Observing Musical Communities Dedicated to Improvisation on TikTok Using Web Scraping
Marc Chemillier (CAMS-EHESS), Yohann Rabearivelo (CAMS-EHESS), Rémi Jaylet (Télécom Paris)


Abstract
In this paper we present a research dealing with musical improvisation practices on TikTok as part of a project devoted to the development of a music improvisation software. Finding musicians to experiment with led us to explore the musical collaborations encouraged by the platform. The duo function appears to be one of its most important feature from this point of view. Tracking relationships between musicians who duet with each other leads to the definition of musical communities that form on the network. These online communities pose new problems to the classic ethnographic approach to communities, and their study can benefit from new methods involving web scraping techniques. This article shows how to use these techniques in the context of TikTok and analyses some properties of the graphs that can be drawn from web scraping data collected within the musical communities that are formed thanks to the duet function. Certain properties of the graph, such as source and sink nodes, are highlighted and linked to musical roles such as soloist and accompanist.

Music plays a fundamental role on TikTok, particularly through dance challenges that promote songs to dizzying levels of listening\(^1\). The music industry has recognized the importance of this phenomenon for a few years, as evidenced by Microsoft's attempted acquisition in the summer of 2020\(^2\). Communities have formed around musical practices fostered by TikTok such as the duo function that allows one to complete a pre-existing video by recording a new video placed next to it. This type of musical collaboration has been studied recently by D. Bondy Valdovinos Kaye\(^3\), but research on music is still scarce among academic works on TikTok which rather deal with video content analysis such as dance challenge videos where visual elements, gestures, facial expressions and paratextual characteristics are studied\(^4\), or economic aspects related to the attention economy\(^5\) or to the question of algorithmic precarity\(^6\).

In this paper we present a research dealing with improvisational practices experimented by the musicians of TikTok. As we shall see, it is related to a project devoted to the development of a music improvisation software. During the pandemic we were driven to look for musical collaborations online to conduct experiments with the software. This is why we started to take an interest in musical practices specific to TikTok such as the duo function mentioned above. These


new practices give rise to communities of people who collaborate with each other musically. We have started an ethnographic study of these musical communities using classical methodological tools such as participant observation. But we also took advantage of the digital nature of the data stored on the TikTok platform in order to explore the possibilities offered by web scraping techniques. This last point is the main subject of this article. In the first part, we recall the context of this research related to computer improvisation issues and we describe the kind of musical collaboration that are made possible using the TikTok duo feature. In the second part, we present the scripts of web scraping that we have designed in order to explore the network of these collaborations. In the third part, we describe some properties of the graph representing this musical community. Such properties concern the different roles that TikTok musicians adopt in their online practice, and the motivations that lead them to invest their time in the platform.

**TikTok duo function for music collaboration**

Our study focuses on the improvisational practices developing on TikTok and on the way they adapt to the duo function and to other characteristics of the application (very short video format, possibility of live broadcasting). Our motivation comes from a research that we conduct on computer improvisation and the development of a music software that can improvise. It relies on the modeling of sequences played by real musicians or stored in a data base. It functions as an interactive system that can play in real time with a human instrumentalist and reply to him by recording what he is playing and improvising with it. We develop this system through fieldwork studies that give us the opportunity to interact with expert musicians. Our interest focuses on idiomatic improvisation such as music from Madagascar as we shall see below. This requires ethnographic survey with both observer and participant approaches. Computer science and ethnographic fieldwork are thus combined into a pluridisciplinary research. The participant aspect of our ethnographic studies implies that we ourselves play music with the computer in collaboration with musicians from a given cultural context. Let us give a short example.

The video (figure 1) shows an improvisation by the computer playing with the group of Malagasy guitarist Charles Kely Zana-Rotsy. The music generated by the software is controled by a human who press various multicolor pads. Sequences played by the machine are based on a corpus including some transcriptions from the great jazz musician Ray Charles. Thus in this passage, some phrases by Ray Charles are integrated in the musical context of Malagasy music. The video published on TikTok points out this relationship by playing the Ray Charles original recording at the end of the computer improvisational excerpt. The issues of integrating a computer generated improvisation in the context of idiomatic music such as the one played by this group of world music artists is discussed in a paper by Marc Chemillier.

