Analyzing Innovation Dynamics Within K-pop Music

Andy Ouyang

RESEARCH QUESTION

In the vast landscape of music, artists navigate a dynamic journey, marked by artistic evolution, societal influence, and the pursuit of creative expression. Their discography becomes a reflection of their growth, experimentation, and impact on the musical landscape.

K-pop groups start with a debut album and continue to release songs where a musical concept and identity are established. However, as groups traverse the oversaturation of the K-Pop music market, staying stagnant is never always the answer as times will change and people gravitate towards newer and different kinds of music. As such small amounts of experimentation along the way will occur and at some point, their musical identity changes.

This project and experiment aim to explore the evolution of artists' discography throughout their music careers through the lens of innovation dynamics and the patterns of fame framework. **The question that is worth investigating is how a group's sound evolves and whether that innovation brings new success to the group. In the end, are people interested in innovation or similarity in music in regards to a particular group that is meant to have stability but also adapt to the times?**

The working hypothesis will be that as most groups progress in their musical career, each release gets more and more similar to the previous one. There could be a point where they could reach a relatively stable point of similarity or branch out and experiment with their musical identity. Because groups often occupy a musical niche within the vast market, it can be said that people are drawn more to similarity but can react very differently to innovation.

I will be using the K-pop group GFRIEND, who released 14 Korean albums within their 6 active years from 2015 to 2021. The group debuted with an innocent, love-centered theme that progressed into a more mysterious, melancholic sound (around the 6th mini album) that finished with mature and darker themes with the final trilogy series, leading to the hypothesis of two possible divergences in similarity in the later visualizations.

CONNECTION TO CLASS IDEAS

Poetry is always changing its trajectory in terms of what is being innovative and music is similar in that regard. The content of poetry reflects the times and events of the era, but music does not reflect that within smaller frames of time. However, there are other ways to interpret innovation dynamics for music, like from a musical perspective or using context from understanding the general opinion in online discourse communities like Reddit.

This project directly follows the framework of similarity for different features of the medium to determine the innovation dynamics and create explanations for people's curiosities, but within the realm of K-pop music and similarity trends for a singular artist instead of poetry. From the lecture on novelty and transcience, we noted that sameness attracts but novelty disturbs except for certain outliers that flourish. Because music groups have established musical concepts and identities, it would be fascinating to see what happens when they challenge what is expected of them and come out with an unexpected release.

Within the music industry, information cascades also exist where producers and composers all have access to many different sources of information that can provide potential success. Different companies manage different groups and many private and public signals can be interpreted and inferred from to make the next best decision for themselves. As an artist's career progresses, they may be influenced by these information cascades and signals, leading to changes in their musical style and experimentation with new sounds and concepts. So, the idea of information cascades aligns with how artists and producers in the music industry rely on trends and market signals to guide their creative decisions. By analyzing patterns of fame and innovation dynamics, this project aims to uncover how artists navigate these information cascades and adapt their musical identity to stay relevant in an ever-changing industry landscape.

DATA GATHERING

Inspired by Kaggle datasets of Spotify data that already exist on the internet, I tried to make my version of that with all of the songs from the artist GFRIEND.

The Spotify API allowed me to search for all of the albums and gather all of their tracks as well as all of the additional variables. Using that, I created a dataframe that contains all of the tracks for the 14 Korean albums GFRIEND released. I deleted all Intro tracks and instrumental tracks since they are all nonlyrical, cutting down on the total number of tracks in the dataframe.

