Fatima Rabia Yapicioglu, 306627 , rabiayap@uef.fi , Learning Analytics Final Project Report

1. VISUALIZATION AND COMMENT OF INTERACTIONS

1.1. Commentary About the Visualization of The Interactions

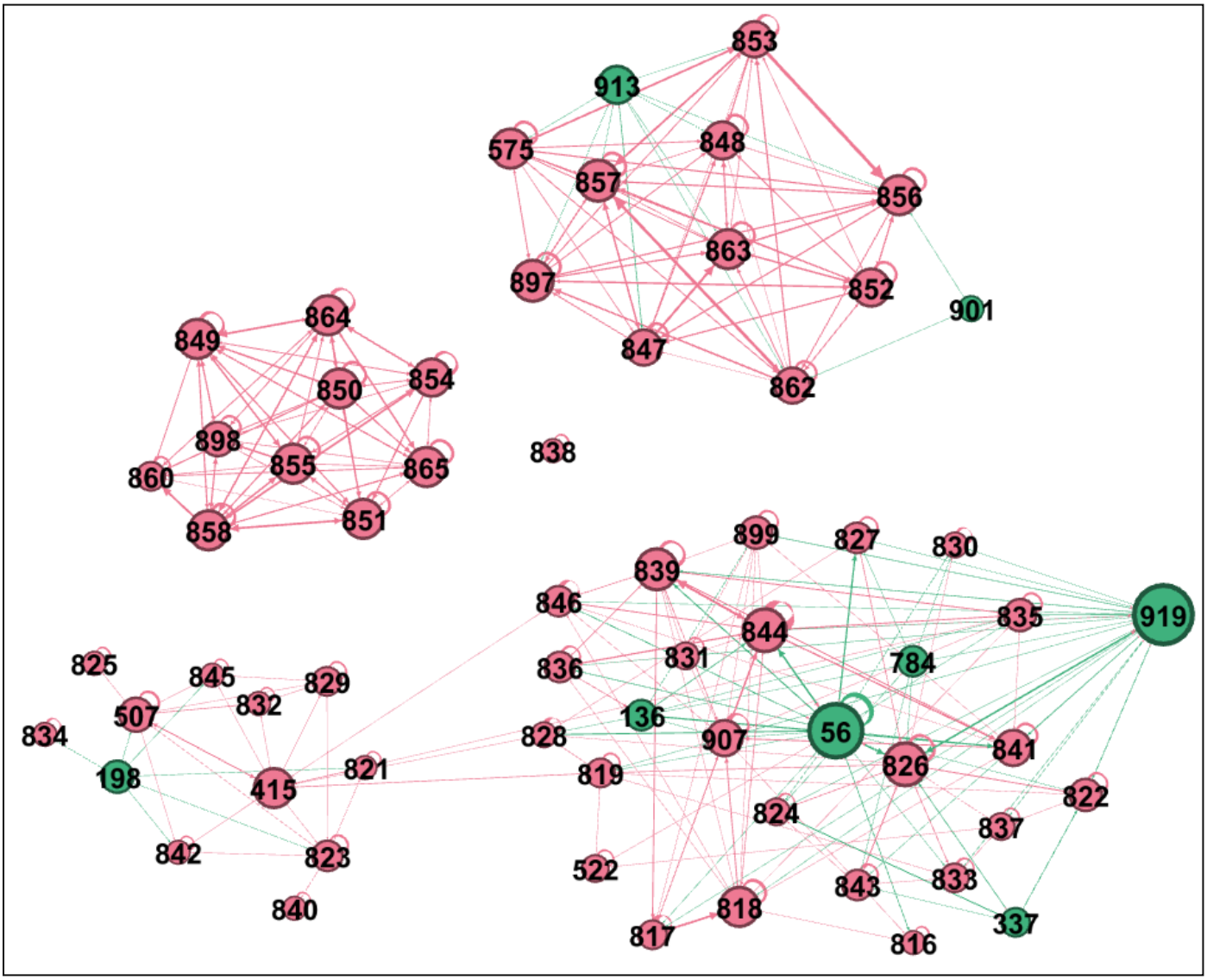


Figure 1.1 Visualization of Interactions Between Students and Teachers, Gephi

Here in the above graph, green nodes represent teachers and the pink nodes represent the students.