How does this research relate to TikTok? During the pandemic, we were not able to make fieldwork with musicians so we began to study music collaboration on TikTok. As part of the participant dimension of our ethnographic approach, we also began to make videos. TikTok duo features promotes musical collaborations, and it is a very important features of the plateform. These collaborations define musical communities on TikTok in the sense that people interested in the same kind of music tends to collaborate. Our own interest lies in improvisation and our presence on TikTok was intended to meet (and try to collaborate with) other musicians sharing the same interest.

---

Let us briefly explain how the duo function works. When you have a video on TikTok and you want to share it, one of the possibilities you have is to use the duo button as shown in the picture on the left (figure 2). When you press this button, the video is reduced to the half part of the screen, and you can record your own duet part as a video which appears in the other half part of the screen. Generally speaking by doing so, you add in the caption of your video some specific informations related to the duet with the hashtag #duetwith followed by the user name of the person who made the initial video. The picture on the right (figure 2) shows the resulting split screen and the corresponding caption. In this case, we added a computer solo on a musical groove played by Mezerg who used a very special interface designed by Playtronics that can connect a synthesizer to fruits such as watermelons in order to play sounds.

There are some pretty sharp issues that arise technically in the production of these duet videos. How to mix the sound in a professional way? TikTok offers interesting tools for video editing, but there are no functionalities related to the mixing of the sound. Thus, TikTok musicians can choose between two methods to create duets from pre-existing videos on TikTok.

The first one is the most direct. It consists in using the duet function available in the parameters of the reference video. The user can directly record his video next to the reference video that plays in the background. Thus it requires users of this function to be able to transmit sound to the phone while listening to the reference video that plays in the background to be synchronous. Some users, especially singers, use a hands-free kit, but instrumentalists have to invest in audio interfaces connected to their smartphone, allowing them to connect their instrument while listening to the sound coming out of the phone. Moreover, this feature only allows one take, and does not permit to save several takes, which implies a successful execution in one go. When the video is finished the author of the reference video is automatically mentioned in the caption of the duet video, as well as the reference sound assigned to the author of the reference video.
The second method used is to download the reference video and edit it with one's own video using specialized sound and video software, and then to upload the resulting video back on the platform. This second method is less spontaneous because you can record several takes on a mixing software, or even synchronize the tracks after recording. When using this method, the musician making the duet will have to manually mention the author of the reference video if he wants to, since it is not done automatically by the platform, and will have to fill in the original sound. Whether this mention is added manually or whether it is put automatically by the platform, in both cases the author of the original video receives a notification that his video has been the subject of a duet. This way, the user mentioned has the possibility to react or not.

The duo function of TikTok has been studied in a paper by D. Bondy Valdovinos Kaye mentioned above\(^8\). The author surveyed a community of TikTok musicians who identify themselves under the hashtag #jazztok. This survey is interesting for our work. First, it traces the history of the duo feature, which is very specific to TikTok and available to its users. Then, he observes that this function was used by the musicians of #jazztok allowing them to play together offline during lockdown. In this way, they were able to add successive musical parts to a given video in collaborative jams sometimes gathering up to 7 or 8 musicians. In such cases, the author speaks of "duetchains", a notion that we shall recall later in the present paper. These are duos of duos, creating complex jams in the manner of an exquisite corpse. The author's investigation is based on the testimonies of these #jazztok musicians with whom he was able to speak in interviews (18 recorded interviews made with Zoom and then transcribed using an automatic device). Indeed, we learn that it is through the online musical meeting including the use of the duo function, that users were able to develop "duet jams", a notion inspired by Sawyer and DeZutter’s theory of distributed creativity\(^9\). So they could locate and connect with each other, and build a relationship continuing with the live function of TikTok allowing them to discuss in real time, or while meeting on other platforms like Discord.

Duets on TikTok are widely used by musicians to expand their audience in the hope of increasing their listenership on streaming platforms\(^10\). A recent article attempts to examine the effects of the

---

8 D. Bondy Valdovinos Kaye, “Please Duet This: Collaborative Music Making in Lockdown on TikTok.”
duet function on the career development of artists, particularly independent artists who would like to become better known. The authors examine three cases including the case of Stacy Ryan, a member of the jazztok community described by Kaye11. Her audience as well as the number of listens she got would have increased on the Spotify and Instagram platforms, a few hours after the publication of her #openversechallenge video where she proposed to sing with her via the duet function. This article shows many coincidences between the use of the duet function at promotional moments by more or less famous artists. A considerable increase in the number of listeners and followers on the music streaming platform Spotify can be observed in the days following the publication of videos using or requesting this function. The article ends by suggesting that “the duet feature should be considered an extremely powerful tool that has the potential to generate virality for creators of all sizes and genres.”