	Album	Artist	Track Name	Track ID	Popularity	danceability	energy	key	loudness	mode	...	valence	tempo	type	id	uri
1	GFRIEND 1st Mini Album 'Season Of Glass'	GFRIEND	Glass Bead	2rFRa1eJ8RqlINbJCZwdtt	55	0.695	0.870	6	-2.660	1	...	0.697	114.022	audio_features	2rFRa1eJ8RqlINbJCZwdtt	spotify:track:2rFRa1eJ8RqlINbJCZwdtt
2	GFRIEND 1st Mini Album 'Season Of Glass'	GFRIEND	Neverland	7GFlY3Aigy7AbhnwX6Mk9d	33	0.741	0.896	10	-3.406	1	...	0.759	126.038	audio_features	7GFlY3Aigy7AbhnwX6Mk9d	spotify:track:7GFlY3Aigy7AbhnwX6Mk9d
3	GFRIEND 1st Mini Album 'Season Of Glass'	GFRIEND	White	7BuGnjX5DpKuJ0OdRUpDnV	39	0.666	0.811	0	-3.448	1	...	0.399	97.007	audio_features	7BuGnjX5DpKuJ0OdRUpDnV	spotify:track:7BuGnjX5DpKuJ0OdRUpDnV
6	GFRIEND 2nd Mini Album 'Flower Bud'	GFRIEND	Me Gustas Tu	6x7fux7bZEfnChKx3nhSZn	63	0.646	0.923	8	-1.309	1	...	0.703	106.086	audio_features	6x7fux7bZEfnChKx3nhSZn	spotify:track:6x7fux7bZEfnChKx3nhSZn
7	GFRIEND 2nd Mini Album 'Flower Bud'	GFRIEND	Under The Sky	3RDnqiCqgA0Vg6z2NAxQqY	33	0.644	0.906	2	-2.834	1	...	0.793	120.002	audio_features	3RDnqiCqgA0Vg6z2NAxQqY	spotify:track:3RDnqiCqgA0Vg6z2NAxQqY
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
106	回:Walpurgis Night	GFRIEND	Night Drive	7azmn7M9mRnHzyqxzjMsNe	43	0.613	0.649	5	-3.899	0	...	0.394	88.956	audio_features	7azmn7M9mRnHzyqxzjMsNe	spotify:track:7azmn7M9mRnHzyqxzjMsNe
107	回:Walpurgis Night	GFRIEND	Apple	1j6VvjYG0YrkMTcyCBHj8F	48	0.711	0.939	9	-4.725	1	...	0.331	119.996	audio_features	1j6VvjYG0YrkMTcyCBHj8F	spotify:track:1j6VvjYG0YrkMTcyCBHj8F
108	回:Walpurgis Night	GFRIEND	Crossroads	0LZK17dJosyY7zftjUY4mK	41	0.376	0.834	1	-1.882	1	...	0.484	169.900	audio_features	0LZK17dJosyY7zftjUY4mK	spotify:track:0LZK17dJosyY7zftjUY4mK
109	回:Walpurgis Night	GFRIEND	Labyrinth	0Z5lHhsYyaAFJ8JeULI4PW	38	0.623	0.898	2	-2.441	0	...	0.858	124.961	audio_features	0Z5lHhsYyaAFJ8JeULI4PW	spotify:track:0Z5lHhsYyaAFJ8JeULI4PW
110	回:Walpurgis Night	GFRIEND	Wheel of the year	2cSPQRjVCkmPDiVfKSOgbj	37	0.598	0.932	6	-2.470	1	...	0.360	131.987	audio_features	2cSPQRjVCkmPDiVfKSOgbj	spotify:track:2cSPQRjVCkmPDiVfKSOgbj

Figure I. A view of the dataframe after dropping all non-vocal tracks

One variable of interest here is the popularity variable, a value ranging from 0 to 100 being the most popular and is calculated by an algorithm that factors in the total number of plays and how recent those plays are. This variable will give us valuable information later on about how much similarity and patterns of fame matter in the larger scheme of things.

For each track, there are audio-related variables such as the danceability of a track or the energy of the track. For each track, I also included lyrical analysis in which I gave ChatGPT a prompt for the Korean and English lyrics, shown below:

> Based on the (English/Korean) lyrics, can u create a feature vector where it has [binary 0 (not a love song) or 1 (it is a love song), emotion evoked (happy, sad, angry, bittersweet, neutral), % Korean lyrics as number of Korean words / (number of Korean words + number of English words), and sentiment intensity from a scale of 1 to 5, with an explanation why the score is given]

On a separate spreadsheet, I manually encoded whether it was a love song or not and each emotion as a binary variable (0 for no and 1 for yes) and encoded the percentage of Korean lyrics (for which I did manually as ChatGPT proved to be unreliable) and the sentiment intensity as integer variables. I later appended this to the original dataframe to complete the data.

