

NetworkSeer: Visual Analysis for Social Network in MOOCs

Tongshuang Wu*

Yuan Yao†

Yuqing Duan‡

Xinzhi Fan§

Huamin Qu¶

The Hong Kong University of Science and Technology

ABSTRACT

The rising trend of MOOCs has attracted wide ranging research interests. Among all the existing studies related to MOOCs, most of them focus on individuals' study behaviors and evaluations (e.g., analysis on click streams for video-watching behavior exploration, etc.) for course design purposes. However, in addition to traditional course materials, MOOCs also provide interactive user forums to encourage students to seek help from peers, which endows the courses with social network formation and interaction. Thus, we present NetworkSeer to help evaluate why MOOC students use forums, and what they do. NetworkSeer visualizes interactions in the forum, including where, when the interactions happen, and why. It also enables filtering out un-targeted groups. A case study is conducted to demonstrate its usefulness.

Index Terms: K.3.1. [Computers and Education]: Computer Uses in Education—Distance Learning

1 INTRODUCTION

Massive Open Online Courses (MOOCs) are online courses that invite large numbers of students to freely enroll. While MOOCs have been recognized for globalizing high-level education, their scalability makes it difficult for the large majority of students to participate in frequent and meaningful interactions with the instructors. Forums have thus been implemented as the primary mode of interaction in MOOCs to compensate for this drawback by encouraging students to seek help from peers, exchange their views, and discuss assignments. Their major role in student collaboration makes them valuable sources for analyzing student learning states: Questions such as how do students with different backgrounds share interactions with each other or how do students' opinions change over time are essential to understanding how students engage in the courses apart from compulsory course activities. The forums help instructors revise their course designs to better fit student needs.

It has been suggested that an organized forum provides a sense of community and engagement [9], and that analyzing the posts in the forum could objectively reflect students' learning features [24]. Therefore, various studies have covered analysis of student participation in MOOC forums, aiming at evaluating its correlations with student performances [4], engagements [18], etc. However, most studies focus on particular aspects of forum participation. For instance, studies would specifically solve problems about the correlation of interactions and performances. However, instructors may wish to understand the interactions from more comprehensive aspects. For instance, they may wish to know how interactions evolve over time, while the current mainstream deals with static social network analysis. On the contrary, visualization is more beneficial,

since instructors could obtain hidden patterns by observing the visual displays. Currently, though various visual analysis for MOOCs has received positive feedback (e.g., [21, 3]), few of them focus student behaviors in the forums.

This study is therefore proposed for more comprehensive visual analysis of student forum participation. Specifically, we design NetworkSeer, which evaluates how students with different demographic characteristics contribute in the forum on a temporal basis, and how they utilize this collaborative study channel to interact. Our contributions could be summarized as follows:

- A new interactive visual design for MOOC forum analysis, which breaks away from the current convention of mere statistical analysis and makes the observations more intuitive. Its interactive views could provide general instructors with easier explorations to student forum behaviors.
- A comprehensive understanding of student forum behaviors. By combining students' time-oriented behaviors and course performance with social interactions, we discovered various factors that motivated students to share interactions with each other. Besides basic understanding, instructors could design courses to stimulate peer collaboration over the long term.

2 RELATED WORK

2.1 MOOC Forum Analysis

Various work has been done analyzing student forum participation in MOOCs. A specifically related field is about the inter-student interactions. Reffay and Chanier [20] computed cohesion with social network analysis (SNA) to highlight isolated people, active subgroups and various roles of the members in the group communication structure. Sinha also processed SNA to identify students who were actively participating in course discussions and those who were potentially at risk of dropping off [22]. Besides these, there is also work on students' online communication activities that are not from MOOC forums. VISM [10] visualizes the sequences of user interactions and subgroup formulations in a radial tree layout, with the arrow between nodes encoding the information flows. ViMoodle [1] presents both the semantic content with word clouds, and a map of relationships and frequencies of activities among students and instructors. However, there was no temporal information embedded in either VISM or ViMoodle. Another course management system, CourseViz [13], presents a 3D plot displaying the discussion posts in conjunction with the topics, participants, and date of post in different axes. While it helps instructors quickly grasp information about social and behavioral aspects of students, its 3D view compromises its scalability as the forum grows in its popularity and the students matrix can be too large to perceive.

