Vietnamese Student Association ACE (Big/Little) Pairing Statistical Analysis

Dan Tran

PURPOSE

Many are probably thinking to themselves, “Why would anyone make a report on an event that happened 3 months ago?” Others may even ask, “Who in the right mind would make a report for fun?” To that question, I answer–me. Honestly, I needed to add C# and dotnet to my coding language bag, and this seemed like the perfect opportunity to do it. I would like to preface a few things at the beginning of the report. First, my only knowledge of probability and statistics is based on a single class I took in the second semester of my first year (APMA 3100), so the accuracy of the stats may have some flaws. Second, there are many pros for me using C#. The main reason was so I could learn the language, but another benefit is that they do not teach C# at UVA. This is a benefit because if I make a coding or algorithmic mistake, only a handful of you can criticize me for it 🙂. I will still add a GitHub link to the code that I used to parse the data at the end of the report (without the rankings files because they’re confidential!). Lastly, do not expect my grammar and English to be perfect–I hate writing. This is an informal report, so it is written in conversational/formal writing.

INTRODUCTION

Given 4 days to match a group of 94 littles with 60 bigs, many combinations and permutations (yes order does matter!) would have led to different big/little pairings. However, Lance and I created pairings that led to this specific event. As full-time students with midterms just around the corner, we did our best to create pairings that not only satisfied us but as many bigs and littles as possible–given the time constraint. Although how well we did on these pairings is subjective, it would be nice to truly know how well we did. Good thing there is! This report will perform raw calculations on both the bigs' and littles’ preference forms to see how accurate Lance and I were.

SHARED BIG AND LITTLE PAIRING DATA

First, we can dive into the logistics of how the pairings were made. Those who are ranked higher on a preference sheet are weighed more than those ranked lower, however when going through the preference sheet, a person's entire preference list is looked at. Next, Lance and I decided to run on a "little's preference" methodology. What does this mean? Let's take a look at a scenario that occurred (using filler names for confidentiality!). In Fig 1 at first glance, it would be easy to conclude that Bob should get James and John should get Robert. However, since the pairings were made on a little's preference" methodology, Bob ended up getting Robert and John ended up getting James. Although the bigs ranked the opposing little higher, they were paired with a little that they ranked lower. This is because we weighed who the littles wanted more.

Bob (Big 1)	John (Big 2)	James (Little 1)	Robert (Little 2)
James	Robert		Bob
Robert		John	John
	James		
		Bob	

Fig 1. Example Preference Sheet

Now that our approach is clear, let us dive into the data itself. Of all the pairings that were made, 66.23% of pairings had a big and little list one another on their preference sheet–we will call this a "match." This 66.23% is slightly inaccurate because some bigs and littles did not have a preference for who they were paired with, so I listed empty slots in their sheet as "n/a." In the code, "n/a" is considered a NONMATCH meaning that the amount of pairings where a big or little received someone on their preference sheet is actually greater than or equal to 66.23%.

This would mean that 33.77% of pairings did not have bigs and littles that matched with one another. From here, our approach shifted. We began to look at littles' and bigs' preference sheets to see the type of people that they ranked and matched them from there. Of the 33.77% of pairings, this would mean either: a big received a little they listed without the little putting them down, a little received a big they listed without the big putting them down, or a pairing occurred where neither the big nor the little put one another down. For now, we will dive into the latter statement since the first two statements will have their own section in this report.

Of all the pairings that were made where neither the big nor the little put one another on their respective preference sheets, 48.89% of them had a similar interest. In all honesty, I did NOT have time to sieve through all the interests and create keywords when parsing the data, so I asked ChatGPT 4's Data Analysis feature. An example of interest transformation is shown in Fig 2. Some problems with this method of data transformation is that it may lead to the statistic obtained being inaccurate since it may have not listed certain traits, or even appended traits that were not listed.

Name	What Was Submitted	→	What Was Turned Into
Person	watch tv/movies, play tennis, pickleball, spikeball, dance, video games, goof around at vsa or other org events, cooking sessions, study hangouts get swole, send infinite reels/tiktoks, hangout with the lineage, random debriefs, help them with their hw	→	shows, movies, cooking, dance, games, lineage, pickleball, spikeball, tennis, gym, tiktok, reels, study, hangout, hw

Fig 2. Example Interest Transformation for Data Parsing

So now you are likely thinking, what happened to the other 51.11% of pairings that occurred where neither the big nor little did not list one another? Lance and I would then look at the people listed and see if there was someone similar to the people listed. It helped a lot

for bigs and littles who came out to fall family week events and went into depth with their interests!