Scrapping the musical communities on TikTok

Around each TikTok musician, the collaborations encouraged by the platform make it possible to define a community of people sharing a common interest. In classical ethnomusicology, the notion of community generally refers to a determined geographical space such as a village or a group of villages, where people speak a common language and participate to common social activities including music. But we know since the arrival of jazz that the appearance of the recorded disc has led to the deterritorialization of this notion of community. This process of creation of new musical communities from the emergence of new practices has been well described about jazz by Jean Jamin and Patrick Williams among others, and it is central in our study since idiomatic improvisation arises from this pooling of musical practices:

“through borrowings and repeats, quotations, shared turns, the improvisations of one which become a theme for the other, jazz musicians can give the impression that they live on a common stock of formulas and that they take part in the same invention. Following the paths of this sharing and showing how all these individual creations relate to each other is an essential work: it is to do ethnology without leaving music as an object. At the end of such an examination, the idea of community reappears; but a community based on a way of conceiving and practicing music – a way of conceiving and practicing society.”12

Video platforms like TikTok push this process even further by creating new musical communities around practices introduced by the platform and shared very quickly by an increasing number of people. The speed at which ideas and cultural traits spread over this type of network is new data that upsets the traditional notion of cultural context, and consequently the notion of idiomatic improvisation which derives from it. Several theoretical approaches have attempted to identify this phenomenon, whether it is the notion of the meme introduced by Richard Dawkins13 or that of the epidemiology of representations described by Dan Sperber14, and these theories find a fascinating field of application on a platform like TikTok15.

To describe the musical communities in which new cultural traits spread via the network, we use the usual methods of video analysis and ethnographic surveys conducted with TikTok musicians, but we also use web scraping techniques to highlight the relationships between musicians and the

---

11 D. Bondy Valdovinos Kaye, “Please Duet This: Collaborative Music Making in Lockdown on TikTok.”
15 Lou-Anne Donguy, “TikTok, l'influence de l'application et des réseaux sociaux sur l'industrie de la musique” (M1 Master thesis, Sorbonne University, Paris, 2022).
musical communities that are formed through their practices. This part of our research constitutes the main subject of the present paper. Thanks to a recursive algorithm, we could get the collaborations of different musicians by studying their network made of the tagged artists in their videos’ captions and those with whom they directly collaborated via the duo feature. By filtering the profiles according to their characteristics and used hashtags, we could target the users inclined to interact with other musicians and to experiment with different forms of improvisation.

More precisely, when a musician duets with another, he usually mentions it in the caption of the video as we have said before: #duetwith @someone. We are interested in this kind of collaboration, and thus we study the general relation "X mentions Y in the caption of a video." Depending on the hashtags, it might indicate various musical relations (#duet, #cover, #reply to a comment, etc.). By following such relations, one can draw the community of people who share some common musical interest.

The advantage of a social network like TikTok is that these links are represented digitally. Thus it is possible for a computer program to follow them automatically and to collect all the people connected to a given musician through these relations of collaboration on the platform. This kind of search can be done thanks to an algorithm of web scraping based on an API (application programming interface) that can provide data extracted from TikTok. Similar studies using web scraping techniques have already been conducted successfully on various subjects such as amateur online film critics\textsuperscript{16} or the analysis and the visualization of political opinions from Twitter data\textsuperscript{17}.

We use the unofficial TikTok API designed by David Teather\textsuperscript{18}. Based on this API, we designed several scripts written in the Python language in order to collect the relationship "X mentions Y in the caption of a video." The program follows these links through the TikTok network and collects the corresponding data (URL of the videos, user names of the Ys, etc.). For a given TikTok user X, it proceeds in a recursive way by computing

-  at first step all the Ys mentioned by X,
-  then all the Zs mentioned by each Y previously found,
-  then all the Ws mentioned by each Z previously found, and so on.