We also deal with interactions in our work. It will focus specifically on MOOC forums, and will be designed to be compatible for massive data. Besides, our NetworkSeer will be more comprehensive in that it not only allows users to evaluate interactions with respect of students' performances or reputations, but also supports pinpointing interactions happened in specific time periods.

2.2 Social Network Visualization

Networking visualization has been thoroughly studied, both for individual's network (e.g., Soyent [6] which visualizes temporal fluctuations in a single user's email) and sociological research (e.g., co-

*e-mail: twuac@stu.ust.hk

†e-mail: yyaoaa@stu.ust.hk

‡e-mail: yduanaa@stu.ust.hk

§e-mail: xfanac@stu.ust.hk

¶e-mail: huamin@cse.ust.hk

authorship analysis [16], social network [8], etc.). [14] reviewed research topics for network visualization. Besides traditional social graphs, researchers have also attempted to integrate auxiliary information for more comprehensive analysis. Xu et al. [25] explored set relations and revealed three relationships simultaneously: The social relationship of people, the set relationship among people's items of interest, and the similarity relationship of the items. SocialAction [17] integrates both statistics and visualizations to enable clearer discovery of social networks. PivotPaths [5] explores faceted and interconnected resources of different aspects of the data. Heer et al. proposed Vizster [7], which combines the node-link social graph with profile attributes and automatically identified community structures. Moreover, [19] went beyond the investigation of networks at a single time point and evaluated the emergence of communities with dynamic graphs. We also link social network with additional information of sub-forums to comprehensively understand the formation of the social network in MOOC forums.

3 TASK AND REQUIREMENT ANALYSIS

Forum participation has been recognized in the literature as one of the major engagements and the primary support in online learning [12]. The MOOC instructors in our university, with whom we have been collaborated since 2014, also confirmed their interests in students' forum behaviors. In our interviews with respect of their needs for analyzing student learning behaviors, they have frequently expressed their curiosity about how students interact with each other. Specifically, their interests could be summarized as:

- T.1 How do students treat the others they interact with?** The forum supposedly serves as a primal platform for students to seek help from peers. In that sense, the forum could be full of random replies to each other, or some students may gradually form study communities and provide mutual support throughout the course period. This kind of feature is important for understanding why students stay in the forum.
- T.2 What are the interactions about?** While sub-forums are classified to accept all kinds of discussions, not all topics are discussed identically. Understanding the focus of students' interactions will help glean their needs regarding the course.
- T.3 What motivates the students?** Students' interaction frequency varies with time, indicating there exists specific reasons why they would or would not dive into the certain forum. The motivations derived from the instruction materials will be specifically helpful for instructors to refine their courses.
- T.4 How do the interaction of different groups vary?** While the former questions analyze students as a whole, the findings may not be representative enough for every individual. We could imagine that, students from different areas may react to the same course materials differently due to cultural diversities. Therefore, pinpointing the students of interest would be necessary for comprehensive assessment.

4 VISUAL DESIGN

We propose to build a visual system, NetworkSeer, to solve the problems mentioned in Section 3. It is built on the forum structure of Coursera, which aligns with our case study. Coursera Forum normally contains multiple sub-forums, which is a primal and default classification of forum topics (e.g., general discussion, assignments). Students could make contributions (threads, posts, or comments) in every sub-forum. Interactions are evoked when students initiate posts (i.e., replies to threads) or comments (replies to posts). For every contribution, we could access its type, original texts, posted date, and its initiator. If it is a post or a comment, we also have the targeted contribution it is replying to. Based on this, our NetworkSeer contains two views, namely Network View, which directly visualizes all the interactions with regard to sub-forums and topics, as well as Feature Statistics View, which serves as an assis-

tive view for perceiving and selecting students' demographic information. The rationales we follow and the concrete designs for the two views and the embedded interactions are introduced below.

