UNIVERSITY OF
BATH

University of Bath

Tutors and gatekeepers in sustainability MOOCS

Danish Mishra, Steve Cayzer, and Tracey Madden

Introduction

For educators concerned with sustainability literacy, we are necessarily both affected by, and effectors of, digital pedagogies. The call for papers for this special issue challenged us to consider whether digital pedagogies are *"supportive of sustainability or perpetuators of unsustainability"*. As might be expected, this question is not a simple binary choice and we have chosen to address it indirectly, by considering the nature of interaction in a global, digitally connected community of learners. In particular, we examine the changing role of tutors in these communities, and the possible implications of this change on sustainable literacy.

MOOCs (massive open online courses) are sometimes credited with the potential to revolutionize distributed learning. Practices such as learner generated resources, system thinking and citizenship education, are characteristic of so-called *connectivist MOOCs*. In such MOOCs, it might be that the educator could take the role of facilitator or even absent themselves entirely from the learning process.

This has implications for sustainability literacy. In our paper we examine the extent to which these patterns of interaction may affect the learning process, and to what extent this may help participants engage with the concept of sustainability. We use Social Network Analysis (SNA) to explore the nature of interaction between learners in a MOOC, particularly the role of the tutors in mediating such interactions. We find that tutors can and do take a central role in some cases. However in other cases the removal of tutor nodes has little effect, suggesting that different modes of learning are possible in a MOOC community. We see participants playing central role in the network as gatekeepers; influencing network learning, learning driven by the participants via conversations among themselves and information flow. The implication for sustainability is that digital pedagogies, when structured correctly, can drive connectivity and engagement.

Related Work

MOOCs (massive open online courses) are sometimes credited with the potential to revolutionize distributed learning. Practices such as learner generated resources, system thinking and citizenship education, are characteristic of so-called connectivist MOOCs. Some authors (Jacobs, 2013; Hew & Cheung, 2014) argue that MOOCs offer a model of democratisation in higher education: courses available to the greatest number of people possible with the lowest barrier to participation. MOOCs differ from traditional classroom learning in their scale and may be change agents. This has potential repercussions for higher education and needs a re-examination of its traditional practices (Gillani & Eynon, 2014; Gašević et al, 2014). Sinha (2014) highlights how MOOCs are prompting participants to rethink learning and it has also been suggested that the real potential is in the new knowledge created through student interaction within MOOCs (Gillani & Eynon, 2014). Traditionally, content delivery has been the focus of education. MOOCs challenge that approach and encourage students to form learning networks where knowledge is constructed by all participants. Traditional roles of student and teacher are challenged (Koutropoulos et al, 2012; Rodriguez, 2013); for instance, in connectivist MOOCs, the educator could take the role of facilitator or even absent themselves entirely from the learning process (Kop, 2011).

Social Network Analysis (SNA) is a key tool to understand interactions in an online environment (de Laat et al 2007) and allows quantitative comparison between different networks and thus between courses (Shen et al 2008). The structure of the network can be pedagogically important: Reuven et al 2003 (quoted by de Laat et al 2007) have found that critical thinking was enhanced in a structured network (rather than an unstructured forum).

In the context of digital pedagogy, there are a number of relevant metrics. We can start by measuring the number of nodes and edges present in a network. This gives a sense of scale, which of course will be a different order of magnitude in a MOOC compared with a traditional classroom environment. MOOCS tend to have low participation as a % of total enrolments, with completion rates around 15% (Jordan 2015). In addition, the vast majority of MOOC participants tend to 'lurk' ; that is, operate in read only mode; Breslow et al (2013) found over 90% of participants were lurkers. Thus *density* (the percentage of all possible edges present) is expected to be low. This gives an initial picture of participation**.** However, as Lipponen et al (2003) point out, high density may be due to one dominant individual. In this way, a teacher's presence may affects the density of a network, perhaps substantially (Martinez et al 2003).

Toikkanen & Lipponen (2011) assert that density is "unrelated to quality or meaningfulness" of online learning. If this is the case then perhaps we need a node centric metric such as **degree** (Rabbany et al 2014; Russo & Koesten 2005) which refers to the number (or relative proportion) of links to a node. The degree can be weighted (e.g. by number of times these nodes have interacted). High degree denotes a node that is highly connected in the network and potentially highly influential; low degree centrality denotes a node that is on the periphery (Wasserman & Faust, 1994) thus SNA can be used to identify isolated participants (Reffay & Chanier 2003). It is common to measure *average degree* over a network or cluster.

Betweenness centrality is a measure which incorporates the importance of nodes as interconnectors. It is measured by

calculating the proportion of geodesic (minimal length) paths between two other nodes in which the node particpates. A node with high betweenness has a large influence on the transfer of items through the network. However this metric is not always correlated with meaningfulness of learning (Toikkanen & Lipponen 2011) nor to course grades (Cho et al 2007).

Calvani et al (2010) advocate measuring **Depth of discussion** – e.g. hierarchic levels in a discussion thread. This approach is clearly only possible where network connections are structured appropriately. Average path length and diameter (Palazuelos et al 2013) are possible alternatives. **Average path length** shows how closely connected nodes are; however this metric can only consider finite path lengths – i.e. connected nodes. So the **number of shortest paths** should also be recorded. **Diameter** is the longest optimal path between nodes. However this is not quite the same thing as depth of discussion (one could loosely claim that diameter measure depth of discussion in participants rather than in messages). What diameter can do is to give some clues about the social nature of discourse. A large diameter implies a potentially loosely connected community; a small diameter may be a very densely connected community, or one in which few connections are present (most nodes unconnected or in small clusters).

