Emergence and Evolution of Network Structure in Rock Music

Ömer Yüksel[1,*] and Leon Danon[2,†]

[1]*Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK*
[2]*Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK*

(Dated: June 25, 2012)

We study the evolution of a musician network, using data obtained from a website containing information about bands, musicians and their recordings. The data is used to create three distinct dynamic networks reflecting different types of collaboration: bands, musicians and albums. We aggregate the networks over time and study the changes in network properties as we vary the aggregation parameters. We also track the evolution of individual nodes in the network over time. Studying the change in the connected components shows the influx of independent groups in the late 60s, and their subsequent addition to the giant component. Analysis of the degree distributions reveals an unusually high number of connections between the bands in the period between 1960 and 1970. Examining the distance related measures in the artist network reveal the emergence of central players with high betweenness in the 70s. Further study of betweenness values shows that the central positions originally held by the generation in the 60s is taken over by the nodes belonging to newer generations. Investigating the change in average distance of the individual nodes reveals that the nodes in the band and artist networks approach a stable value over time. The change in the k-core shows that the bands tend to form new connections as time progresses, whereas artists tend to play with same people throughout their career.

I. INTRODUCTION

Social network analysis is essential to understanding the interactions within communities. While it is recently a popular topic of research, the studies on social networks actually go way back to 1930s, two decades before the term 'social networks' was coined [1]. The study by Jacob Moreno [2] on a friendship network between schoolchildren triggered the studies on the field that eventually became known as 'social network analysis'.

An important milestone in the field was the 'small world' experiments by Stanley Milgram [3]. The experiments suggested that the shortest path length in the real world networks is quite small compared to the size of the network. This phenomenon eventually led to construction of network models with 'small world' properties, with the most famous one being the Watts-Strogatz model [4]. Another important property of social networks is that the people who join the network are attracted to the highly connected individuals. This led to several network generation models using 'preferential attachment' mechanism, such as Barabási Albert model [5].

The widespread use of the Internet has a role in the recent popularity of social network analysis in the literature. The websites like Twitter, Facebook and MySpace created new environments where the social networks could be studied by researchers. In addition, the world wide web opened up access to databases to the researchers [1], which reduced the time and financial resources required to obtain large amounts of data.

Music, being a collaborative process, also leads to formation of social networks. But the area of music-related networks is relatively untouched: the best known works are about communities of Jazz musicians by Gleiser and Danon [6], and rappers by Smith [7]. The work on Jazz musicians focuses on the collaboration between bands and musicians, and reveal the community structure, mainly determined by the musicians' race and their geographic location. The work on rappers also focuses on community detection using weighted edges and finds the geographic location to be essential in the formation of communities. There are also studies about the relationship between the music and the audience, revealing listening habits and formation of genres [8–10].

Both of the previous works on musician networks deal with static networks. However, social networks have dynamic properties. As time progresses, new actors may join the system and new connections can

*Electronic address: o.s.yuksel@warwick.ac.uk
†Electronic address: l.danon@warwick.ac.uk

be formed. Conversely, some of the connections and actors may cease to exist. In more complex cases, the properties of these connections and the actors may also change. A band, for example, can have new members, break-up, or come back together after several years. The network, and the members of the network, evolve in time. We aim to investigate these time-varying properties in a network of musicians.

The lack of subsantantial amount of work on dynamic nature of social networks is largely related to the expense of data collection in the past, and the complexity of such an analysis [11]. However, there have been a number of works on dynamic networks as well: few notable examples would be the works on face-to-face contacts [12], mobile phone calls [13], citations [14–16], e-mail messaging and affiliations in a university [17].

A number of the aforementioned studies deal with the evolution of network structure, which is also the main subject of our work. The work by Leskovec et al. [15] is focused on the change of the diameter and density in different types of networks over time. Their study is on citation, scientific affiliation and internet router networks. The work revealed densification and shrinking diameters in these networks. Since it deals with the changes in the network properties over time, we found this work particularly of interest, and made comparisons with the findings on the musician network.

The work of Kossinets and Watts [17] studies the evolution of a social network in a university, using e-mail and class affiliation data. The work investigates the changes in the average network properties, individual properties and their distributions over time. The average network properties and the distributions are found to be relatively stable, while the ranking of the individuals change significantly. Our work differs from the university affiliation/contact network in the terms of time scale(yearly data vs. daily data) and the nature of the contact between the individuals, but the results are nevertheless noteworthy.