4.1 Design Rationale

Based on task analysis, we compile the following design rationales:

- R.1 Assess the features of the interactions:** As a study focusing on social network analysis among MOOCs forums, we need to first identify and present the general characteristics of interactions, which will serve as indicators for further exploration.
- R.2 Align interactions with their triggers:** According to T.2 and T.3, assessment on the biased distributions of the interactions and their triggers will have a big impact on evaluating the course material. Therefore, our visualization should help review when and on what topic the interaction occurs.
- R.3 Allow exploration among different student groups:** We would like to explore the interactions for different student groups, as discussed in T.4. Rather than exploring some pre-defined groups, it is necessary to provide an interface that allows users to arbitrarily select a group based on their features.

4.2 Network View

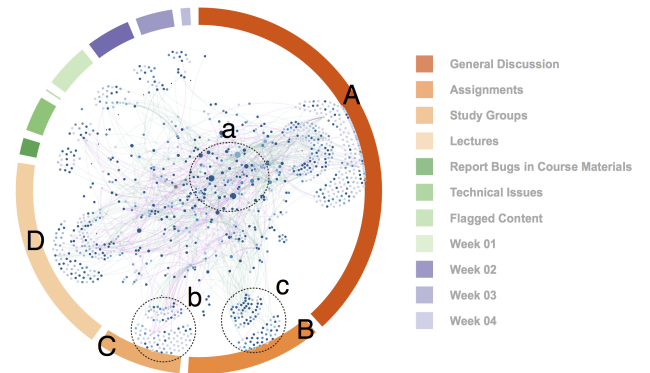


Figure 1: Network View for NetworkSeer, with colored arcs delineating sub-forums (e.g., A-D). Area *a* shows students with high contributions have less specific interests in any sub-forums. *b* and *c* shows students focusing on *Study Groups* and *Assignments* respectively.

Network View is the main view for evaluating the forum interactions between students (R.1). We use the directed graph, a commonly used method for social network analysis (R.1), to represent such interactions (Fig. 2). The detailed design is as follows.

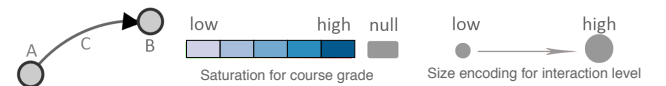


Figure 2: Encodings for student interactions, with node *A* and *B* being students, *C* being (v_A, v_B) , the response from *A* to *B*. Node saturation encodes the grade of the corresponding student, and size encodes the interaction level, $w(v_i, v_j)$.

Denote every student that has contributed at least once in the forum as a node v_i . For v_i and v_j , if v_i replies to v_j , whether in v_j 's thread or post, we add a directed edge (v_i, v_j) between v_i and v_j . The weight of (v_i, v_j) , $w(v_i, v_j)$, is the number of replies from v_i to v_j . This eventually results in a directed graph $G(V, E)$, where V contains all v_i and E all the (v_i, v_j) .

We define a student's level of interaction as his or her corresponding out-degree, i.e., how willing the student is to respond to other students. We encode it with the size of the corresponding node: The larger the out-degree, the bigger the node. Moreover, we encode the course grade with the saturation of the node to investigate the correlation between students' interactions and their grades,