These possibilities can be teased apart by looking at the clustering of the graph. Connected nodes tend to form in clusters, which may be weakly or strongly connected. **Weakly connected clusters** are those in which every node is connected somehow to every other (perhaps via other nodes in the cluster). It is common for the entire graph, with the exception of isolated nodes, to form a single weakly connected cluster. However, when link direction is taken into account, this **strongly connected cluster** will lose any nodes without both in-links and out-links to nodes in the cluster. The **clustering coefficient** gives a clue as to the tightness of that cluster, or clique.

The number of connected components is another useful metric, closely related to **modularity,** which is the tendency to form subcommunities. According to the Gephi documentation *"A high modularity score indicates sophisticated internal structure ... into sub-networks (or communities)"* More precisely, modularity compares links within partitions against links between partitions. The result is calculated as a scalar with +1 representing perfect modularity. The algorithm proposed by Blondel et al (2008) starts with individual nodes (ie clusters of size 1) and then performs a hill climbing search, clustering nodes one by one until modularity can no longer be increased by the addition of one more node. The algorithm repeats on the clusters produced by the first pass, and so the process repeats until a local maximum is reached.

A modular structure might result in deeper, or perhaps just more fragmented, community discussion. On the other hand, Toikkanen & Lipponen (2011) find that communities with low hierarchy and few cliques are conducive to learning.

These SNA metrics will be used to build up a rich picture of network structure with and without tutors, and to relate this structure to pedagogic outcomes.

Methods

We focus on the 'Sustainability for Professionals' MOOC delivered by the University of Bath on the FutureLearn platform which hosts the 'Inside Cancer' MOOC, also from Bath. 'Sustainability for Professionals' is pedagogically connectivist, with 'Inside Cancer' being more traditional and instructor led.

The FutureLearn platform provides a dataset containing comments posted by the participants and tutors during the course. These comments are timestamped, with each commentator and comment being assigned a unique identification number. Comments may be either directly associated with course content, or may be a response to another user's comment.

For our analysis we used Gephi, an open source network analysis and visualization software. We used Gephi as the syntax for metrics in Gephi is accessible and can be altered to reflect the desired structural changes as well as new metrics can be accommodated within the Gephi framework. Gephi has become an industry standard for visualising networks and producing a rich variety of metrics (Jacomy et al., 2014; Burns, 2012) .

Using Python (code in Appendix I), we parsed through the FutureLearn dataset to create node and edges file for Gephi. The definition of nodes and edges are as follows:

- A node represents a participant who has posted at least one comment

- An edge represents a response to a comment. The FutureLearn dataset does not support hierarchical responses, so the recipient node is taken to be the owner of the original comment and not any intervening response.

- Edge weight is the number of interactions between 2 nodes (participants). Directed edges are permitted and recorded; for simplicity the results reported here use undirected edges (ie the sum of in-links and out-links).

Table 1 shows number of nodes and edges for each run of the courses.

Table 1: Statistics for different runs of the "Sustainability for Professionals" and "Insider Cancer" MOOCS.

MOOC	Sustainability March 2014	Sustainability January 2015	Sustainability August 2015	Cancer January 2014	Cancer September 2014	Cancer March 2015
Nodes	962	1109	1177	1512	1099	1069
Edges	2312	2279	1822	1215	1255	1141

These files were imported into Gephi, and various metrics (see table 2) were run to create the final dataset containing metric values for each node and the graph as a whole

Table 2: Metrics used for SNA analysis of the MOOCS

Metric	How Calculated
Nodes (Participants)	Simple count
Edges (Interaction among participants)	Simple count
Network Diameter	Longest finite optimal path between nodes using undirected edges
Graph Density	Fraction of all possible undirected edges present
Modularity	Calculated using Gephi algorithm, based on Blondel et al (2008).
Weakly Connected	Number of maximally sized clusters in which each node is reachable from every other node along undirected edges
Strongly Connected	Number of maximally sized clusters in which each node is reachable from every other node along directed edges
Average Degree	Average number of undirected, unweighted edges per node
Average Weighted Degree	Average sum of weights on undirected edges per node
Average Path Length	Average path length (along undirected edges) between all connected nodes
Number of Shortest Paths	Number of optimal routes found between nodes.

Interviews with Tutors

In addition to Social Network Analysis, structured interviews were carried out with the lead tutors on the Cancer and Sustainability MOOCs to contextualise the network analysis and discuss the metrics. The interview questionnaire is attached in Appendix II. Interviews highlight the approach of tutors to the online learning environment and their expectations (or lack of) concerning network learning. Tutors also remark on differences between a traditional pedagogic environments, such as university classroom, and MOOC.

