Memorandum

To: Prof. Peng Shi

From: Alex Chang, Modan (Eileen) Wang, Gyan Prakash, Irem Kahyaoglu, Ruhan Dong

Date: March 6, 2018

Subject: Interim Progress Report for USC Marshall Class Scheduling Project

Executive Summary

The objective of this project is to apply optimization techniques to improve the current class scheduling system at USC Marshall School of Business. Based on our analysis of the data and interviews with students, we concluded that the biggest opportunity for improvement lies in reducing the number of times a student has to go to campus in a given week.

Currently, departments are assigned time slots en masse such that they can schedule department classes based on their preferences and/or feedback from professors. In an ideal scenario, when a particular student is required to take courses only from the department that runs his/her program, this system can be quite efficient. However, programs run by USC Marshall are becoming increasingly interdisciplinary and cross-department core/elective courses are very common.

Therefore, we decided to dig into '*Student Course Selection'* dataset. This dataset had over 100,000 entries corresponding to six terms from Fall 2015 through Summer 2017. It had enrollment information about 449 unique courses run by USC Marshall over this period. To begin with, we looked at pairs of courses in which the highest number of students enrolled in a given term.

Item Set	Support
{BUAD-425, BUAD-497}	3.75%
{BUAD-310, ECON-351}	3.12%
{BUAD-304, ECON-351}	2.36%
{GSBA-510, GSBA-542}	1.91%
{BUAD-310, ECON-352}	1.81%
{BUAD-304, ECON-352}	1.76%
{BUAD-307, ECON-352}	1.74%
{BUAD-497, WRIT-340}	1.59%
{BUAD-304, BUAD-310}	1.50%
{BUAD-425, WRIT-340}	1.44%

Association analysis on the course selection data gave us paired itemsets with highest *supports* i.e., the percentage of enrollments with contain a given pair of courses. The top 10 such pairs are given alongside. It can be seen that 7 out of 10 most popular pairs of courses (shaded in grey color) are offered by different departments. This makes it all the more important that departments coordinate with each other to make the schedules more streamlined such that students don't have to make multiple trips to campus over the week in order to attend such classes.

We further analyzed all pairs of courses which had (association) support of at least 1%. There are 73 such pairs. We found that three-fourths of such pairs have at least one class on the same day of the week. However, less than one-fourth of such pairs had all the classes on the same days in a given term. This leaves us with a lot of scope for improvement. It is our objective to schedule classes in such a way that pairs of courses with highest supports tend to have all the classes on the same days of the week. We plan to prioritize scheduling based on these supports.

It must also be noted that we understand that this data is biased as in students were able to enroll in only those courses which didn't have time conflicts. It is possible that course enrollments were swayed by factors such as whether certain courses were scheduled on common days of the week. For example, if a student is interested in two different courses but those courses require her to make four trips to the campus per week (two for each course), she might choose a pair of courses that minimize her commuting time. It will, therefore, be very helpful to have student preference data instead of student selection data as the latter has been biased by historical scheduling restrictions. Such data would give us an accurate picture of course baskets whose schedules need to coordinated such that number of trips to campus and possibly, gap between classes are optimized.

Measure of Goodness

As discussed earlier in this text, we used '*Student Course Selection*' dataset to study which courses should be kept clear of any possible time conflicts, both in terms of days of the week the courses are scheduled on as well as the exact time slots. Once, we found out most popular pairs of courses that are taken together, we compared it against the '*Course Enrollment*' data for the same period. Thus, our metric measured the consonance between pairs of courses that tend to be taken together and whether they are scheduled on the same day. We found that three-fourths of such pairs have at least one class on the same day of the week but less than one-fourth of such pairs had all the classes on the same days in a given term.

Our metric conforms with CASE approach and a detailed explanation has been given below.

- **<u>Computable</u>**

  Historical student course selection data and student course preference surveys can be used to generate a ranked list of paired courses with high preferences. Such paired courses should be carefully scheduled to avoid time and/or even spatial conflicts, wherever possible. In case, an expansive survey is unavailable, (association) support can be used to assign weights to pairs of courses and prioritize them during scheduling.

- **<u>Actionable</u>**

  The current scheduling approach is largely manual and is, therefore, very limited its capability to cope with such a multiplicity of factors. However, with the optimization approach discussed in this text, scheduling can be made very easy and accommodative of the interests of different parties. A very small fraction of courses would need to be manually scheduled and only under special circumstances.

- **<u>Simple</u>**

  By prioritizing students as the primary stakeholders, we have found the most direct way of incorporating historical inputs from students to improve scheduling and the process involved. Based on anecdotal evidence, it can be said that most students pick courses based on available schedules, however if we allow them to learn whatever they would rather be learning, this will not only improve students’ satisfaction but also, improve their career prospects (since, they would be able align courses and career interests).

- **<u>Enlightening</u>**

  Although the preferences derived from the historical student course selection data may not prove to be transformational for students right away (given that the data is biased), it will be over many iterations of this new scheduling approach that the real benefits will begin to manifest.