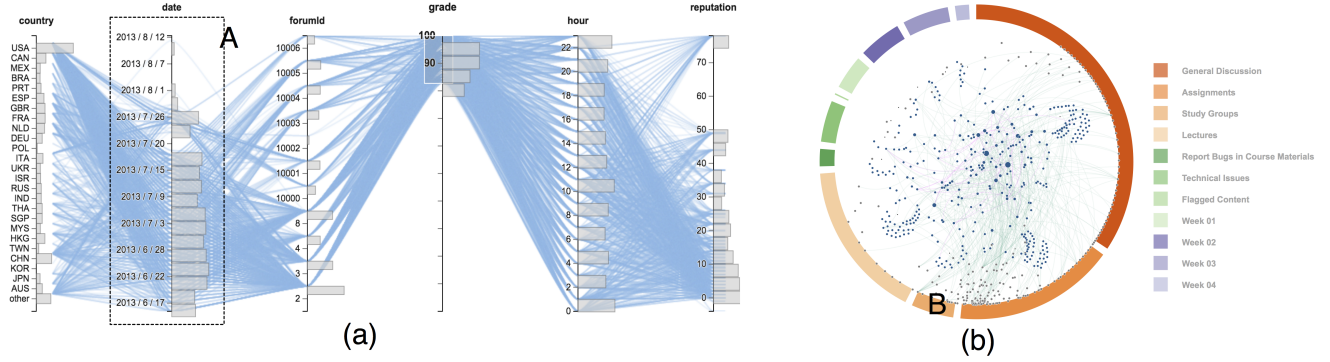


Figure 3: Filtering students with high grades (≥ 85), with (a) being the Feature Statistics View, and (b) the corresponding Network View.

since it strongly indicates how the course requirements have been carried out. Specifically, the higher the grade, the more saturated the node is. We identically segment the grades into five saturation levels to simulate the Likert Scale commonly used in Psychology [23]. As for the edges, we use colors to illustrate if two students interacting are from the same country (purple edge) or not (cyan).

To understand the triggers of the interactions, we incorporate the division of topics into the view. Here, “topic” refers to sub-forums, since they are the primary classification of topics in MOOC forums. Students’ obsession with these forums could effectively reflect their interest. Specifically, we encircle the network graph with a ring graph. Each sub-forum is drawn as an arc of the ring and is colored differently. The sub-forum hotness is another important indicator for the triggers of interactions. Thus, we encode the total number of posts received in the corresponding sub-forum as the center angle of these arcs. For instance, in Fig. 1, A and D represent mainstream sub-forums, while B and C less popular ones.

With the ring graph embedded, instead of following the common practice to automatically layout the social graph, we articulate the node positions with respect to the corresponding student’s obsession on specific sub-forums and topics (R.2). We achieve this with the metaphor of force-balance: For every topic a student participates in, it generates an attractive force \vec{F} on the corresponding node toward the center point of the topic arc, whose quantity $|\vec{F}|$ depends on n , the amount of posts the student contribute to the topic, and r , the distance between the arc and the node. The node will be placed where the force is balanced. Though the force layout could suffer from the overlaid nodes and edges, we still stick with it because its node distributions and edge connections have been proven to be informative for encoding the features of networks [11].

4.3 Feature Statistics View

This view is designed to display all related features of students so that analysts can conduct a visual query to select a specific group of students they are interested in. A parallel coordinate is chosen to show the multidimensional data. (Fig. 3(a)). Several related features that have been proven in previous literature to be related to user behaviors are extracted from the user log data. These include profile related features (i.e., geographical information [2, 3] and grade [15, 3]) and the forum activity feature (i.e., reputation [9], the sum of square roots of votes across all contributions made by a student who participates in the forum). Moreover, because the dates of the forum interactions could reflect how course materials motivate these interactions (R.2), we collect the dates of every post (i.e., post date) as one of the features. We also include its specific time in a day setting, to see if students with similar study habits will share interactions (e.g., logging into the course during the midnight). The sub-forums that contain the contributions are further added to closely link it with Network View.

The brushing interaction on each axis helps to intuitively filter

the user group according to their values on different features (R.3). Both the Network and the Feature Statistics View will be refreshed accordingly, only showing the data for the selected users.

One drawback of the traditional parallel coordinates visualization is that it cannot show the user distribution on different features. Thus, we further add a histogram on each axis, which is crucial to help conduct user selection on this view.

5 CASE STUDY

We verify the usability of NetworkSeer with the data from a course offered by our university on Coursera, which mainly discussed the history of China, and summarize our findings below.