Interviews highlighted tutors' approach to the MOOC and helped contextualise social network analysis findings within the broader framework of pedagogy.
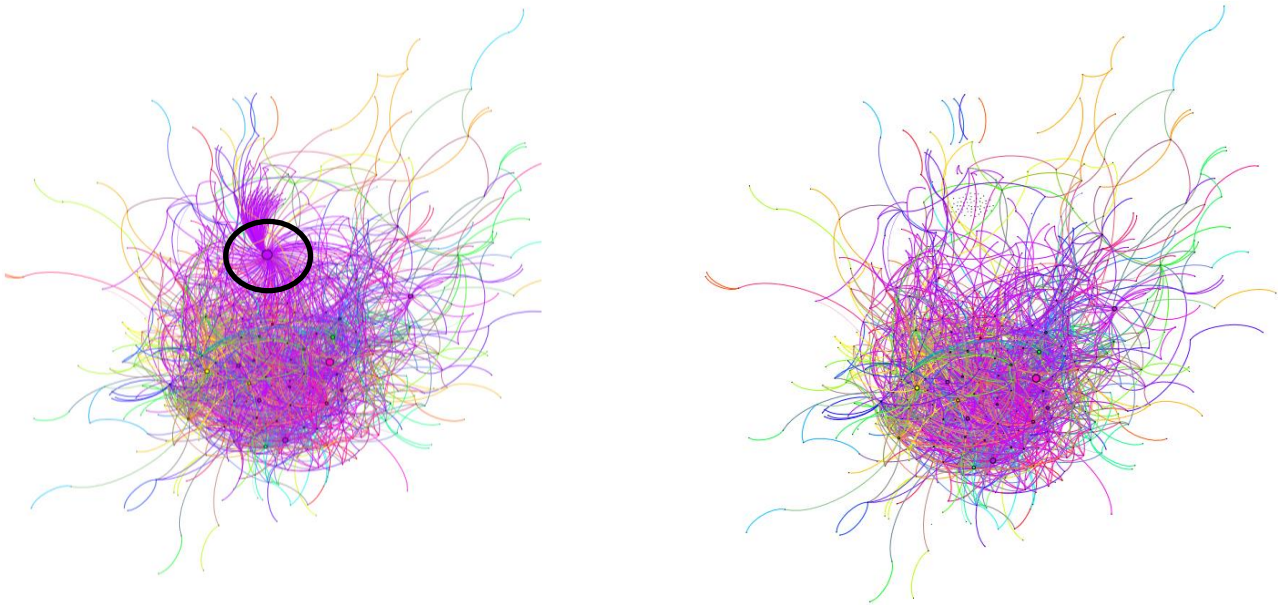
Results

Figure 1 Sustainability March 14 Network Graph with (left) and without (right) Tutors. One of the tutor nodes is circled on the left hand picture. Node size and colour is based on Betweenness Centrality

Figure 1 illustrates the effect of removing tutors from a network. A number of well-connected nodes (including the one circled in the picture) have been removed; and yet the network still appears richly connected. Of course, this is only a visual impression and a deeper inspection can be achieved through the use of SNA metrics as shown in table 3.

Table 3: Metric Table. Figures in brackets are metric value of the network without tutors. Where no figure is given, the removal of tutors had a negligible effect on the value of the metric.

MOOC	Sustainability March 2014	Sustainability January 2015	Sustainability August 2015	Cancer January 2014	Cancer September 2014	Cancer March 2015
Nodes (Participants)	962 (955)	1109 (1103)	1177(1172)	1512 (1503)	1099 (1093)	1069 (1063)
Edges (Interaction among participants)	2312 (1959)	2279 (2165)	1822 (1715)	1215 (1020)	1255 (1126)	1141 (1075)
Network Diameter	10	12	12	14 (12)	11	11
Graph Density	0.003 (0.002)	0.002	0.001	0.001 (0)	0.001	0.001
Modularity	0.305 (0.338)	0.373 (0.367)	0.401 (0.42)	0.501 (0.542)	0.401 (0.422)	0.402 (0.405)
Weakly Connected	356 (423)	470 (480)	586 (589)	972 (1008)	664 (669)	669 (673)
Strongly Connected	764 (786)	887 (891)	987 (985)	1435 (1450)	1011 (1008)	976 (973)
Average Degree	2.403 (2.051)	2.055 (1.963)	1.548 (1.463)	0.804 (0.679)	1.142 (1.03)	1.067 (1.011)
Average Weighted Degree	3 (2.53)	2.705 (2.596)	2.038 (1.933)	0.997 (0.844)	1.669 (1.525)	1.843 (1.751)
Average Clustering	0.048 (0.031)	0.029 (0.028)	0.024 (0.023)	0.01 (0.008)	0.024	0.023