	A	B	C	D	E	F	G	H	I	J	K	L	M
		Album	Artist	Track Name	love song	happy/positive/h	sad	calm/affectionate	bittersweet/yearr	anger	korean percent	sentiment intensity (1 to 5)	
1		GFRIEND 1st Mini	GFRIEND	Glass Bead	1	1	0	0	0	0	1	4	
2		GFRIEND 1st Mini	GFRIEND	Neverland	1	1	0	0	0	0	0.6567	4	
3		GFRIEND 1st Mini	GFRIEND	White	1	1	0	0	0	0	1	4.5	
6		GFRIEND 2nd Mini	GFRIEND	Me Gustas Tu	1	1	0	0	0	0	0.9	5	
7		GFRIEND 2nd Mini	GFRIEND	Under The Sky	0	1	0	0	0	0	0.9762	3	
8		GFRIEND 2nd Mini	GFRIEND	One	1	1	0	0	0	0	0.8526	5	
9		GFRIEND 2nd Mini	GFRIEND	My Buddy	1	1	0	0	1	0	0.8176	3.5	
12		GFRIEND 3rd Mini	GFRIEND	Rough	1	0	1	0	1	0	1	4	
13		GFRIEND 3rd Mini	GFRIEND	Say my name	1	0	0	0	1	0	0.94	3.5	
14		GFRIEND 3rd Mini	GFRIEND	Luv Star	1	0	0	0	1	0	1	5	
15		GFRIEND 3rd Mini	GFRIEND	Someday	1	1	0	0	0	0	0.9581	4	
16		GFRIEND 3rd Mini	GFRIEND	TRUST	1	0	0	0	1	0	1	4	
19		GFRIEND The 1st A	GFRIEND	Fall in Love	1	1	0	0	0	0	1	4.5	
20		GFRIEND The 1st A	GFRIEND	NAVILLERA	1	1	0	0	0	0	1	4.5	
21		GFRIEND The 1st A	GFRIEND	LOL	1	1	0	0	0	0	0.8675	5	
22		GFRIEND The 1st A	GFRIEND	Distance	1	1	0	0	0	0	0.9762	5	

Figure II. Manual encoding of lyrical data

I then created another dataframe where all of the musical variables were normalized within their respective columns for easier comparison across different scales as seen with the raw audio data. I then averaged the z-scores for each album to create the feature vectors that will be used in the similarity score computation.

Album	Popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	Album_ID
GFRIEND 1st Mini Album 'Season Of Glass'	0.599551	0.643530	0.053063	-0.030990	-0.212078	0.794491	-0.156036	-0.265361	0.136763	-0.134018	0.067934	-0.504614	1.0
GFRIEND 2nd Mini Album 'Flower Bud'	0.324866	0.321339	0.455818	-0.148750	0.106968	0.794491	-0.257550	0.188732	-0.117877	0.266041	0.652889	-0.586745	2.0
GFRIEND 3rd Mini Album 'SNOWFLAKE'	0.223852	-0.270262	0.136212	0.133875	-0.057626	0.386509	-0.047496	-0.041864	-0.117877	0.482859	-0.011083	-0.872411	3.0
GFRIEND Summer Mini Album 'Sunny Summer'	0.032459	0.537967	0.664212	0.345844	0.548314	0.386509	0.300800	-0.206577	-0.117877	-0.196720	0.781771	0.127982	9.0
GFRIEND The 1st Album 'LOL'	-0.297164	0.093053	0.097755	-0.042766	0.063386	-0.021473	0.312139	0.007240	-0.116127	0.085650	0.145991	0.592858	4.0
GFRIEND The 2nd Album 'Time for us'	-0.251088	0.335910	-0.452764	0.204531	-0.223113	-0.395456	-0.114849	0.312715	-0.117877	-0.133902	0.140813	-0.295173	10.0
GFRIEND The 4th Mini Album 'THE AWAKENING'	-0.180201	0.243624	0.371803	0.263412	0.044854	-0.225464	-0.182202	0.331050	-0.074970	-0.295524	0.452470	0.115301	5.0
GFRIEND The 5th Mini Album <PARALLEL>	-0.197923	0.178862	0.207236	-0.207630	0.122168	-0.565449	-0.421087	-0.231923	-0.117877	0.561017	-0.306054	-0.107657	6.0
GFRIEND The 5th Mini Album Repackage <RAINBOW>	-0.592229	-0.062376	0.257039	-0.192910	0.178546	-0.225464	-0.463607	-0.381181	-0.117877	0.463030	-0.440306	-0.209147	7.0
GFRIEND The 6th Mini Album <Time for the moon night>	0.316005	-0.781234	-0.574024	0.263412	0.059323	0.114521	-0.403643	0.465110	-0.113783	-0.493063	-0.411538	0.153323	8.0
GFRIEND The 7th Mini Album `FEVER SEASON`	-0.438431	0.401135	0.501476	-0.098281	0.487732	0.211660	0.548504	-0.190848	-0.117877	-0.032426	0.580557	-0.166419	11.0
回:LABYRINTH	0.280561	-1.126091	-0.667567	-0.089870	-0.296992	-0.225464	-0.202069	0.598861	-0.117877	-0.254511	-0.596614	0.676689	12.0
回:Song of the Sirens	0.386891	-0.480091	-0.468355	0.440052	-1.165857	0.454506	0.822761	-0.562048	-0.111792	0.221784	-0.844021	0.900425	13.0
回:Walpurgis Night	0.525442	-0.003207	-0.028197	-0.502032	0.215904	-0.132741	0.061396	-0.160231	0.800159	-0.296177	-0.011397	-0.217628	14.0

Figure III. Average z-scores of each album (audio data only)

I will then conduct the comparison of each of the nth album's tracks with the (n-1)th album. I computed the similarity score using the cosine similarity formula and plotted them in chronological order (see Figure IV).