In the overview (Fig. 1), we first notice that students with a significant number of contributions (Fig. 1a) are located in the middle of the graph, with a slight offset to *General Discussion* (Fig. 1A). During browsing, we further find that students represented by these larger nodes tend to have higher reputations. This indicates students being highly acknowledged in the forum are not specifically interested in any of the sub-forum. Also, the filtering result in Fig. 4 shows that they tend to have higher grades.

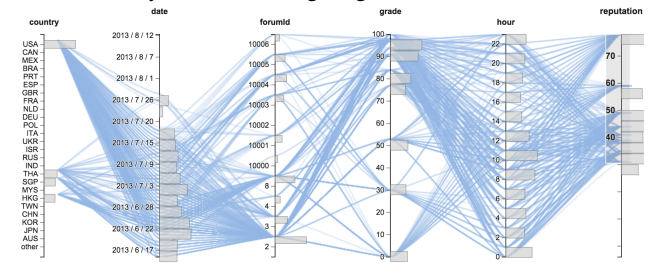


Figure 4: Feature Statistics View for posts from students with high reputation. These students generally have high grades.

Besides them, we also easily identify clusters around sub-forums such as *General Discussion* (Fig. 4A), *Assignments* (B) as well as *Lectures* (D), which indicate that students are more interested in discussing course materials instead of technical problems such as bug reports. The latter not only lack such clusters, but also hold much smaller arc angles. Among them, *General Discussion* is the most popular, which might be because it provides more freedom for unclassified topics to exist. It is also possible that students are sometimes confused about which forum to post in, especially for questions concerning multiple topics. This suggests that *General Discussion* can be further divided or students should be allowed to propose their own desired topics. In addition, since the course materials are the biggest concern of students, instructors and professors can be more involved in these sub-forums to avoid repetitive posts and to improve course outcomes.

The distribution of students in these sub-forum-oriented clusters is interesting. For instance, we notice that nodes close to *Assignments* tend to be more saturated, as in Fig. 1c. According to this,

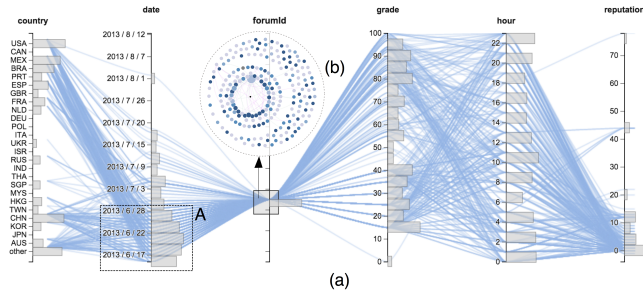


Figure 5: Filtering posts in *Study Groups*, with (a) being the Feature Statistics View, and (b) the corresponding network graph.

we further filter students with high grades. We define *high grades* as those ≥ 85 , which accounts for the top 20% students with effective grades, and confirm our finding on their obsessions: These students, compared with the general group (i.e., all the students participated in the forum), focus more on course related sub-forums, and much less in *Study Groups*, as shown in Fig. 3(b)B. The biased attention makes sense, since students are graded based on assignment performances. We also notice that the heights of *Date* bars in Fig. 3(a) are mostly identical, indicating high-grade-students post in the forum stably, instead of popping up randomly. Also, interestingly, they tend to interact with those in different countries. This might be worth deeper investigation.

In the meantime, those nodes surrounding *Study Groups* in Fig. 1b are colored more lightly, indicating most students, though intended to form study groups, either do not perform well in the course, or have dropped out quickly. To verify if *Study Groups* is an early focus, we query this sub-forum as shown in Fig. 5, and find from the bar chart that most of the contributions are indeed made in the early stage of the course. The corresponding network graph in Fig. 5(b), which contains various purple edges, shows students tend to discuss with people from their own country to form groups here.



Figure 6: Student interactions in different periods: (a) Jun 14 to 28, when *Study Groups* is heavily used, and (b) from Jun 28 till the end.