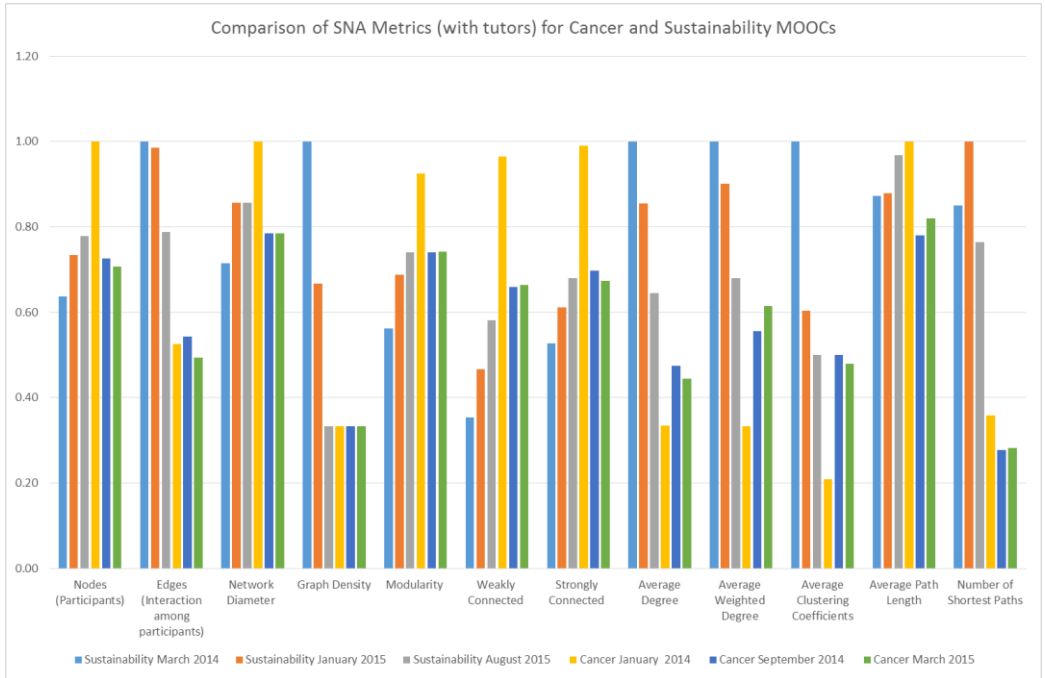| Coefficients						
Average Path Length	4.155 (3.922)	4.189 (4.157)	4.61 (4.644)	4.766 (4.58)	3.719 (3.703)	3.907 (3.937)
Number of Shortest Paths	128329 (95859)	150918 (141036)	115241 (112381)	54160 (37074)	41945 (38210)	42523 (39890)

Fig 2: Comparison of SNA Metrics for Sustainability (1st 3 columns in each group) and Cancer (final 3 columns in each group) MOOCs. The metrics are normalized against the maximum for that particular metric.
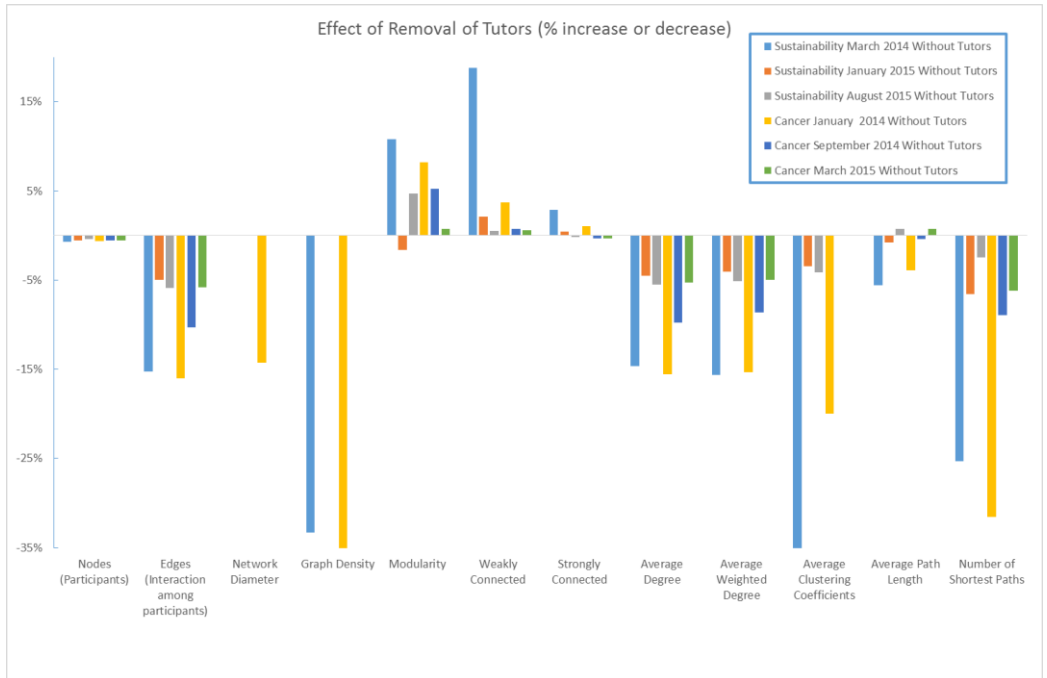

Figure 3. Effect of removal of tutors (% increase or decrease). The bars are arranged as for figure 2. The bar for the density in the first run of Cancer has been cropped from -100%.

Table 3 presents the summary results for the SNA metrics considered. Figure 2 presents the data in a relative form (normalised against the maximum value for that metric across all 6 MOOCs). The number of active participants (nodes) increases for subsequent runs of the sustainability MOOC but with the number of connections (edges) decreasing. A contrasting effect occurs for Cancer, with the number of nodes decreasing and the edge numbers holding steady. Naturally, removal of tutors had a negligible effect on the number of nodes (less than 1%), but the effect on edges was more marked, particularly on the first run.