Krings et al. [13] studied the effects of time window size and placement on the network structure using mobile phone data. While the musician network has significant differences from the phone call network, there is a similarity in the methods used to analyze the network. In order to study both the evolution of the network, and the effects of the past nodes' exclusion, it was necessary to aggregate the network in different time windows. As a result, we employ similar tools and techniques to analyze the network properties, with different goals.

Our work deals with a community of musicians, with the data obtained from the web. The general properties of this kind of network are expected have similarities to the citation and scientific collaboration networks, as music is a collaborative process. We take the dynamic properties of the network into account and study the changes in the network properties over time, as well as those in the individual nodes. By doing this, we aim to explain how the network is formed, how the central players acquire their position and how the structure of the network changes over time.

This report's structure is as follows: first, we give information about the data obtained and the methodology that has been followed to analyze the data. After that, we show how certain properties of the network change over time, marking the changes in the structure and comparing it with other studies. We then focus on the changes in the individual nodes in the network using centrality measures. Lastly, we show the results of the case study done on best-selling bands and albums.

II. DATA

A. Overview

The data is obtained from 'BandToBand.com' [18] on April 8, 2012 in HTML format. The website aims to map out a family tree of the bands related to Rock'n Roll music. Despite the emphasis on rock music in this report's title and in the website, it covers a variety of genres ranging from pop to jazz.

The data consists of structured information of bands, their members, the subset('line-ups') of the band members that have played together, and the albums produced by those members of the band. We also have the year for the formation of the line-ups, and the year of production of each album. In total, we downloaded 19,058 pages of bands.

The website conforms to a certain format, knowledge of which proved useful when constructing our networks:

1. Each musician belongs to a band. Solo artists have self-titled bands.
2. An album is produced by one or more line-ups within the band.
3. An album is produced by **one band only**. If two bands collaborate on one piece of work, a new band is created in the website containing the collaborating artists.

4. There is a path between every band in the website, meaning that the full network consists of the giant connected component only.

5. The temporal properties of the data is given in years. This determines the precision level of the temporal analysis. We can track changes year-by-year, but cannot track the monthly or daily changes within a year.

B. Limitations

As in many works on social networks, we are dealing with incomplete data. The network of bands given in the website is only a subset of the real life network, and we don't know to which extent it does capture the real life network. As the website's content is maintained by a community, the data is inherently biased towards the known bands.

The selection of bands is also biased in a way that constructs a giant connected component. That means, the authors have to dismiss all new bands that have no connection to the existing bands in the network. This could be an explanation to the unexpected decline in the number of nodes and edges created in the last decade, which can be seen in Figure 1.

In addition to the biased selection of the bands by the authors, our initial web crawling missed 10 web pages. We also found exceptions to 'giant connected component' rule: two bands were disconnected because the bands that are supposed to 'bridge' them to the main network were missing. One of those bridge bands were among the bands we failed to download. The other one was not added to the website at the time of web crawling.

Finally, it should be noted that the only activity considered by our data is album production. The year attribute for bands and artists are defined as the time they started to work on albums, not the actual start of their music career.

III. METHODOLOGY

A. Accessing the Data

Pattern [19] and *lxml* [20] libraries for Python were used to crawl the website and parse the HTML files, respectively. The extracted information was saved into a relational database. Our database consists of tables containing the information on bands, artists, albums, line-ups, and association tables('lineup-album','lineup-artist','album-artist'). The latter helped us to determine the edges when constructing our networks.

B. Network Construction

We created three different networks to capture the different levels of interaction: band, album and artist. Studying the network of bands gives us information about bands sharing members, whereas studying the artist network allows us to study the collaborations on a more personal level. The connections in the artist network are reflected in the band network, as producing an album together requires playing in the same band according to the website's rules. The network of albums is more abstract, as albums are not interacting agents like bands and artists, but the products of the collaboration. The definitions for the networks are the following:

Band: The bands are connected to each other by the musicians that played in both bands. Therefore, the bands are represented by nodes and the common artists are represented by edges.

Artist: In our network, artists form connections by producing an album together. The nodes represent the artists and the edges represents common albums.