VISUALIZATIONS

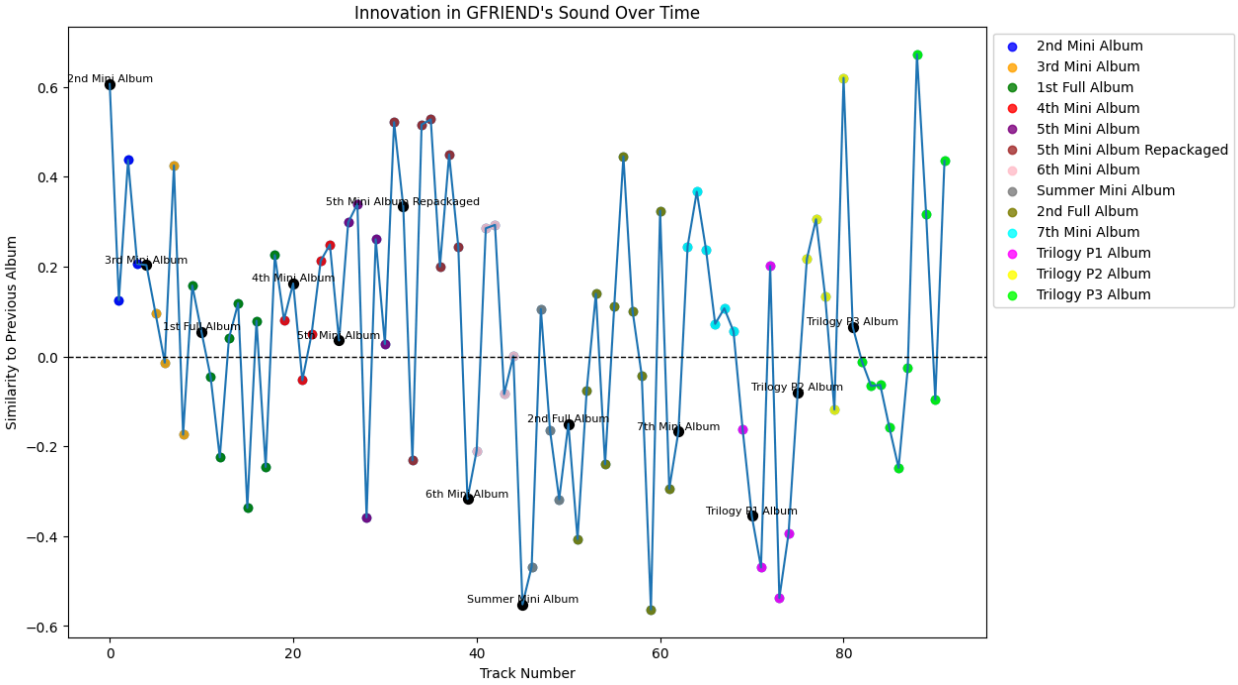


Figure IV. Innovation of GFRIEND's entire discography throughout time, by comparing the scores of individual tracks of the nth album to the (n-1)th album's average score