BIGS PAIRING DATA

In this section, we will cover the specific case where a big received a little that they placed on their preference sheet–meaning that it does not matter what the little had on their preference sheet. To jump right into the data, 63.33% of bigs received someone they listed in their top 3, and 76.67% of bigs received someone they listed in their top 6! Many bigs were highly contested, the percentage of times the 5 most wanted bigs were listed on the overall little's preference sheet is shown in Fig 3. The 5 most wanted guy bigs are shown in Fig 4 and the 5 most wanted girl bigs are shown in Fig 5. It is important to note that these percentages are not additive. A single little can rank multiple bigs, so these percentages are likely to overlap. The percentages of how frequently every big was listed on the little's preference sheet are in ascending order below in Fig 6.

Big's Name	Percentage That They Appeared
Big 1	22.34%
Big 2	19.15%
Big 3	18.09%
Big 4	17.02%
Big 5	14.89%

Fig 3. Percentage of 5 Most Wanted Bigs Listed on Overall Little's Preference Sheet

Big's Name	Percentage That They Appeared
Guy Big 1	19.15%
Guy Big 2	17.02%
Guy Big 3	14.89%
Guy Big 4	13.83%
Guy Big 5	13.83%

Fig 4. Percentage of 5 Most Wanted Guy Bigs Listed on Overall Little's Preference Sheet

Big's Name	Percentage That They Appeared
Girl Big 1	22.34%
Girl Big 2	19.15%
Girl Big 3	18.09%
Girl Big 4	14.89%
Girl Big 5	13.83%

Fig 5. Percentage of 5 Most Wanted Girl Bigs Listed on Overall Little's Preference Sheet

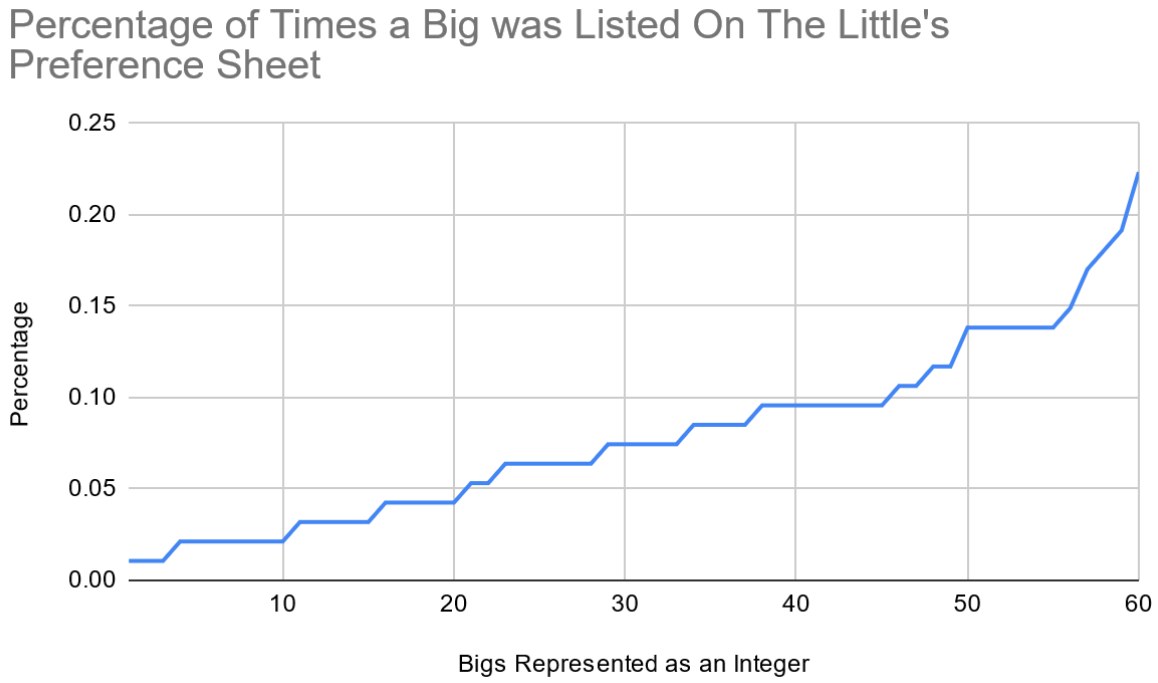

Fig 6. Frequency Percentage of Every Big Listed on Little's Preference Sheet

LITTLES PAIRING DATA

Similar to the previous section, we will cover the specific case where a little received a big that they placed on their preference sheet–meaning that it does not matter what the big had on their preference sheet. The percentage of littles that received someone in their top 3 was 65.96%, and the percentage of littles that received someone in their top 6 was 79.79%. When comparing the top 3 and top 6 percentages of the littles to the bigs–the littles'

percentages are slightly higher. This statistic agrees with me and Lance's original methodology of rolling on a "little's preference" basis. Similar to bigs, many littles were highly contested, the percentage of times the 5 most wanted littles were listed on the overall bigs' preference sheet is shown in Fig 7. The 5 most wanted guy bigs are shown in Fig 8 and the 5 most wanted girl bigs are shown in Fig 9. Of the top 15 most wanted littles, 10/15 of them were girls, indicating a heavy preference for girl littles. Since the ratio of guy to girl littles was close to 50/50, we can conclude that many people picked from the same pool for girl littles. The percentages of how frequently every little was listed on the big's preference sheet are in ascending order below in Fig 10.