Album: Albums are connected by the common artists that produced them. As a result, albums are represented by nodes, and common artists are represented by edges in the network.

Using these definitions, we constructed three multi-graphs, which allow multiple edges between nodes.

C. Dynamic Properties

Each node and edge has a 'year' attribute. Once the nodes and edges are created, they stay in the network until the end. The year of the nodes are determined as follows:

Band nodes: The year of the earliest line-up the band contains.

Artist nodes: The year of the earliest line-up the artist is a member of.

Album nodes: The year of the album's production.

The definition for edges are slightly more complicated:

Band edges (b_1, b_2): Let b_1 and b_2 two bands the artist a has been a member of. Let t_{b_1} the year of the earliest line-up of band b_1 common artist a has been in, and let t_{b_2} be the same for the band b_2. The 'year' attribute of the edge is then defined as follows:

$$t_{b1,b2} = max(t_{b_1}, t_{b_2})$$

In other words, it is the year of the artist that played in the band b_1 joining the band b_2.

Artist edges (a_1, a_2): Let a_1 and a_2 be the nodes denoting two artists that worked together on the album l. Let t_l denote the year of the album's production. The year attribute of the edge (a_1, a_2) is equal to the album's production year: $t_{b1,b2} = t_l$.

Album edges (l_1, l_2): Let l_1 and l_2 be the nodes denoting two albums produced by the artist a. Let t_{l1} and t_{l2} denote the year of the l_1 and l_2's production, respectively. The year attribute of the edge (l_1, l_2) is defined as:

$$t_{l1,l2} = max(t_{l_1}, t_{l_2})$$

In other words, the edge year is equal to the production year of the newer album of the two.

These definitions ensure that we do not end up with edges connected to the nodes that are yet to be created.

D. Network Aggregation

Multi-graphs are convenient for storing the data in our case, since new edges can be added between two nodes over time. But most of the graph algorithms are designed to work on single-link graphs. Therefore, when aggregating our network, we project it onto a weighted single-link graph. The edge weight between two nodes is equal to the number of edges between these nodes in the multi-graph. Consequently, the edge weights in our graphs represent the strength of the connection between two nodes.

We define two values, t_{start} and Δt, for aggregation. They denote the starting year of the aggregation, and the time window size in years, respectively. For convenience we also define the ending year as $t_{end} = t_{start} + \Delta t$. Due to the low number of nodes prior to 1950(Figure 1), we only aggregate in the time windows of $t_{start} \geq 1950$ during network analysis. We also use t_0 synonymously with t_{start} in some of the plots.

When aggregating, we take only the nodes $v \in V$ $v_{year} \in [t_{start}, t_{end}]$. For edges, we take the pairs $e = (v_1, v_2)$ with $(v_1, v_2 \in V)$ and $e_{year} \in [t_{start}, t_{end}]$. It should be noted that we have an extra constraint for edges: we check both the year for both the edges and the nodes it connects. This constraint is necessary to prevent edges from referring to a node outside the given time window.

Keeping the value of t_{end} constant while decreasing Δt means excluding the nodes from the past $|\Delta t_2 - \Delta t_1|$ years. This allows us to see the exclusion of which nodes affect the network properties significantly. Using this information, we can determine the time periods where the structure of the network has changed.

Table I shows the number of nodes and edges for the whole network. The album network has a considerably higher density than the other two networks. The relative compactness of the band network allows for more detailed analysis with different algorithms.

	Number of nodes	Number of edges
Band	19058	81341
Artist	48078	186574
Album	44470	546286

TABLE I: Basic statistics for the networks.