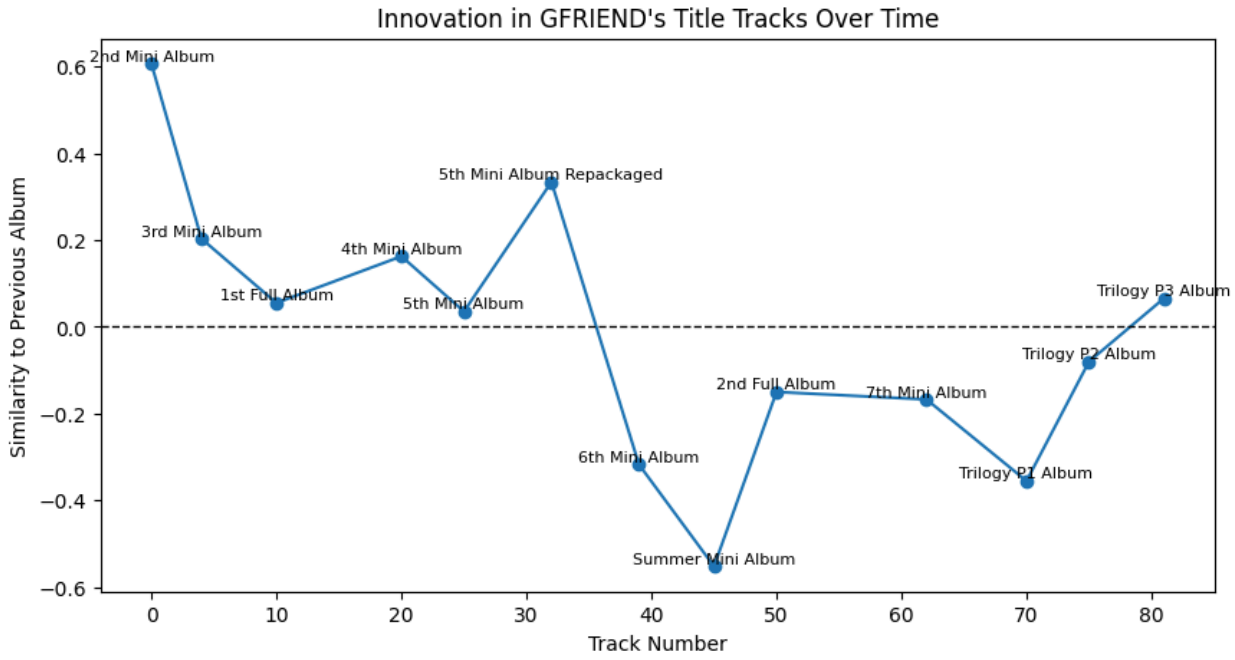


Figure V. Title track innovation over time for GFRIEND

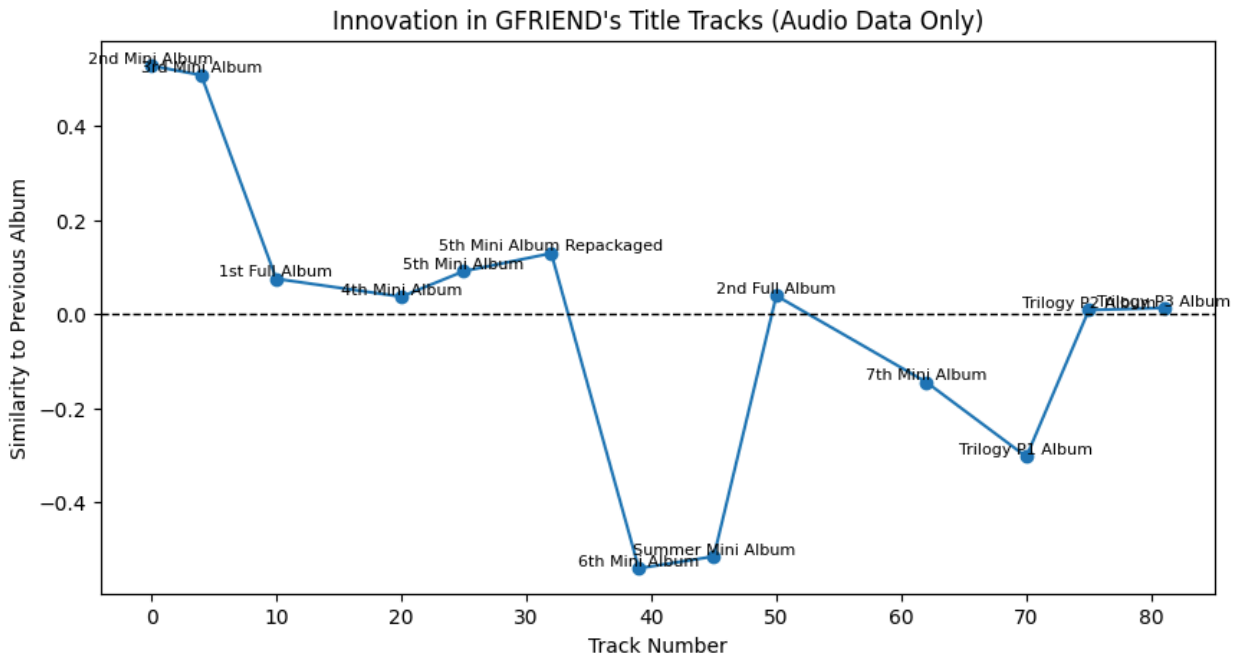


Figure VI. Title track musical innovation over time for GFRIEND

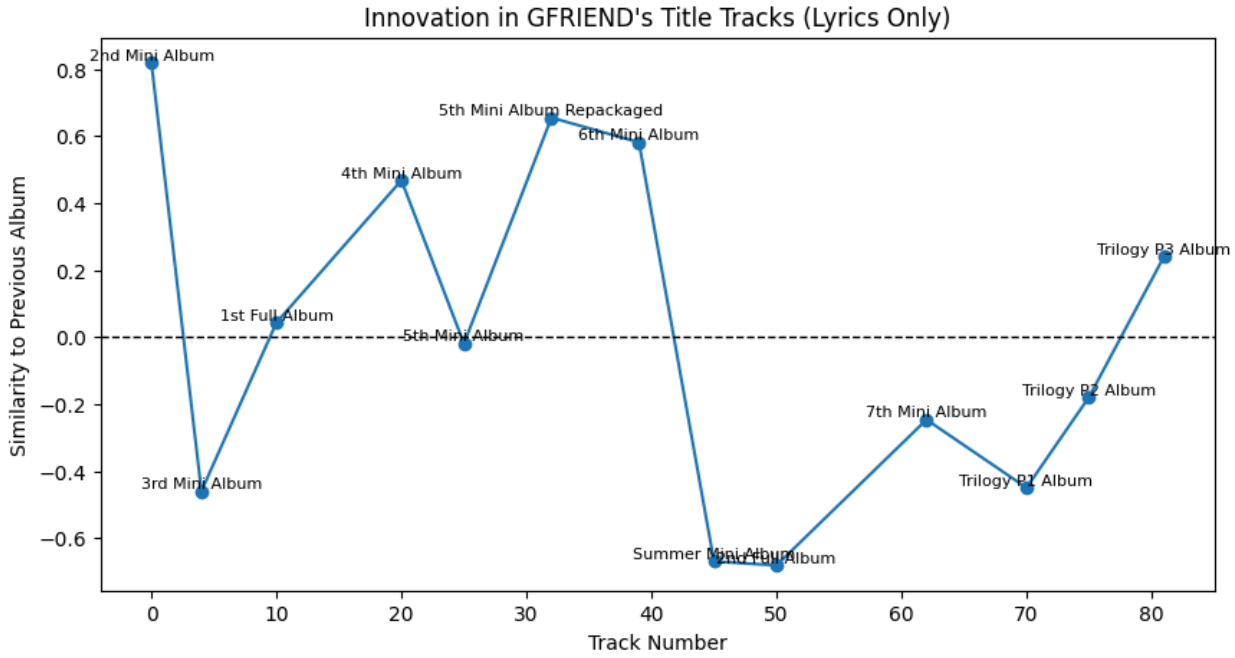
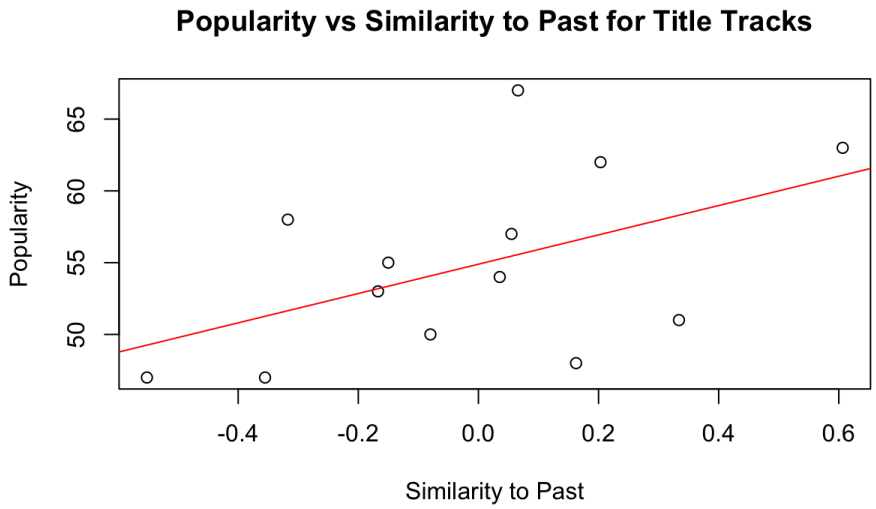


Figure VII. Title track lyrical innovation over time for GFRIEND




```
Call:
lm(formula = popularity ~ similarity_to_past)

Residuals:
    Min      1Q  Median      3Q     Max
-8.5553 -4.0776 -0.1856  1.9093 11.4327

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         54.894      1.617  33.956 1.73e-12 ***
similarity_to_past  10.217      5.450   1.875   0.0876 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.824 on 11 degrees of freedom
Multiple R-squared:  0.2422,     Adjusted R-squared:  0.1733
F-statistic: 3.515 on 1 and 11 DF,  p-value: 0.08759
```