In the bottom-left cluster, as it can be seen, the most active student who is trying to interact with other students is 415. It may be showing a scene after assignment grading and feedback. The teacher 198 is giving feedback to the students who completed their homework submissions (834,507,845,415,823 and 842.) This may be an homework that students can work individually(

almst all nodes have cycle to itself) or at most two student, because the probability that the students 825-507, 840-823 and 507-415 are in the same group.

(Thick interaction line between 507-415)Also, just the students 415 and 821 are trying to interact with another cluster, the may have dependent sub-problems in one big project.

In the bottom-right cluster, node 919 has the highest out-degree, we can identify him or her as a coordinator and has too many interaction lines with the rest of the cluster. He may be sending a lot of e-mails about the new updates to both the students and teachers. This may be a scenario about the teacher's and students' interactions during the transition from one period to another period. Some courses are about the end and some courses are about the start, so some teachers like 136,337 and 784 are distributing and receiving less than others. But when we look at the 56, first we can see easily that 56 is spending much more time also working self. (cycle to itself), this is because of he or she is still trying to set up a moodle page or preparing some lecture notes, etc and at the same tie trying to inform the rest of the students about the course(his or her interaction lines are thick, it means that he or she is frequently contacting their students.)Finally, the 839 and 844 has a thick interaction line, maybe they are helping each other about selecting courses.

In the top-left cluster all of the students have cycle to itself and they have very strong interaction lines with each other.This is an isolated group for now, maybe the deadline would be approaching for some assignment, and they are working hard on that both individual and as a group.

In the top-right cluster, teacher 901 is just contacting two students from the class. The reason maybe he wants to especially work with those students in another further Project of him or herself. So we can say that the students 856 and 862 are very successful and got good grades from this course. Also, other students like 853 are asking too many questions to 856 about the course or assignments. (The interaction line is thick, and from 853 to 856).And again the teacher 913 is not receiving just distributing, so we can say that he has given homework previously to students and forgot to share data with them, and now she or he is updating the assignment by sending the data files to the students who are taking that course.

And finally the 838 is a student who is isolated from that environment.But we can say that maybe he or she is not enrolled to his or her own school for this semester, but have gone to perform exchanging studies to another school.Because he or she seems still studying(cycled to itself) but no active interaction line with the other students of this class.

So, these are my oservations about the graph above which shows interactions between students and teachers in one school.We have investigated them cluster by cluster because in every school this is the nature of students and teachers.

For detailed commentary about the interactions between the students and teachers, we can use the demographic data. We have the latitude and longitude of each student and teacher, we can visualize this data by using matplotlib as following, and then we can observe how long-distance interactions affect the student grades or activity.