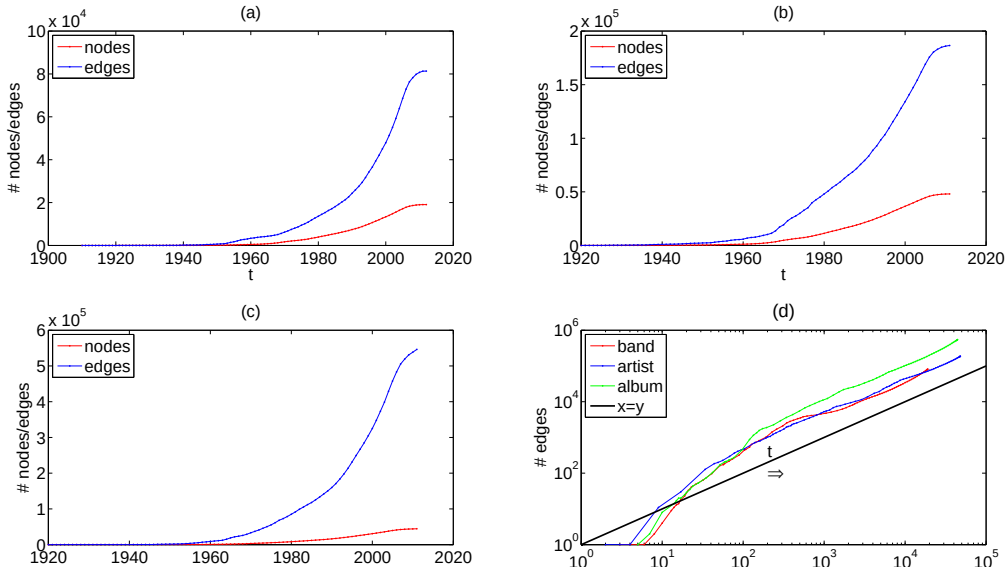

FIG. 1: The number of nodes and edges in the full network when we aggregate from the earliest $t_{start} = 1910$. (a) Band network (b) Artist network (c) Album network (d)Nodes with relation to edges as they grow in time.

E. Network Analysis

We used *NetworkX* [21] library for Python was to analyze the network. The majority of the network analysis was done with the values of t_{start} between 1950 and 2000 and Δt between 0 and 50. It should be noted that the combination of networks and time windows means a large number of plots for each time-varying attribute. Therefore, we only included the most relevant ones in the main report for the convenience of the reader. The appendix contains all results for completeness and interested readers.

IV. NETWORK PROPERTIES

A. Connected Components

It was mentioned earlier that there exists a path between all bands in the network. This holds true for the whole network when its dynamic properties are ignored. When we start aggregating from 1950 to 2012, however, it appears that the network initially consists of more than one connected component. The giant connected component(GCC) grows over time, eventually encompassing the whole network.

Figure 3 shows the proportional size of the largest connected components over time window. All three networks follow the same pattern, except the artist network near 1955, where it differs from bands and albums, and falls to its minimum. The drops in the proportions and their eventual rise to 100% describe the birth of independent bands and artists, which eventually become a part of the mainstream scene. It can be said that the bands are compact representations of artists, therefore the drops and rises in the proportion are reflected more strongly in the artist network. We can also see that all three networks are largely dominated by the GCC. The second largest connected component's size is comparable to the GCC in the beginning, but its proportion drops quickly over time.

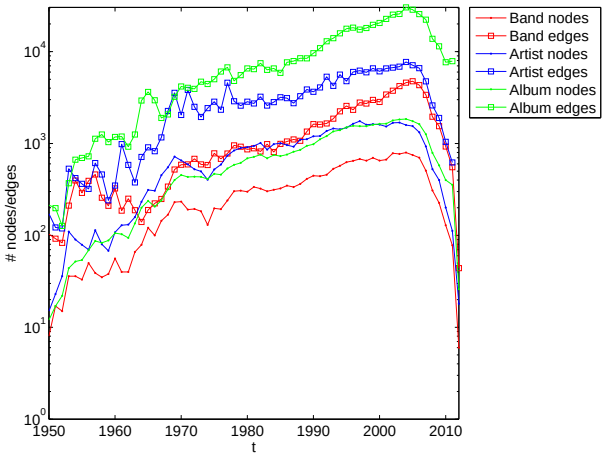

FIG. 2: The number of nodes and edges joining the networks per year.