The density for MOOCs tends to be low with Sustainability having a higher density then Cancer. The removal of tutors (fig 3) appeared to have little effect on overall density; we observed an appreciable effect only in the first run of each MOOC where removal of tutors from the network resulted in more than 33% drop in density. In case of degree, we observed that Sustainability MOOCs had higher average and weighted out-degree than Cancer MOOCs. Again, subsequent runs appeared to have opposite effects, with degree increasing in Cancer MOOC but decreasing in the Sustainability MOOC. Nevertheless, there appears to be some persistent difference between the MOOCS, since a higher proportion of nodes have high betweenness centrality in the Sustainability MOOC (figure 4).
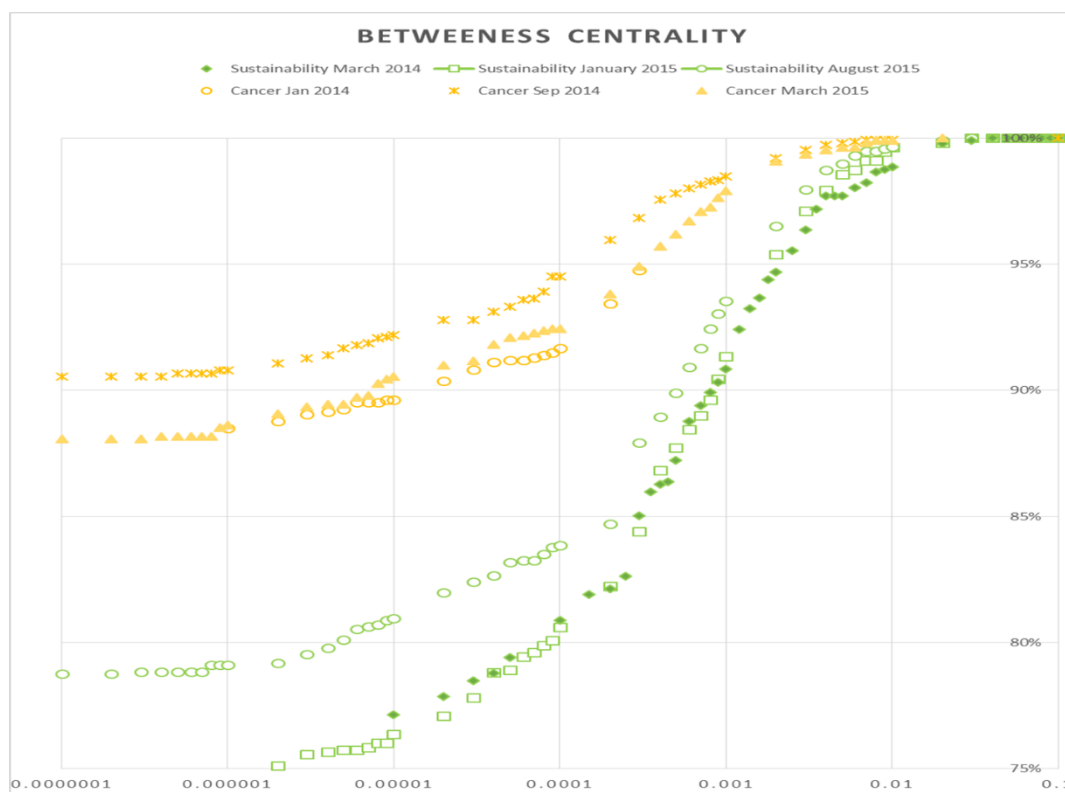

Figure 4. Cumulative distribution of betweenness centrality for three runs of the Sustainability and Cancer MOOCS. Betweenness centrality is plotted on a log scale, and the graph is expanded to show the top 25% participants ranked by betweenness centrality.

The number of shortest paths for Sustainability MOOCs are almost double of those available in Cancer MOOC courses. While removal of tutors substantially reduces the number of shortest paths (particularly on the first run), the difference between the MOOCs persists. The network diameter increases on subsequent runs of the Sustainability MOOC, but decreases on subsequent runs of the Cancer MOOC. A similar pattern is shown for the average path length. Only on the first run of both MOOCS (and for number of shortest path lengths, just the Cancer MOOC) did the removal of tutor nodes have an appreciable effect.

Looking at connected components we found two distinct trends. First, in both courses, there are almost twice as many strongly connected components as weakly connected components. Second, cancer courses have higher number of connected components, both weakly and strongly. This difference reduces with subsequent runs to the extent that it is barely noticeable on the third run. Modularity also displays the same trend. The first run of the Cancer MOOC has a much higher modularity compared to its Sustainability cousin, but this difference disappears by the third run of the course. The clustering coefficient exhibits the same trend as modularity but in reverse (with values increasing rather than decreasing and vice versa). Both

modularity and number of connected components are increased (and clustering coefficient decreased) by removal of tutors, particularly on the first run of the Sustainability MOOC.

Interviews with Tutors

The lead tutor for Cancer MOOC remarked on the nature of interaction among learners and how the level of engagement was "*surprising*". She also commented on the role of tutor in *clarifying* and *verifying* information posted by learners. The Cancer MOOC is a content driven course and needs tutor intervention on matter of technical knowledge and veracity of information within the network. The tutor felt that the MOOC offers more robust engagement for students in terms of involvement of patients, experts, and consultants when compared to the traditional classroom environment. The tutor also highlighted a different approach to aid student learning in MOOC with focus on *end of the week wrap-ups* to assist student learning.