The resulting set of people gathering X and the Ys, the Zs, the Ws... build some kind of musical community.

Here is an example of the execution of the script with initial parameters given at the beginning. In this case, the search is done by following the links from only one user (@digitaljazz), by iterating the search with only 10 of the people mentioned by each user (n_user=10) and by visiting only 30 videos for each user (n_vid=30). These limitations are necessary to avoid searching for too long. The first iteration searches for all people who have been mentioned by @digitaljazz (limited to 10 people). The other iterations will repeat the same search from people found in the previous step and so on.

```
% python3 tiktok_music.py
Press 'enter' to validate the default values
Number of users to initialize with (start = 1): 
Usernames to initialize the algorithm: digitaljazz
Number of users to end the algorithm (n_user = 10): 
Max number of visited videos (n_vid = 30): 
```


\textsuperscript{17} Nicolas Gaumont, Maziyar Panahi, David Chavalarias, “Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election,” \textit{PLOS ONE} 13, no. 9, (2018), \url{https://doi.org/10.1371/journal.pone.0201879}

\textsuperscript{18} David Teather, Unofficial TikTok API in Python (2022), \url{https://github.com/davidteather/TikTok-Api}
Max number of followers (max_followers = 50000):
Keep only duo relationships ? (only_duo=False):

iteration # 1 : digitaljazz
etc.

The result is written in a CSV file. In the first column, you have all the user names that have been found. In the corresponding rows, you have the biography of their TikTok account, then in a further column the list of user names that have been mentioned in the captions of their videos. The script tries to selected among them the ones that may correspond to musicians (this is done by taking into account the hashtags chosen by these users as we describe below) and we also give the URL of the videos where the mention can be found. One can imagine that this search is far from being perfect. On the one hand it is not exhaustive since one has to restrict arbitrarily the search to avoid excessively long calculation time (that is done by the parameter n_user). One the other hand it brings too many results that are not interesting from a musical point of view (since a musician can mention other users who are not musician themselves).

Hashtags provide a possible way to refine the web scraping algorithm. They are keywords preceded by the symbol ‘#’ which are written by Tiktok users in the caption of the videos they produce. They allow to reference their content and constitute very interesting metadata accessible to the web scraping API. Thus, the script also contains a certain number of keywords that allows it to select or not a content to analyze. In our case, if a video contains at least three of these keywords related to music (for example: #synth, #groove, #jazz) it will be selected as a music video and the script will continue working.

Participant observation led us to enrich and diversify the list of keywords included in the script. Indeed, we could observe that the musicians of TikTok have different uses of hastags which vary according to the beliefs and practices of each one with regard to the algorithm of Tiktok. Some use a lot of keywords to catch the attention of the algorithm and try to diversify them as much as possible in the hope of appearing on the page for you from a large and varied set of users. Others use very specific hastags to fit into more specific communities. Finally, others think that Tiktok favors content that is not very voluminous in text and uses little or none in the idea of producing fluid and spontaneous content. All these conditions also frame the relevance of the results brought by the script and its exhaustiveness. They are also at the origin of relevant questions concerning the role of the hastags in the identification with communities by TikTok musicians.

At the end of the web scraping process, one has to refine manually the CSV file obtained. Moreover, the calculation is performed at a given time. It is obvious that the result would have been different if it had been done at another time because TikTok network is constantly evolving. Thus any calculation should be regarded as a kind of snapshot. From this instant image, one can load the resulting CSV file into the Gephi application which can display the graph of the relationship ”X mentions Y in the caption of a video" for all the users gathered by the script during the web scraping process.

One of the biggest issues of the scrapping part is due to the many changes in the privacy policy of TikTok. Indeed, the numerous modifications of the application makes the maintenance of the algorithm quite complicated. The access to some of the parameters can be suddenly removed because the algorithm depends on an unofficial and open-source API that is reliant on the updates made by the different persons involved in the project to try to counter the various modifications of TikTok. Moreover, even when everything is working, the interface is able to quickly detect recurrent calls from the same user. It appears therefore necessary to regularly change the connection token as well as to carry out a proxy rotation to make sure not to be blocked by the platform.
Furthermore, it is quite difficult to have access throughout the API to certain of the features. For example, different forms of links are potentially exploitable, however a large part of them are in practice not usable like the subscribers’ list which is impossible to scrap or the list of liked videos, which is by default on private in most accounts. This is why we only focused on the mentioned links (a user mentions with ‘@’ another user on a particular video). The relationships between users are separated into two forms of link: duets (interaction with the video of other users) and direct mentions. This type of link is not the most common but is actually the most significant as it shows a real collaboration between different users.