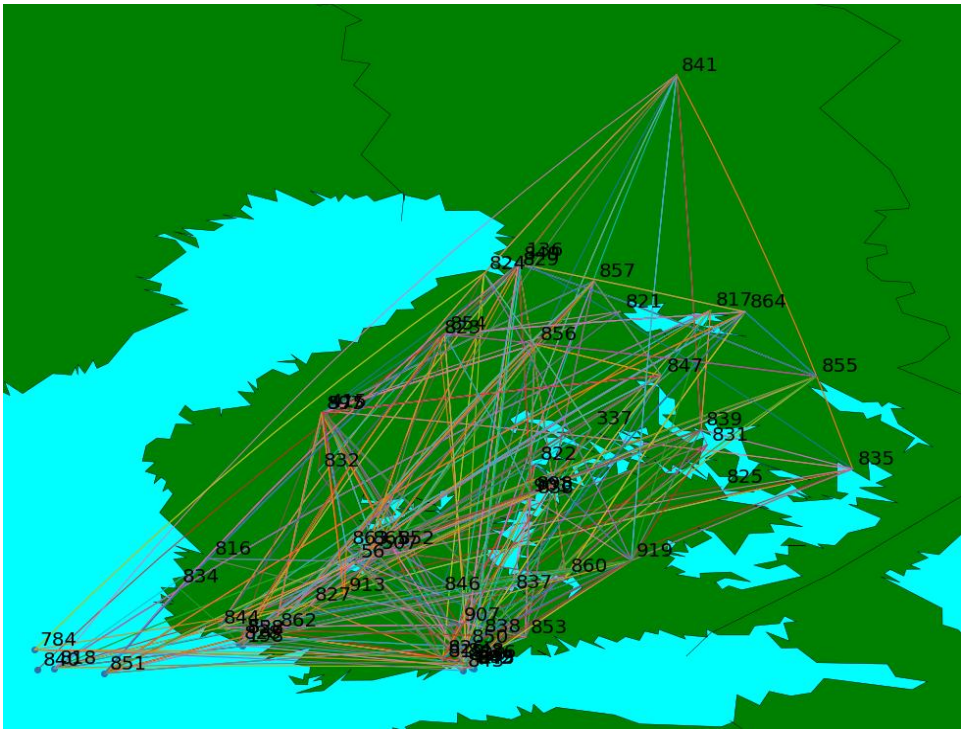


Figure 1.2. Finnish Students and Teachers Interaction Visualization on the Map

But this map shows all the interactions at the same level.So, to investigate the interactions detailed let's use the line weights in the following map.If we can use the Size property in the interactions file we can comment on the interactions easily.