Appendix

1. **Pairs of courses with greater than 1% (association) support**

Item Set	Support
{BUAD-425, BUAD-497}	3.75%
{BUAD-310, ECON-351}	3.12%
{BUAD-304, ECON-351}	2.36%
{GSBA-510, GSBA-542}	1.91%
{BUAD-310, ECON-352}	1.81%
{BUAD-304, ECON-352}	1.76%
{BUAD-307, ECON-352}	1.74%
{BUAD-497, WRIT-340}	1.59%
{BUAD-304, BUAD-310}	1.50%
{BUAD-425, WRIT-340}	1.44%
{BUAD-302, BUAD-306}	1.40%
{GSBA-510, GSBA-511}	1.38%
{BUAD-306, BUAD-311}	1.34%
{BUAD-302, BUAD-307}	1.33%
{BUAD-306, BUAD-307}	1.31%
{BUAD-280, ECON-352}	1.23%
{BUAD-307, ECON-351}	1.21%
{BUAD-304, BUAD-307}	1.18%
{ACCT-370, BUAD-302T}	1.17%
{ACCT-370, ACCT-371}	1.17%
{BUAD-302, ECON-352}	1.16%
{GSBA-540, GSBA-542}	1.15%
{GSBA-516, GSBA-540}	1.15%
{GSBA-516, GSBA-542}	1.15%
{BUAD-307, BUAD-310}	1.14%
{GSBA-510, GSBA-516}	1.14%
{GSBA-510, GSBA-540}	1.14%
{GSBA-504A, GSBA-516}	1.14%
{GSBA-504A, GSBA-540}	1.14%
{GSBA-504A, GSBA-542}	1.14%
{GSBA-580A, GSBA-580B}	1.14%
{GSBA-504A, GSBA-521A}	1.13%
{GSBA-504A, GSBA-533}	1.13%
{GSBA-504A, GSBA-552}	1.13%
{GSBA-516, GSBA-521A}	1.13%
{GSBA-516, GSBA-533}	1.13%

Item Set	Support
{GSBA-516, GSBA-552}	1.13%
{GSBA-521A, GSBA-540}	1.13%
{GSBA-521A, GSBA-542}	1.13%
{GSBA-533, GSBA-540}	1.13%
{GSBA-533, GSBA-542}	1.13%
{GSBA-533, GSBA-552}	1.13%
{GSBA-540, GSBA-552}	1.13%
{GSBA-542, GSBA-552}	1.13%
{GSBA-521A, GSBA-533}	1.13%
{GSBA-521A, GSBA-552}	1.13%
{GSBA-580A, GSBA-580C}	1.13%
{GSBA-580B, GSBA-580C}	1.13%
{GSBA-504A, GSBA-510}	1.13%
{GSBA-504A, GSBA-545}	1.13%
{GSBA-510, GSBA-521A}	1.13%
{GSBA-516, GSBA-545}	1.13%
{GSBA-540, GSBA-545}	1.13%
{GSBA-542, GSBA-545}	1.13%
{GSBA-510, GSBA-533}	1.12%
{GSBA-510, GSBA-552}	1.12%
{GSBA-521A, GSBA-545}	1.12%
{GSBA-533, GSBA-545}	1.12%
{GSBA-545, GSBA-552}	1.12%
{BUAD-285B, BUAD-306}	1.12%
{GSBA-510, GSBA-545}	1.12%
{GSBA-521A, GSBA-521B}	1.11%
{GSBA-504A, GSBA-521B}	1.10%
{GSBA-516, GSBA-521B}	1.10%
{GSBA-521B, GSBA-533}	1.10%
{GSBA-521B, GSBA-540}	1.10%
{GSBA-521B, GSBA-542}	1.10%
{GSBA-521B, GSBA-552}	1.10%
{GSBA-510, GSBA-521B}	1.09%
{GSBA-521B, GSBA-545}	1.09%
{ACCT-371, BUAD-302T}	1.08%
{BUAD-302, BUAD-311}	1.08%
{GSBA-528, GSBA-548}	1.04%

2. Python Code

a) Importing libraries and reading files

```
import numpy as np
import pandas as pd
import os, re
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats

os.chdir("C:\\Users\Gyan Prakash\OneDrive - USC Marshall School of
Business\Spring 2018\DSO 570\Final Project")

ce = pd.read_excel("Marshall_Course_Enrollment_1516_1617.xlsx")
cc = pd.read_excel("Cancelled_Courses_1516_1617.xlsx")
rc = pd.read_excel("Marshall_Room_Capacity_Chart.xlsx")
ss = pd.read_excel("Summary_Special_Session_Codes_1516_1617.xlsx")
da = pd.read_excel("Department_Allocations_20171.xlsx")
cs = pd.read_excel("Student_Course_Selection_1516.xlsx")
```

b) Creation of unique identifiers

```
cs['UID'] = cs['Randomized Unique
Identifier'].astype(str)+'_'+cs['Term'].astype(str)
ce['UID'] = ce['Term'].astype(str)+'_'+ce['First Days'].astype(str)

ce.to_csv("CE With UID.csv")
cs.to_csv("CS with UID.csv")
```

c) Horizontal stacking of array of courses for association analysis

```
cs.groupby('UID')['Course'].apply(list).to_csv('Course
Association.csv')
ca = pd.read_csv('Course Association.csv', names = ['UID', 'Course'])
ca_array = np.hstack(ca['Course'])
```

d) Sample Python code for association analysis

```
for line in ca['Course']:
  for item in line:
    if not item in freqMap:
      freqMap[item] = {}

    for other_item in line:
      if not other_item in freqMap:
        freqMap[other_item] = {}

      freqMap[item][other_item] = freqMap[item].get(other_item, 0) + 1
      freqMap[other_item][item] = freqMap[other_item].get(item, 0) + 1

ca_array_df = pd.DataFrame(freqMap).T.fillna(0)
```

Disclosure: We aided our association analysis using JMP.