Properties of the music collaboration graph

We visualized the relationship between musicians in the form of a graph using the Force Atlas 2 algorithm as implemented in the Gephi software. We can thus observe some specific properties as for instance the dense points, the sparse areas, the possible separation of the graph into connected components, the sub-graphs corresponding to specialized mentions (duets, covers, etc.). We are also interested in the motivations that push musicians to invest in these new practices.

The resulting graph is directed so that we use arrows in the display of the collaborations between musicians by distinguishing "X mentions Y" from "Y mentions X." When it applies to duos, the relationship involves two distinct musical roles (soloing, accompaniment) corresponding respectively to source and sink nodes in the graph. A node is considered a source if there is no arrow coming from another node to this one. It mainly corresponds to solo instruments like sax, voice, etc. who can improvise on a given video. Likewise, a node is considered a sink if there is no arrow going from this node to another one. It is often associated with beatmakers who propose grooves that solo instruments can improvise with. In the graph displayed below (figure 3), there are two source nodes and one sink node that are surrounded by a rectangle.

![Figure 3: A music collaboration graph with source and sink nodes](image)

The fact that a user is a source node means that every arrows connected to him are coming out of this node. From a musical point of view, it means that he makes duos with other people but no one makes duos with him. Generally speaking, it corresponds to soloist musicians but in the current case, the two source nodes correspond to computer improvisation softwares: one is the computer program we are working on as described above @digitaljazz, and the other one is an expert musician from Bulgaria who works on a similar program @ivan.dragolov and that we have met...
through TikTok. In the graph (figure 3), you also have a sink node @jonah.nilsson with every arrows connected to this node being oriented towards the node. In this case, it corresponds to the famous keyboard player Jonah Nilsson from the group Dirty Loops. He published a video with a nice groove associated with a challenge called #badsolochallenge. Many musicians were stimulated by this challenge and made a duo with him including @ivan.dragolov.

These roles soloist / accompaniement should probably be studied in relation to popularity. On can presume that a TikTok user who wants to increase his popularity will tend to make duos with other users having a bigger number of followers. This might be the case in general, but it is not an absolute rule. As a counterexample, @haywyremusic (7615 followers) proposed a challenge in one of his videos with the message “Trade solos with me on my new track!”. Then @oblvyn (70.3K followers) who is a quite skilled female synth player responded to the challenge by making a duo with him (figure 4).

![Figure 4](https://www.tiktok.com/@oblvyn/video/6950542822944574726)

There exist duet chains when someone duets a video that is already a duo as we have seen earlier by recalling the principle of exquisite corpse. This can be written in this way:

\[ A \rightarrow (B \rightarrow (C \rightarrow D)) \]

It means that C made a duet video with D, then B added a part to the resulting video, and then A also added a part to this trio. In the graph, one can find an example of duet chain going from @kingdavidsimien to @jonah.nilsson (figure 5).
It is based on the same #badsolochallenge mentioned above with musicians accumulating their parts to the initial video by @jonah.nilsson beginning with @alexengelberg on solo keyboard, then @patplaysdrums on drums, and then @kingdavidsimien on bass.

But graph chains may not be duet chains. They may correspond rather to role reversal. When you have the following chain $A \rightarrow B \rightarrow C$ in the graph, it generally means that $B$ plays two different
roles and that the video of his duet with $A$ is not the same as the video of his duet with $C$. More precisely, one has:

$A \rightarrow B$ as accompaniment

$B$ as soloing $\rightarrow C$.

In the graph displayed as an example above, we can find a path $A$ to $B$ to $C$ of this kind involving two different videos where $B$ is represented by @jacobcollier who plays two different roles (figure 7).

![Figure 7](image)

Figure 7: A chain with role reversal by @jacobcollier

This chain is an example of role reversal. Jacob Collier, the famous talented young British musician who is very active on TikTok and makes a lot of duets, proposed in this case the #allIneedchallenge for which @digitaljazz made a duet. But Jacob Collier did himself a duet for the #badsolochallenge proposed by Jonah Nilsson, and therefore changed his role from accompaniment to soloing.