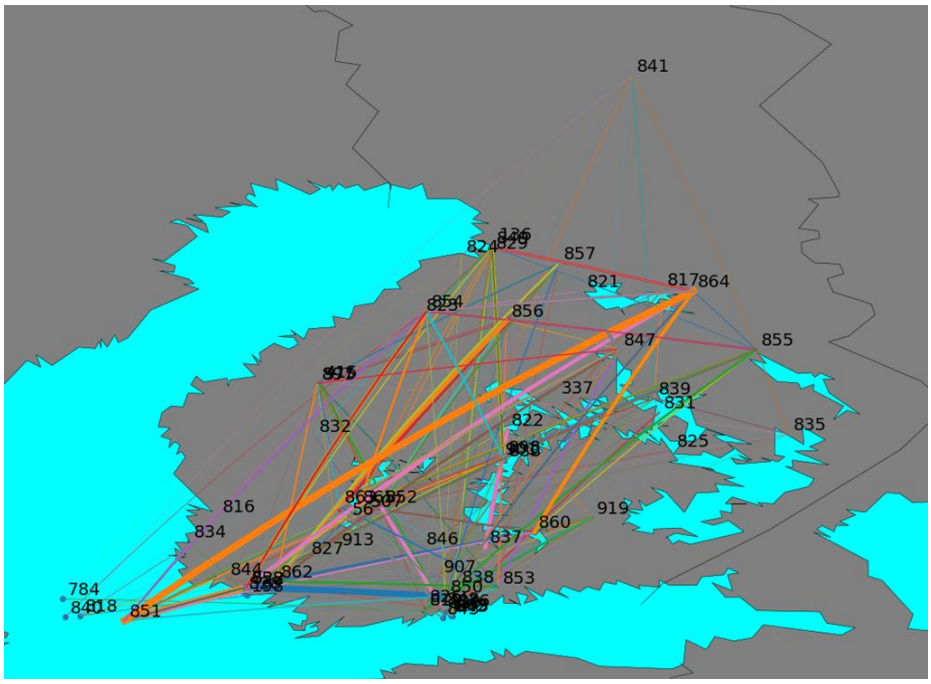


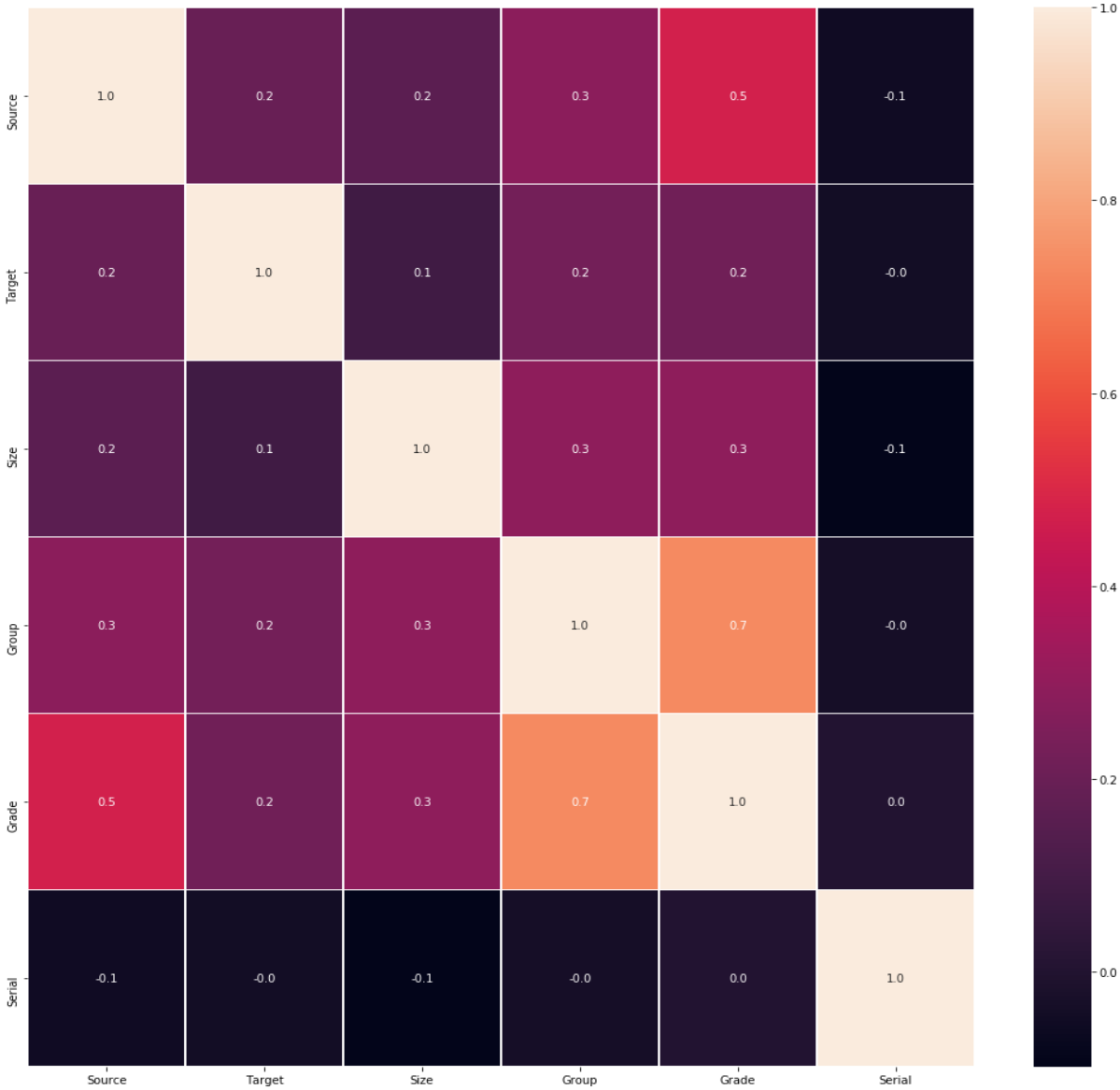
Figure 1.3. Weighted Interaction Lines Between Students and Teachers

When we look at the above graph we can see that even if the students were living far way from each other as 851-864, it doesn't affect the interaction quality and frequency.But for 841 it is not so,841 has weak relationships with the rest of the network.Finally we can say that it depends on the student's or teacher's itself.We have investiated the demographics and interactions together and commented on that, now we will look at the grades and interactions realationship in the folowing section.