Querying the dates when *Study Groups* is heavily used and those otherwise respectively, we find that the interactions between different countries are significantly different in these two periods (Fig. 6). We could see from Fig. 6(a) that more purple links appear at the beginning of the course. However, cyan links surpass the purple ones in the latter period (Fig. 6(b)), indicating that though students start their discussion with those from the same country, they tend to interact more with foreign students. This can be explained by their intentions to know more about and to make friends with students who have different backgrounds. This may inform the instructors to strengthen the social impact on student engagements.

The timing distribution of *Assignments* is also interesting. As in Fig. 7, during the several days before the deadlines of Assignment 2, 3 and 4, the number of posts in the sub-forum Assignment is significantly larger than that at other times, especially for Assignment 4. The reason might be that the difficulties students encountered when doing these assignments motivated them to join the discussion in the sub-forum Assignment. The discussion before the deadline of Assignment 1 was not as heated, perhaps because the first assignment was easy enough to be completed well individually.

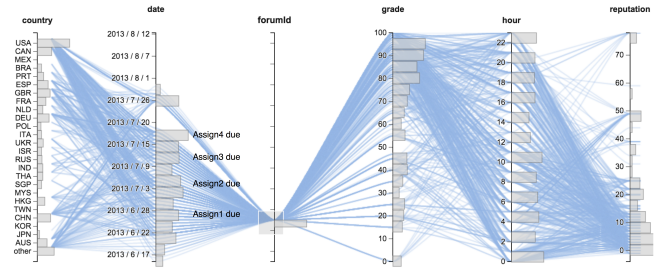


Figure 7: *Assignments* in Feature Statistics View, posts in which experience a periodical variation, and meet its local maxima almost always around the assignment due dates.

By filtering different regions from which students emanate, we find differences among their study focuses: The three sub-forums (except for *General Discussion*) students from North America (Fig. 8(a)) tend to participate in are *Lectures*, *Study Groups* and *Assignments*. For those from East Asia (Fig. 8(b)) are *Lectures*, *Assignments* and *Study Groups*. Europeans (Fig. 8(c)) visit *Assignments*, *Lectures* and *Study Groups* the most. The sequences of the top-three sub-forums show the different emphasis of students from distinct areas. For instance, East Asians attach more importance to forming study groups and Europeans put less attention to lectures than other students. Their daily posting times also vary, which could be due to the time zone differences. Moreover, North Americans discuss the most with students from the same country. This could result from the large number of students from the USA.

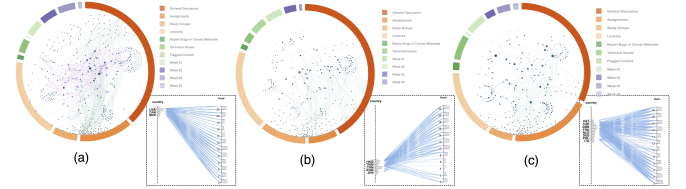


Figure 8: Network View of students from (a) North America (CAN, USA, MEX), (b) East Asia (CHN, HKG, TWN, JPN, KOR), and (c) Europe (PRT, ESP, GBR, FRA, DEU, NLD, POL, ITA)

6 CONCLUSION AND FUTURE WORK

In this paper, we propose NetworkSeer, an interactive visual system to help users, especially instructors, understand the students' interactions in MOOC forums. We characterize common concerns of forum interactions, and design two views for analyzing a forum's overall activeness and the social interactions involved. A case study exemplifies the usefulness and the effectiveness of the design.

NetworkSeer still has room to improve. A primal improvement needed is to append the system with more specific topic analysis. While it could reveal temporal interaction variations, it contains no reference to the original texts. In other words, understanding why they would be specifically obsessed on certain sub-forums, or post content actively on a particular day is difficult. Thus, we will need to add more topic-oriented analysis, and also display the original post content to provide deeper insights into "what the students are talking about". For instance, we may support hierarchical exploration in Network View, and further display students' obsession on more specific sub-topics when selecting one specific sub-forum. We will also support statistical verifications to make it more reliable than pure manual observations.