Figure VIII. Linear regression of popularity vs similarity to past with output

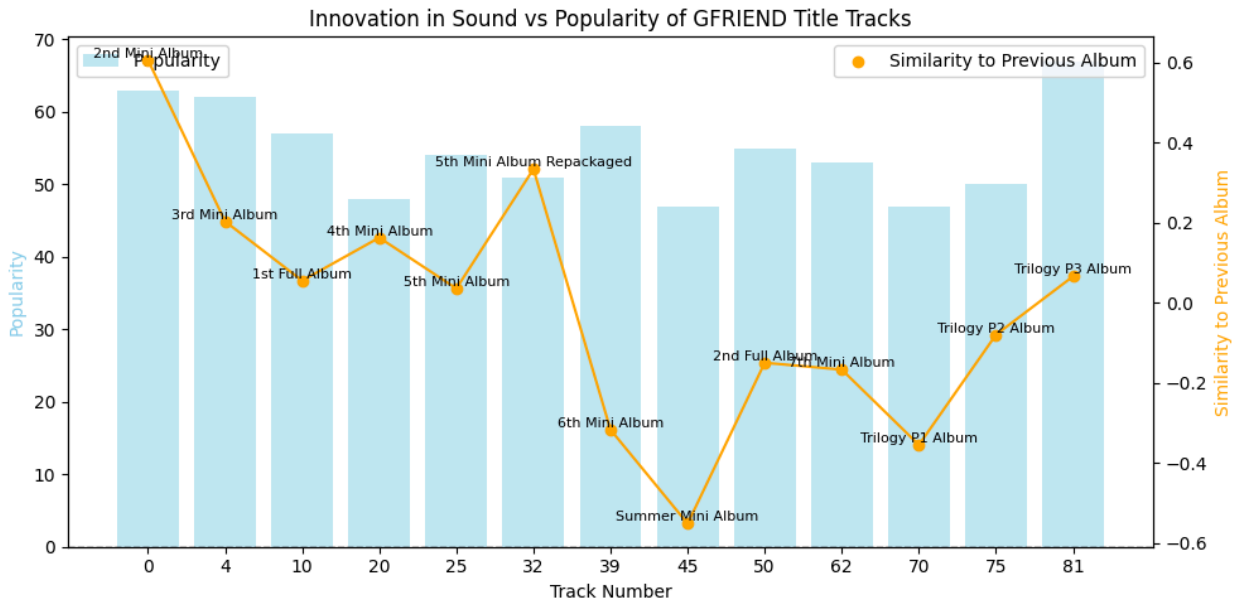


Figure IX. Staggered view of similarity trend and popularity trend

RESULTS

Out of the 95 lyrical songs GFRIEND has in their discography, 74 are love songs, 60 evoke positive, happy, and hopeful emotions, 4 evoke sadness, 15 are calm and affectionate, 42 capture bittersweet and reminiscing feelings, 2 evoke anger and 3 evoke sensual and seductive feelings. The average percentage of Korean lyrics overall is around 85% and the average sentiment intensity is around 3.95 on a scale of 1 to 5.

From Figure IV, we can see how noisy the entire discography could become when considering all of the different audio and lyrical variables. Within K-pop, only the title tracks are usually promoted and paid attention to as that is usually what is considered the main song that encapsulates the visual and musical themes of a single album. As a result, taking the title tracks will create a more clear visualization of how the public perceives the sound to change.

From Figure V, compared to the hypothesis we presented at the beginning, there seems to be much more dissimilarity than similarity occurring and seems to deviate a lot more than the predicted shape. From the 2nd mini album's title track having a 0.6 similarity to the 1st mini album's title track, that similarity quickly drops to around 0.2 and then close to 0 at the 1st full album's title track. This drop was not to be expected theoretically as usually, a group develops their sound and increases in similarity, not decrease. Also, it should be noted that the 5th mini album repackaged is the same as the 5th mini album except for a new title track and one new additional track. This title track seems to follow the same ambiance in sound as the 5th mini album.

The 6th mini album takes a large drop at around 0.3 dissimilarity, marking the biggest departure in sound. As mentioned earlier, the group has been known (at least on Reddit) to have two big changes in musical identity and the 6th mini album Time for the Moon Night was one point where people began to notice that, supported by the data we have for the visualization. A summer album right after that album takes more artistic freedom and deviates further from the previous album and the next full album embraces the new sound as shown with the rise. The 7th album in a way shows slight variation from the second full album and we experience a second noted change in sound.

GFRIEND's company had been acquired as a subsidiary of HYBE CORPORATIONS during the same month as the release of their 7th mini album, and as a result, changes can be expected as it was not the same people in charge anymore. They released a trilogy series of albums in 2020, which follow the same musical trends and show the development of a new kind of musical identity as the similarity scores increase and reach around 0.1 with their final album release.