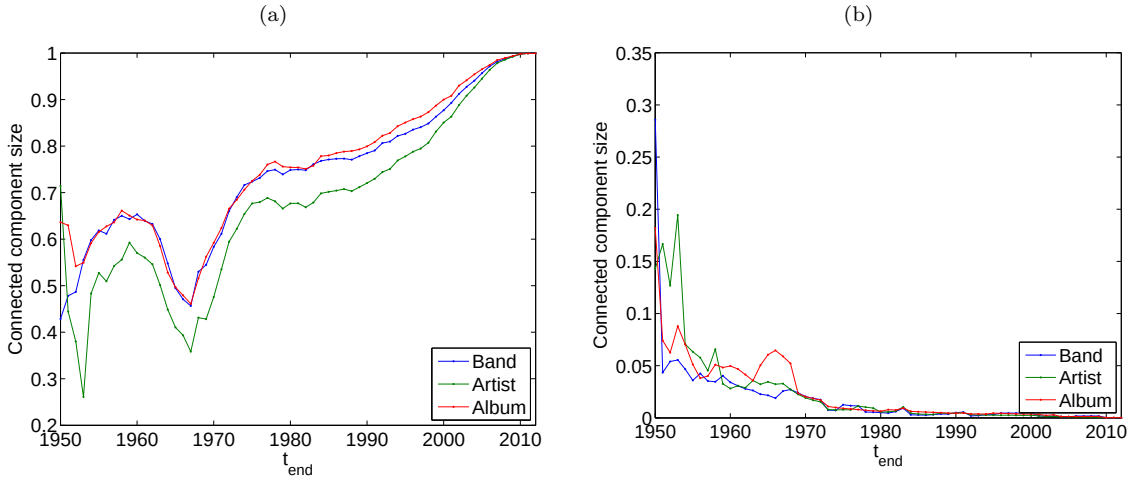

FIG. 3: The number of nodes in the largest connected components divided by the total number of nodes, aggregated from $t_{start} = 1950$. (a) Giant connected component (b) Second largest connected component.

B. Degree, Weight and Strength

Degree of a node is defined as the number of edges that connects it to the other nodes [1]. Degree distribution is one of the main indicators of a network's characteristics. It can reveal possible hierarchies and the existence of the 'rich get richer' phenomenon [22]. Figure 4a shows how different types of interactions lead to different degree distributions. Figure 4b shows degree distributions from different time window positions for comparison.

Bands: With a t_{start} of 1950, the distributions are nearly overlapping: they converge towards the same distribution as the window size is increased. But the removal of the nodes between 1950 and 1960 causes the network structure to change: with t_{start} values of 1960 onwards the distributions start spreading out as the time window size is increased. The distributions are monotonically declining.

Artists: The distributions overlap for the values of $t_{start} \in [1950, 1970]$ The distributions start becoming more broad after $t_{start} = 1980$, similar to the distributions in the bands network. The peak of the degree distribution is between 4-7 in various time windows, unlike the distributions for the artist and album networks.

Albums: When $t_{start} = 1950$, the distributions overlap until k=100, and the values differ after that point. As t_{start} is increased, the distributions become more broad, with less number of overlaps. The distributions are monotonically declining.

Our results can also be verified in the Appendix section A 1, where we included the distributions for