ACKNOWLEDGEMENTS

The authors wish to thank the MOOC instructors in HKUST for their feedbacks on our related studies, Dr. Conglei Shi for his advice on the visual design, and the anonymous reviewers for their valuable and constructive comments. This research was supported in part by HK RGC GRF 16208514.

REFERENCES

- [1] D. A. G. Aguilar, R. Therón, and F. J. G. Peñalvo. Understanding educational relationships in moodle with vimoodle. In *ICALT*, pages 954–956, 2008.
- [2] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton. Studying learning in the worldwide classroom: Research into edx’s first mooc. *Research & Practice in Assessment*, 8(1):13–25, 2013.
- [3] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu. Peakvizor: Visual analytics of peaks in video clickstreams from massive open online courses. *IEEE Transactions on Visualization and Computer Graphics*, 2015.
- [4] J. Davies and M. Graff. Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, 36(4):657–663, 2005.
- [5] M. Dork, N. H. Riche, G. Ramos, and S. Dumais. Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2709–2718, 2012.
- [6] D. Fisher and P. Dourish. Social and temporal structures in everyday collaboration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 551–558. ACM, 2004.
- [7] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Information Visualization*, pages 32–39. IEEE, 2005.
- [8] N. Henry and J.-D. Fekete. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, 2006.
- [9] J. Huang, A. Dasgupta, A. Ghosh, J. Manning, and M. Sanders. Superposter behavior in mooc forums. In *Proceedings of the first ACM Conference on Learning@Scale*, pages 117–126. ACM, 2014.
- [10] S. Jyothi, C. McAvinia, and J. Keating. A visualisation tool to aid exploration of students’ interactions in asynchronous online communication. *Computers & Education*, 58(1):30–42, 2012.
- [11] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo. Topicpanorama: A full picture of relevant topics. In *Conference on Visual Analytics Science and Technology (VAST)*, pages 183–192. IEEE, 2014.
- [12] S. Mak, R. Williams, and J. Mackness. Blogs and forums as communication and learning tools in a mooc. 2010.
- [13] R. Mazza and V. Dimitrova. Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, 65(2):125–139, 2007.
- [14] M. J. McGuffin. Simple algorithms for network visualization: A tutorial. *Tsinghua Science and Technology*, 17(4):383–398, 2012.
- [15] S. O. Nesterko, S. Dotsenko, Q. Han, D. Seaton, J. Reich, I. Chuang, and A. Ho. Evaluating the geographic data in moocs. In *Neural Information Processing Systems*, 2013.
- [16] M. E. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5200–5205, 2004.
- [17] A. Perer and B. Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 265–274. ACM, 2008.
- [18] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor. Modeling learner engagement in moocs using probabilistic soft logic. In *NIPS Workshop on Data Driven Education*, 2013.
- [19] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, and T. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. In *Computer Graphics Forum*, volume 30, pages 1061–1070. Wiley Online Library, 2011.
- [20] C. Reffay and T. Chanier. How social network analysis can help to measure cohesion in collaborative distance-learning. In *Computer Supported Collaborative Learning*, pages 343–352, Bergen, Norway, June 2003. Kluwer Academic Publishers.
- [21] C. Shi, S. Fu, Q. Chen, and H. Qu. Vismooc: Visualizing video click-stream data from massive open online courses. In *Conference on Visual Analytics Science and Technology (VAST)*, pages 277–278. IEEE, 2014.
- [22] T. Sinha. Supporting mooc instruction with social network analysis. *arXiv preprint arXiv:1401.5175*, 2014.
- [23] J. S. Uebersax. Likert scales: Dispelling the confusion. *Statistical methods for rater agreement*, 31, 2006.
- [24] M. Wen, D. Yang, and C. P. Rosé. Sentiment analysis in mooc discussion forums: What does it tell us. In *Proceedings of Educational Data Mining*, 2014.
- [25] P. Xu, F. Du, N. Cao, C. Shi, H. Zhou, and H. Qu. Visual analysis of set relations in a graph. In *Computer Graphics Forum*, volume 32, pages 61–70. Wiley Online Library, 2013.