We can try finding the source of these divergences, whether it is found in the lyrics or in the music itself. From Figure VI, we can see that musically, the 2nd and 3rd albums are similar but we witness a drop when comparing the first full album's title track with the third album, which builds a new trend of similarity until the expected decrease with the 6th mini album's title track and the summer mini album's title track. The 2nd full album's title track is significantly more similar to the previous release before it but it only drops further from there onwards with the 7th mini album and the first album of the trilogy series. From the trilogy releases, we can see the penultimate album bearing a slight resemblance to the previous release and the final album replicating that resemblance as well.

From Figure VII, we can see that lyrically, there are a lot more interesting changes. There are several decreases and increases in similarity relative to previous releases, with the two most drastic ones being the one from 2nd to 3rd mini album and the one from 6th to summer mini album. From the lyrical data, it can be seen that the 2nd mini album was very happy sounding and used minimal amounts of English words, and comparing that with the bittersweet, fully Korean lyrics of the third mini's title track, it can be seen why such a big dissimilarity score was given. The opposite was seen in the second drop where the sentimental and sadder feelings of the 6th mini album were compared to the bright happy title track for the summer mini album. One thing that cannot be refuted is the consistency of the trilogy to become more and more similar in terms of lyrics to the previous album before it, with the final album having a similarity score of around 0.3.

We can analyze the relationship between similarity to the past and popularity through a regression analysis. From Figure VIII, we can observe a weak positive correlation between similarity to the past and popularity, as indicated by the r-squared of 0.2422 and the trendline. There does not appear to be a significant relationship between similarity to past and popularity as indicated by the p-value of 0.0876. For the entire discography, there was a slightly negative relationship where an increase in similarity correlated with a decrease in popularity but the relationship was not significant by a larger p-value than the previous one.

In the end, it may not be that people's interests entirely lie within being attracted to innovation or similarity as there are no drastic trends shown in Figure XI as the drops seen at the first full album and the 6th mini album do not show extreme changes. Some increases in similarity can be seen with rises in popularity and the same can be said for decreases and drops in popularity. One clear trend is the trilogy releases being an upward trend for both popularity and similarity. We can take this as a sign that innovation for GFRIEND has brought them consistent success but not anything that severely underperformed or anything that became a complete gamechanger. We can argue that innovation has brought success as, MAGO, their final title track, is their most streamed song, surpassing the second most popular one by 40,000 streams at a total of 126,000,000 streams. With this track being the height of their departure from their debut musical

identity, it can be hypothesized that people can welcome innovation after long periods of consistency, but further dives into their major departures in music identity can support this claim, possibly looking into sparking conversations on Reddit.

CONCLUSIONS

A group's sound evolves through time, and as shown in the graph, that evolution is not very consistent with the human perception of music. Despite human perception showing pretty good consistency in music, using the similarity score with feature vectors has shown that there is much more than meets the eye.

With the combination of audio and lyrical data, GFRIEND has been through a lot more musical variation than what is commonly and publicly perceived by Reddit and humans in general. Their beginnings were rougher than what people would anticipate for songs with relatively similar-sounding vibes. The departure from their original innocent sound was noted and supported by the data, with a consistent drop for the 6th mini album for the overall evaluation and the audio-only version. The second departure in sound was seen with the release of the trilogy albums, which stayed consistent throughout all three different figures featuring different subsets of the data. For the audio data, we saw three main drops that were reflected in the final data that incorporated the lyrics as well whereas the lyrical data saw 2 main drops. Their innovation in music brought them relatively consistent success in terms of the popularity of the songs are not too dispersed in distribution.

We cannot say for certain how people react to similarity or innovation from this data because of the weak relationship between similarity and popularity and the lack of effect that the popularity and similarity had when compared side by side in the last figure. While GFRIEND's similarity to past albums can be shown as their commitment to their group image and musical identity, there is no doubt that had they continued their musical careers as a group, they would have traversed far different paths than those recorded on this paper. Had they continued off the success of MAGO, we could have seen more of their innovation in music and witness whether people will continue to welcome their innovation.

Within the larger scheme of things, this project may prove the point that there are too many outside factors to consider when using this type of methodology to examine the medium of music. Pop music is very popular, with many listeners and many releases coming out constantly. While this project sufficed for examining the evolution of music within the scope of one artist, it can be harder to say based on the larger scale of the entire pop music market.

These results and interpretations may also potentially downplay the importance and impact of information cascades. There may be a lot of insider knowledge regarding the success of music because people know what is trending and popular and know that capitalizing on that will bring success. However, not everyone will choose to act on the signals that are being sent and implied. Album makers may just simply look at the previous album to determine the next album they will

make and how successful the previous one was, basically sticking to what they know is best regardless of outside factors.

Links

Jupyter Notebook with all of the data processing and analysis

Lyrical Data Encoding