The lead tutor for Sustainability MOOC pointed out design of the course as critical factor for an online learning environment and how it leads to greater participation. In case of Sustainability MOOC, the lead tutor already had experience of distance learning programme and found that helpful in delivery of the MOOC. Though MOOC is similar to online distance learning programmes, MOOC generates more positive feedback and is more engaging and enjoyable for the tutors. MOOCs also appear to have more potential for network learning compared to distance learning as students involvement appeared to be more. MOOC learners have distinct set of motivations compared to traditional pedagogic environment learners, such as university students. The tutor added that students in university may be more driven by "*value for money*" aspect of the course whereas MOOC participants with their different backgrounds and professions are more interested in learning for the sake of it. The tutor also pointed out how conversations in MOOCs are initiated by highly confident people before spreading through the network.

Discussion

Social Network Analysis (SNA) is a key tool to understand interactions in an online environment (de Laat et al 2007) and allows quantitative comparison between different networks and thus between courses (Shen et al 2008). The structure of the network can be pedagogically important: Reuven et al 2003 (quoted by de Laat et al 2007) have found that critical thinking was enhanced in a structured network (rather than an unstructured forum). Table 5 shows a comparison between sustainability MOOC and a traditional pedagogic environment course.

Table 4: comparison between sustainability MOOC and a traditional pedagogic environment course (Palazuelos et al 2013).

Characteristic	Introduction to Multimedia Methods	Making an Impact: Sustainability for Professionals
Size	~70	962 - 1177
Density	0.05 - 0.07	0.001 – 0.003
Average degree	1 - 2	0.5 – 1.5
Diameter	11-19	10-12
Components	1-3	356 - 586
Average size of component	20-70	2-3 ! BUT largest 550-650 nodes

Although the number of participants did vary across MOOC and occurrence, the number of active nodes was usually between 900 and 1200 (with the first run of Cancer picking up 1500 active participants), thus allowing comparison between the datasets.

Density for MOOCs tend to be low, perhaps because a majority of MOOC participants are predominantly in a 'read only' mode. Density can be problematic as the presence of even one dominant node can greatly skew its value. We observed that trend only in the first run of Sustainability and Cancer MOOCs (this aligns with the conclusions of Martinez et al 2003.).

Average degree and diameter are perhaps slightly reduced in MOOCS; however this difference is negligible in comparison to the explosive growth of clusters, many of which are in fact isolated nodes. This suggests that the connectivity in MOOCS is worth exploring further.

One common finding is that the Sustainability MOOCS are better connected that the Cancer MOOCs, with higher numbers of edges, degree, clustering coefficient and number of shortest paths. These metrics, taken together, suggest that Sustainability MOOCs networks are more 'dense' than Cancer MOOCs. Sustainable MOOC networks appear to be sparse but with dense neighbourhoods and high number of shortest paths. Toikkanen & Lipponen (2011) have shown in their paper that students with high in and out degree have a more meaningful learning experience.

Betweeness centrality presents an interesting picture. Frequency distribution (figure 4) curves suggest that sustainability course nodes usually have a wider distribution and higher values for betweeness centrality metric. This suggests a gatekeeper phenomenon where certain nodes are important for network wide interactions, whereas in case of cancer courses, communication is more widely spread and not dependent on certain critical nodes. This trend holds on removal of tutors from the network suggesting a more participant driven community learning environment. It can be again attributed to the course content where Cancer MOOC resembles a more traditional pedagogic approach where Sustainability MOOC is designed to encourage network learning. *We define network learning as the learning driven by the participants via conversations among themselves.* In network learning; participants acting as information brokers attract other participants based on their reputation and are hubs for flow of information flow (Kop, 2012)

The lead tutor interviews confirm these intuitions; Cancer is more traditional in its structure and involves tutor involvement on technical subject matter. In the case of Sustainability, the lead tutor remarked on how network learning was facilitated via the design of the course.

On subsequent runs of the MOOC, the MOOCs become more similar in terms of modularity, connected components and degree. However the number of shortest paths is still appreciably higher in the sustainability MOOC. A common finding was that the effect of tutor removal was ameliorated in subsequent runs. Particularly in the case of clustering, modularity and connected components, tutor removal goes from having a very strong effect to an almost negligible effect. This suggests a move to network learning in which the critical nodes (or gatekeepers) are participants and carry on the conversation irrespective of the presence of tutors. This tendency of MOOCs to adapt and encourage network learning even when the course is not specifically designed to do so stands out in our study.
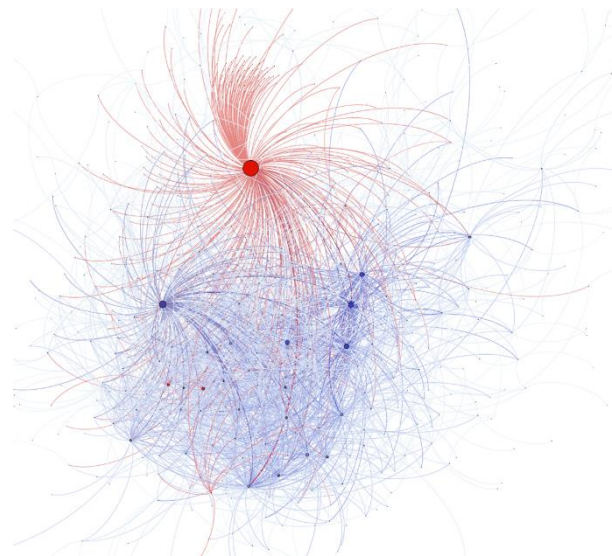

