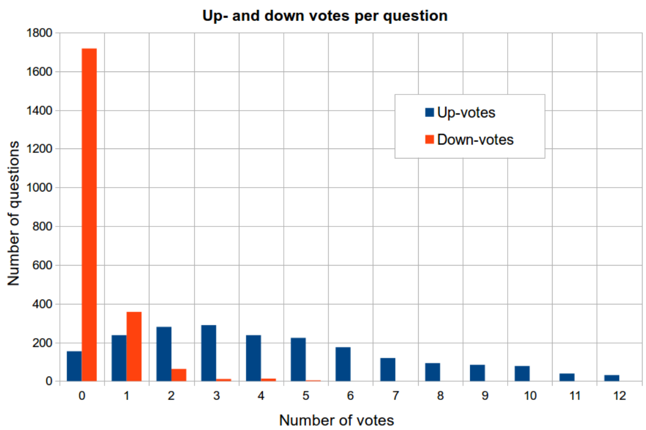

Figure 3: Users tend to give much more up-votes than down-votes to questions. Generally speaking, down-voting is only used to remove off-topic, duplicate questions or questions that are either too specific or broad.

of the music Stack Exchange platform, 68% of the questions received an answer within the day, and 20% even within the hour. This may seem odd, but closer inspection reveals that – once again – this is due to the small-group of ”super-users” that are very engaged.

4. CRITICAL DISCUSSION

Based upon the analysis done in the previous section, some critical remarks and suggestions are offered to improve the pedagogical nature and collaborative learning

4.1 Remarks

4.1.1 *Limited to no instructional design*

The data set on Stack Exchange music's forum, is an amalgam of questions (1) with different levels of granularity, typically with a small scope, (2) on a wide range of topics, for learners (3) with different learning goals and (4) different levels of expertise. The activities are not designed to elicit collaborative learning, and as the data is unstructured, without sufficient scaffolding of the learning content (e.g. through hyper-linking), it is no natural fit for learning but rather provides **ad-hoc answers to appease short-term narrow personal learning goals**.

4.1.2 *A heterogeneous community*

Above remarks wouldn't be so *problematic for collaborative learning*, if proficient communities existed within the Stack Exchange platform that had more or less the same goals, expertise and engagement. In the current case, there's a risk of frustration and boredom in expert users that don't see their questions answered and who have to answer straightforward questions. For novice members, on the other hand, their learning remains limited because they do not get sufficient guidance and do not really construct knowledge.

Although *the group of super-users* makes sure that questions get answered quickly and perform the largest part of moderation, they are *potentially harmful to collaborative learning* as they distort the natural formation and dynamics of collaborative communities. From the other side, their interventions may bootstrap ”young” forums.

4.1.3 *Strong preference for "liking"*

The dataset revealed a *very strong preference for voting up rather than down*: only two users gave more down votes than up votes and of all the people that have ever cast a down vote (72 users out of the roughly 1500 active users), 80% gave more than five times as much up-votes in return. 80% of the questions had *no down vote*, compared to less than 10% without up-vote. Figure 3 shows the distribution of up- and down-votes. This effect was even more pronounced in the answers: the number of down-votes is typically zero or very small, whereas the up-votes reach a maximum at about 3 up-votes, then slowly attenuates. A further analysis of questions with more down than up-votes, revealed that these questions where either off-topic (40%), too vague, broad or specific (35%), not real questions (10%) or Duplicate questions (8%).

4.2 Suggestions

4.2.1 *Sub-communities*

Allowing users to organise themselves in smaller active sub-communities with common or similar learning goals, may prove an elegant solution to manage or exploit the variety in expertise of the users. Also, the concept of reputation would make more sense. A similar idea was proposed by Santos [13].

4.2.2 *Knowledge construction*

Good feedback should provoke critical thinking by asking sensible questions, provide a clue to ”what's next” and allow to construct knowledge through scaffolding and coupling back to acquired knowledge. Though the concept of freely asking questions is very accessible, the content stays rather ad-hoc and unstructured. A way to organise and link different questions in order to guide learners would be very useful.