Little's Name	Percentage That They Appeared
Little 1	23.33%
Little 2	16.67%
Little 3	15%
Little 4	15%
Little 5	15%

Fig 7. Percentage of 5 Most Wanted Littles Listed on Overall Big's Preference Sheet

Little's Name	Percentage That They Appeared
Guy Little 1	15%
Guy Little 2	15%
Guy Little 3	13.33%
Guy Little 4	13.33%
Guy Little 5	13.33%

Fig 8. Percentage of 5 Most Wanted Guy Littes Listed on Overall Big's Preference Sheet

Little's Name	Percentage That They Appeared
Girl Little 1	23.33%
Girl Little 2	16.67%

Girl Little 3	15%
Girl Little 4	15%
Girl Little 5	15%

Fig 9. Percentage of 5 Most Wanted Girl Littles Listed on Overall Big's Preference Sheet

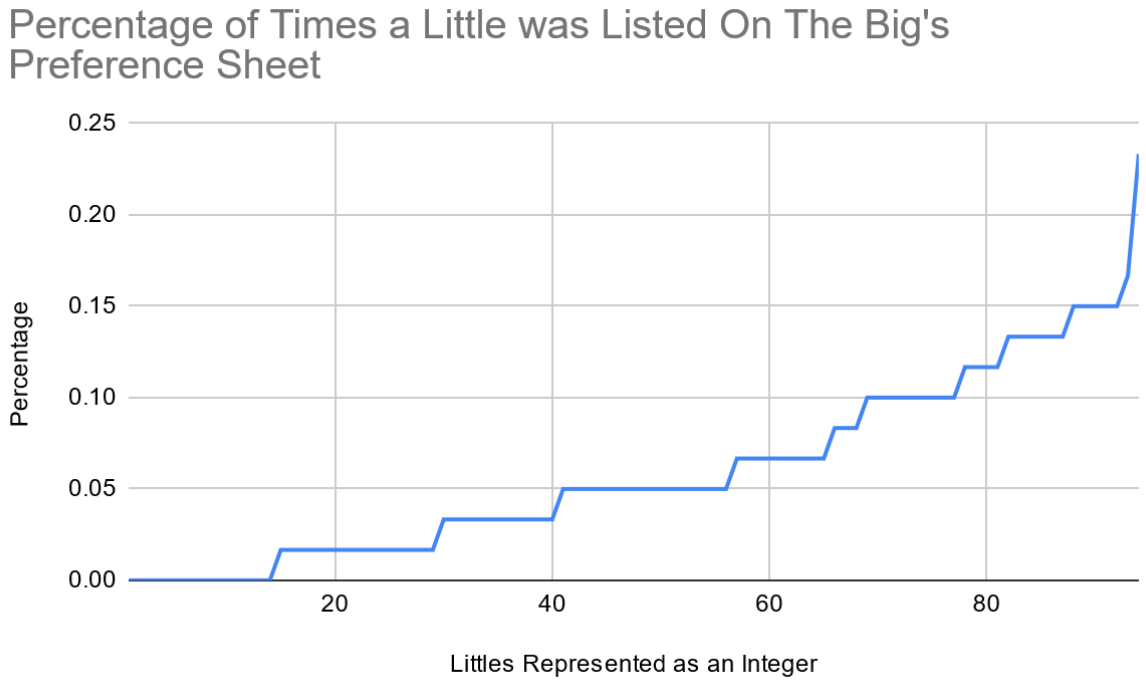

Fig 10. Frequency Percentage of Every Big Listed on Little's Preference Sheet

CONCLUSION

In this statistical analysis of the Vietnamese Student Association's ACE (big/little) pairing event, we have observed significant insights into the preferences and outcomes of the pairing process. Our methodology, focusing primarily on the littles' preferences, resulted in a high percentage of mutual selections and satisfied pairings. It's essential to acknowledge the limitations of the analysis, primarily the reliance on a single statistical approach and potential inaccuracies in interest matching. However, the overall success rate of the pairings, as indicated by the high percentage of top-ranked matches, suggests that our methodology was effective in meeting everyone's expectations. This report not only sheds light on the

effectiveness of Lance and I's matching algorithm but also provides a foundation for future improvements in the Big/Little pairing system.

CODE

https://github.com/auroradan/VSA-Big-Little-Report