The last graph (figure 8) was produced automatically while the previous one was constructed both automatically and manually by refining the data and selecting only the duet relationships. The latter graph is obtained from users identified from an online ethnographic survey that we are conducting and they regularly practice duets. The presence of many source nodes reflects their online music activity and their connection to other musicians, via features like TikTok duets. Here you can see double arrows, which means that roles are not clearly separated and that role reversal is frequent within these duet communities: two musicians can make a video with one role distribution and make another video where this distribution is reversed. Indeed, the survey reveals the real nature of these role inversions. They are often reciprocal duets (a musician makes a duet with another, and then the latter makes a duet with the first one), but there may correspond to other kind of relationship. It might be mentions by users who repost the videos of other users who participated in their duets. And it might also be video responses to comments (for example @garnetwaltersmusic responds to @tylarhairston as we shall see). Webscraping provides a broader view of these relationships between TikTok musicians.
The video comment response is not the result of a deferred collective creation between musicians as in the case of a duet, but it is nevertheless a feature that allows exchanges about music and is used by TikTok musicians. Here the pianist @garnetwaltermusic responds to the guitarist @tylarhairston who makes a remark in the comments space of a previous video in which @garnetwaltermusic proposes a duet by playing a sequence of chords to be reused. Then @tylarhairston suggests that what @garnetwaltermusic is playing may be reminiscent of a Kirk Franklin piece. In his response video @garnetwaltermusic plays the part again which indeed sounds similar confirming @tylarhairston's observation. Web scraping here highlights an exchange that demonstrates a real musical relationship between the two musical users. Web scraping allows to access to an overview of the relationships between TikTok musicians, but an ethnographic survey still seems important in order to verify the nature of the exchanges recorded by the API.

Conclusion

As we have seen in this article, musical practices on TikTok considerably renew the issues related to musical communities. The work presented here is part of a project dedicated to idiom music against computer, and it is interesting from this point of view, to deepen the study of the characteristics specific to the musical communities appearing on the network. One of the richest features of TikTok in terms of collaboration between musicians is the duet function which allows one to add a musical part to an existing video. This article takes a fresh look at the links that are woven through this practice by implementing a web scraping computer program. This allows us to

---

19 Original video by @garnetwaltermusic with comment by @tylarhairston: [https://www.tiktok.com/@garnetwaltermusic/video/7087303631313194283](https://www.tiktok.com/@garnetwaltermusic/video/7087303631313194283)

20 Video comment respons by @garnetwaltermusic whi the mention “Reply to @tylarhairston”: [https://www.tiktok.com/@garnetwaltermusic/video/7091693576127630634](https://www.tiktok.com/@garnetwaltermusic/video/7091693576127630634)
draw the graph of all the musicians who collaborate with one or more given musicians thanks to the
duo function of TikTok.

The visual representation of these graphs in the form of nodes and arrows reveal certain
characteristics specific to this type of musical collaboration. It brings up sources (i.e. musicians who
do a lot of duets with existing videos) and sinks (i.e. musicians offering videos that many others
duet with). This distinction must undoubtedly be linked to the popularity of users because intuition
suggests that those who offer duets are better known than those who make duets with them and who
see this as a way to increase their popularity, but this point deserves further quantitative study. From
a musical point of view, these two categories roughly cover the distinction between soloist musician
and accompanying musician, but there too the reality is more subtle (for example an a cappella
singer, therefore a priori soloist, can give birth to a duo with someone one wanting to accompany
his singing). All sorts of nuances exist and the fact that these categories are not clear-cut is perhaps
linked to the observation that many young musicians tend to be poly-instrumentalists. In the graph,
some configurations show these cases of confusion between categories. For example, the duet
chains correspond to duets on existing duets, in which the roles are mixed. Finally, the paths in the
graph during which the arrows alternately change direction reflect role reversals, that is to say that a
soloist playing with a musician becomes an accompanist with another.

**Fundings**: This research is supported by the European Research Council (ERC) REACH project,
under Horizon 2020 programme (Grant 883313) and by Agence Nationale de la Recherche (ANR)
project MERCI (Grant ANR-19-CE33-0010).