2. Investigating the Effects of Interactions on the Student's Grade

2.1. Correlations of Interactions Among Students and Their Effect on the Grade

In any statistical program, we can see the correlation values between attributes of one dataset by using heatmaps, so we created the following heatmap on this purpose.

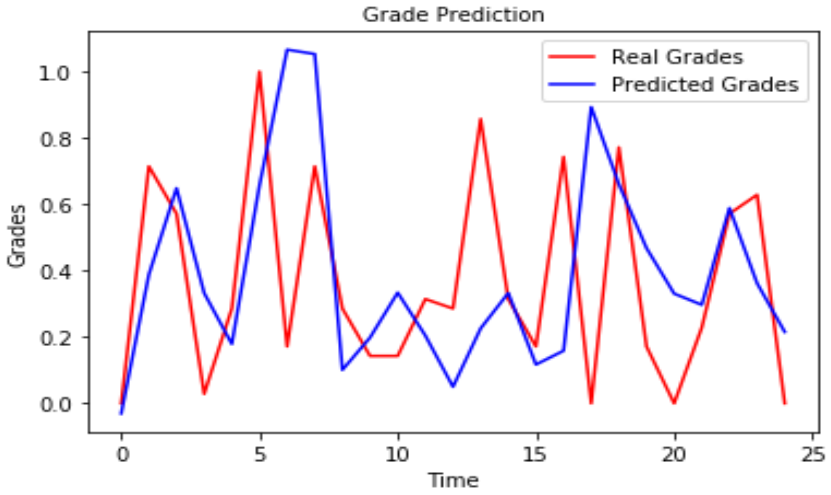


So, now we can see the correlation values on the above graph.For example on the x-axis let's take Grade and compare with the each attribute on the y-axis.So, there there is no relation between the Serial and Grade(0.0), but when it comes to Group and Grade the correlation value is 0.7, it means that there is a high correlation between groups and grades, we can say that group performance has been an effective property on the student's grade.When we take the Size property on the y-axis and compare with the Grade on the x-axis, we can see the value 0.3, that

means that the frequency of interactions is effective on the grade but not that much.Maybe the students arguing the unrelated topics or contacting each other for some other reasons not for their group work.And Finally if we take a look at the Source and Target properties on the y-axis and compare with the Grade of the x-axis, we can say that the success of the student 20% dependent on his/her friend/teacher who is contacted with, and 50% dependent on student his or herself.So, we can understand that the most important properties that affects student's grade are the Group which the student is member of,(it is 70%) and then student's his or her own studies or searches(it is 50%), and finally the Size, we can think of size as the frequency of the interaction between students(it is 30%).

2.2. Predicting the Grades of Students by Using the Merged Data

I avoided to use sensitive data like names, surnames, address, etc.I combined;
calculated interactions centrality file from Gephi,
grades file,
stats file,
detailed interactions file.
Because we can train our model by using those required, useful properties for predicting grades.



So, I have written the code of my model by using Scikit-Learn Library and Linear Regression for prediction. Divided the data into two-part train/test and after training my data I tested by using test data, I got the above visualization. So still there must be some useless properties on predicting grades because we couldn't get that much better results.Or we can train our model with much more epoches to get better results.
We can use interactions data in a much better way, for example, we can find the total number of interaction size for each student and then use it as a property.If we use the total number of interactions or the total size of the interactions we would get a better property for predicting grades and train our model.So, my observation is that we need to use those interactions by producing two new property called,

- Total number of interactions
- Total size of all interactions

for each student before training our model.

3. References

3.1. Matplotlib Basemap Toolkit documentation, https://matplotlib.org/basemap/
3.2. Basemap Utility functions, https://basemaptutorial.readthedocs.io/en/latest/utilities.html