Figure 5: Example of a gatekeeper; the highlighted node (in red) is highly connected but is not a tutor. Example taken from Sustainability MOOC

Studies done using social network analysis (Xie et al., 2011) have shown how these highly influential gatekeepers or information brokers, who themselves resist external influence, are committed to spreading their point of view and when they reach a inflection point, their 'followers' adapt their point of view. Boyd et al. (2010) describe how these gatekeepers filter the information and resources that reach to other participants within the network. These gatekeepers do not seem to have responsibility to validate the information that they feed into the network unlike a tutor or information gatherer might. This is where digital pedagogies adapt a contrasting approach to that of traditional learning environments.

However this does not mean that tutors are *not essential* to such a learning environment. Tutors play a valuable role in delivery of the course and when approached by participants, they can play an important role in validating the information that is being filtered via the gatekeepers. Tutors' limited role in initiating conversations within the network is due; in parts to the design & content of the courses and composition of the network participants. A MOOC course designed with focus on

professionals (Sustainability) will encourage network learning with minimum input from tutors whereas a MOOC designed with highly technical details with a focus on experts and students in the field (Cancer) will behave differently and interactions within such a network will not be on a similar scale as can be seen in table 3. Participation, as shown by number of edges, in Sustainability is greater than in Cancer. The importance of design is highlighted in the interviews with the tutors as well.

MOOC networks differ depending on the subject matter and have a tendency to adapt to the course content. As shown in our analysis, most metrics tend to similar values with each subsequent run of the course. MOOCs offer a network learning approach based on connection, communication, and collaboration. Participants build connections and communities leading to collaborations and information flows. Role of participants, as shown in our analysis, is critical and in contrast with traditional pedagogies, it is the participants, especially those who act as gatekeepers that direct the learning process and access to resources.

This implies a different type of learning is possible in a MOOC community where the network of learners, instead of the tutor, facilitate learning. For sustainability learnings, if the course is designed and delivered in a manner consistent with the medium, digital pedagogies are an effective medium. The role of tutor in such a case is limited to the delivery of the '*lecture*' and participants learn via network learning where gatekeepers act as hubs for information flow and greatly influence the network.

Our analysis shows how connectivist MOOCs can play a constructive role in sustainability pedagogy. Sustainability courses that are *designed* and *delivered* in MOOC environment can lead to network learning with knowledge sharing and conversations being driven by participations. Participations in connectivist MOOCs tend to form communities with certain individuals playing the role of gatekeepers. As our analysis highlighted, these gatekeepers are distinct from tutors who play a limited role in driving the conversation. This type of 'network learning' where participants take a leading role has implications for sustainability pedagogies.

Limitations and Future Implications

Our study is based on analysis of two MOOCs, Sustainability for Professionals and Inside Cancer, run by University of Bath. We analysed three runs of each course and thus limited in sample size. Although we believe our findings have high *generalizability,* our analysis was not carried out in a *controlled environment*. Platform for MOOC represent another limitation to our analysis. We made use of FutureLearn platform which is one of many platforms available for MOOCs and it is possible that different platforms will have different effect on network learning.

A major limitation of study is we had no way to interact with participants of the course after they had finished the course. To derive most robust conclusions, students need to be interviewed on their experience of the MOOC and how it facilitate/obfuscate the learning process. Our study's pedagogical implications are limited by the lack of a controlled environment and sample.

Future studies can elaborate on the role of gatekeepers and their motivations with a focus on how it affects network learning. To better understand the role of gatekeepers, studies need to be conducted in a controlled environment with the same population sample and size, involving subject matter delivered in a traditional pedagogic environment and via MOOC. Gatekeepers, participants, and tutors would need to be interviewed at the end of both runs alongside the social network analysis to arrive at conclusive findings on the role and motivations of gatekeepers and difference between MOOCs and traditional classroom approaches.

Conclusion

Our study has implications for sustainability literacy. We examined the extent to which patterns of interaction in the network affect the learning process, and how this can help participants engage with the concept of sustainability. We used social network analysis to explore the nature of interaction between learners in a MOOC, particularly the role of the tutors in mediating such interactions. We found that tutors can and do take a central role in early runs of the MOOC however with the subsequent runs the removal of tutor nodes has little effect, suggesting that different modes of learning driven by participants are possible in a MOOC community. This type of network learning has implications for sustainability.

In our study we found that MOOCs are different in their network structure but tend to adapt to the subject matter. Digital pedagogies for sustainability result in a qualitative as well as quantitative change in learning where course design affects the learning process and gatekeepers are critical for information flow. These gatekeeper are distinct from tutors in the network. In such a network, tutors' role is limited to course delivery and verifying, depending on course content, the information within the network. Our analysis shows that network learning is dependent on course design and content and gatekeepers exercise influence over the information within the network.