4.2.3 *Collaborative interfaces*

In the modern ages of web technology, users could benefit from a collaborative interface in which knowledge is constructed together, in a way similar to for example Google Docs where one single entity is shared by all users. So, rather than preserving the strict question/answer or learner/teacher dichotomy, one would go for a situation in which knowledge – not only answers but also questions – is constructed live in an interactive way.

5. CONCLUSIONS

In this paper, the case for collaborative learning in open-ended auto-didact Q&A environments like Stack Exchange is investigated. Based upon the criteria put forward by Dillenbourg, we can state that though there are *technically no hindrances towards collaborative learning, the nature and dynamics of the community that organically form on Stack Exchange, do not support the case for collaborative learning.*

It was observed that the *symmetry of action was distorted* due to a small group of ”super-users” that answered the majority of questions and a large group of ”silent users” that do not really interact with the platform. Inspection of the degree distribution of the user interactions reveals that the community network is scale-free, which means that *symmetry of knowledge is very unlikely*. The reputation system seems insufficient as a measure of expertise and a strange kind of symmetry of status is observed, in the sense that *no one really builds up reputation*, except for a small group of users.

Lastly, the limited possibilities to instructional design, elicits *short-term narrow and personal learning goals*. Also, the very heterogeneous nature of the community is not favourable for learning. Suggestions were made to adapt these interesting and popular platforms to learning, like *creating sub-communities with common learning goals*, extend the possibilities for *organising and structuring the content* and employ *collaborative interfaces*.

As future work, these results should be validated by means of other communities on Stack Exchange as well, and on different modes of feedback, rather than only text-based.

6. ACKNOWLEDGEMENTS

This research has been supported by the EU FP7 PRAISE project #318770.

7. REFERENCES

[1] A. R. Anaya and J. G. Boticario. A data mining approach to reveal representative collaboration indicators in open collaboration frameworks. *International Working Group on Educational Data Mining*, 2009.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] C. G. Brinton, C. Mung, S. Jain, H. Lam, Z. Liu, and F. Ming Fai Wong. Learning about social learning in MOOCs: From statistical analysis to generative model. *arxiv.org*, abs/1312.2159, 2013.

[4] M. A. Chatti, M. Jarke, and D. Frosch-Wilke. The future of e-learning: a shift to knowledge networking and social software. *International journal of knowledge and learning*, 3(4):404–420, 2007.

[5] M. Csikszentmihalyi. *The Evolving Self: A Psychology for the Third Millennium*. Harper Collins, New York, 1993.

[6] L. Cuban and L. Cuban. *Oversold and underused: Computers in the classroom*. Harvard University Press, 2009.

[7] P. Dillenbourg et al. Collaborative-learning: Cognitive and computational approaches. Technical report, Elsevier, 1999.

[8] J. Fenn and M. Raskino. *Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time*. Harvard Business Press, 2008.

[9] J. Hattie. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, 2009.

[10] T. Lewin. After setbacks, online courses are rethought. *New York Times*, (December 11), 2013.

[11] T. P. Padilha, L. M. Almeida, and J. B. Alves. Mining techniques for models of collaborative learning. In *Designing Computational Models of Collaborative Learning Interaction, workshop at. ITS*, pages 89–94, 2004.

[12] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, 2007.

[13] O. C. Santos, A. Rodríguez, E. Gaudioso, and J. G. Boticario. Helping the tutor to manage a collaborative task in a web-based learning environment. In *AIED2003 Supplementary Proceedings*, volume 4, pages 153–162, 2003.

[14] M. Sharples, P. McAndrew, M. Weller, R. Ferguson, E. FitzGerald, T. Hirst, and M. Gaved. Open university: Innovating pedagogy. 2013.

[15] V. Strauss. Are MOOCs already over? *http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/12/are-moocs-already-over/*, (December 12), 2013.