References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. http://doi.org/10.1088/1742-5468/2008/10/P10008

Boyd, D., Golder, S., & Lotan, G. (2010) "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter." HICSS-43. IEEE

Burns, A. (2012). "How long is a tweet? Mapping dynamic conversation networks on twitter using GAWK and GEPHI. Information, Communication & Society. 15(9).

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles, and learning performance in a CSCL community. Computers & Education, 49, 309–329.

Gillani, N. and Eynon, R. (2014) Communication Patterns in Massively Open Online Courses. The Internet and Higher Education 23: 18-26.

Gasevic, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. The International Review Of Research In Open And Distributed Learning, 15(5)

Hew, KF & Cheung, WS (2014) Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. Educational Research Review, Volume 12, pp 45-58

Jacobs, A J (2013) Two cheers for web u!, New York Times, vol 162, no 56,113, pp 17, available at: http://massiveopenonlinecoursesrepository.yolasite.com/resources/13_04-20-13-J.%20   Jacobs%20-%20Two%20Cheers%20for%20Web%20U.pdf

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLoS ONE, 9(6), e98679. http://doi.org/10.1371/journal.pone.0098679

Kop, R. (2011). The Challenges to Connectivist Learning on Open Online Networks: Learning Experiences during a Massive Open Online Course.

Kop, R. (2012). The Unexpected Connection: Serendipity and Human Mediation in Networked Learning. Educational Technology & Society, 15 (2), 2–11.

Koutropoulos, A., Gallagher, M. S., Abajian, S. C., deWaard, I., Hogue, R. J., Keskin, N. Ö., & Rodriguez, C. O. (2012). Emotive vocabulary in MOOCs: Context & participant retention. European Journal of Open, Distance and E-Learning. Retrieved from http://www.eurodl.org/?p=Special&sp=init2&article=507

Laat, M. de, Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. International Journal of Computer-Supported Collaborative Learning, 2(1), 87–103.

Palazuelos C., Garcia-Saiz D., Zorrilla M. 2013 Social network analysis and data mining: An application to the E-learning context Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8083 LNAI   651-660

Rabbany, R., Elatia, S., Takaffoli, M., & Zaïane, O. (2014). Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective. In A. Peña-Ayala (Ed.), Educational Data Mining SE - 16 (Vol. 524, pp. 441–466). Springer International Publishing. http://doi.org/10.1007/978-3-319-02738-8_16

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. European Journal of Open, Distance and E-Learning. Retrieved from http://www.eurodl.org/?p=Special&sp=init2&article=516

Russo T.C., Koesten J. 2005 Prestige, centrality, and learning: A social network analysis of an online class Communication Education 54 3 254-261 Reuven, A., Zippy, E., Gilad, R., & Aviva, G. (2003). Network analysis of knowledge construction in asynchronous learning networks. Journal of Asynchronous Learning Networks, 7(3), 1–23.

Shen D., Nuankhieo P., Huang X., Amelung C., & Laffey J. (2008) Using Social Network Analysis to Understand Sense of Community in an Online Learning Environment, Journal of Educational Computing Research 39: 17-36

Sinha, T. (2014b). "Supporting MOOC Instruction with Social Network Analysis". arXiv preprint arXiv:1401.5175

Toikkanen, T., & Lipponen, L. (2011). The applicability of social network analysis to the study of networked learning. Interactive Learning Environments, 19(4), 365–379. http://doi.org/10.1080/10494820903281999

Xie, J., Sreenivasan, S., Korniss, G., Zhang, W., Lim, C., & Szymanski, B.K. (2011) Social consensus

through the influence of committed minorities. Physical Review E 84(1):011130

Corresponding author

Steve Cayzer can be contacted at: s.cayzer@bath.ac.uk

Appendix I
Python Code

```python
import csv
import collections
import argparse

parser = argparse.ArgumentParser(description='Restructure MOOC CSV.')
parser.add_argument('input_csv', type=open)
args = parser.parse_args()

data_in = {}
with args.input_csv as csvin:
    reader = csv.reader(csvin)
    next(reader)  # skip titles
    for cid, author_id, parent_cid in reader:
        data_in[cid] = (author_id, parent_cid.strip())

data_out = collections.defaultdict(int)
for source_author_id, parent_cid in data_in.values():
    if len(parent_cid) > 0:
        data_out[(source_author_id, data_in[parent_cid][0])] += 1

with open('output.csv', 'w') as csvout:
    writer = csv.writer(csvout)
    writer.writerow(['source', 'target', 'weight'])
    for k, v in data_out.items():
        writer.writerow(list(k) + [v])
```

Appendix II
Interview Questionnaire

- How would you describe your role in the MOOC?
    - (Follow up) Does your role in the MOOC differ from your role in the classroom?
- How would you describe the extent of your involvement in the community of students?
    - (Follow up) Did you find your involvement levels changing with the progression of the course?
    - (2nd Follow up) How does it differ from the traditional classroom environment?
- Did you take on the same role in each run of the course?
    - (Follow up) If yes, why? Does it differ from the classroom?
    - (Follow up) If no, why?
- How would you describe the network learning and involvement of the participants?
    - (Follow up) Did you find the levels of participation changing with the course progression?
    - (2nd Follow up) Does this stand in contrast with the classroom setting?
- What were your expectations regarding student participation and network learning before the course commenced?
    - (Follow up) Did these expectations change with the subsequent runs of the course?
    - (2nd Follow up) If yes, why?