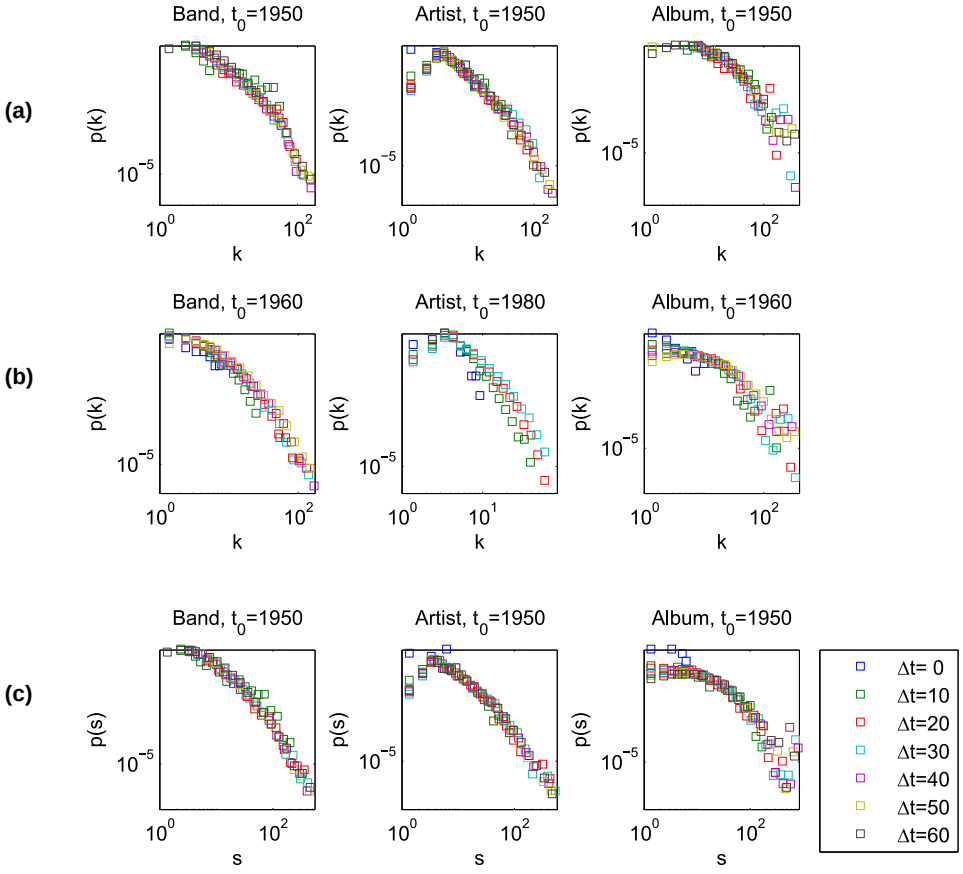

FIG. 4: Degree and strength distributions in various time windows, log binned. The legend in (c) is valid for all of the plots in the figure. (a) Degree distributions with $t_{start} = 1950$. (b) Degree distributions with t_{start} of 1960, 1980 and 1960; for band, artist and album networks, respectively. (c) Strength distributions with $t_{start} = 1950$.

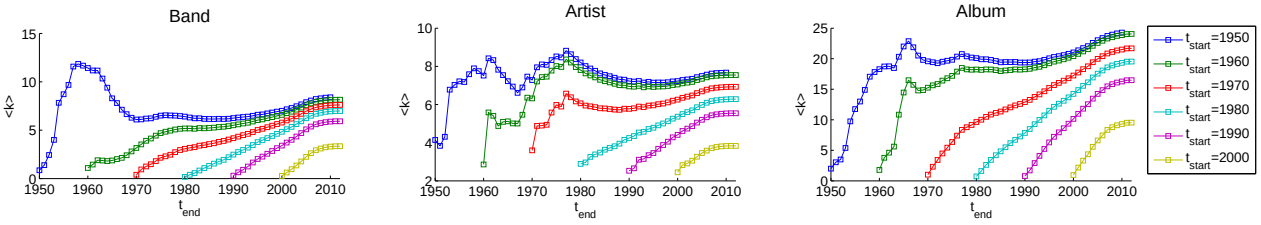

FIG. 5: Average degree in the networks compared to the time window. Colors denote different values of t_{start} , starting position of the aggregation window.

all values of t_{start}.

Figure 5 shows the changes in the average degree. In the band network, the peak of the graph is in the period between 1950-1960. There is an unusually high number of connections between bands in that period of time, and it gets balanced out as the network grows. Removal of the nodes in 1950-1960 causes the average degree to drop in all three networks, but the effect is not as strong as in the band network.

The interpretation of the results gives us insight on the change in network structure: the peak of average degree at $(t_{start} = 1950, t_{end} = 1960)$ in the band network marks a period of unusually high level of collaboration. There were more connections formed between the bands than between individual artists: a considerable number of artists in this period played with the same people under the name of different bands. Table II shows the 10 bands with the highest degree in the period. Interestingly, the name of Miles Davis appears three times in the list. All of the listed bands are jazz groups.

Rank	Degree	Band
1	63	Sonny Rollins Quartet
2	62	Horace Silver Quintet
3	61	Miles Davis Sextet
4	58	Lou Donaldson Quintet
5	53	Miles Davis And The Modern Jazz Giants
6	52	Thelonious Monk Septet
7	51	Jackie Mclean Quintet
8	50	Kenny Dorham Quintet
9	50	Kenny Drew Trio
10	50	Miles Davis Quartet

TABLE II: Bands with the highest degree when $t_{start} = 1950$ and $t_{end} = 1960$.

The degree and band networks have monotonically declining degree distributions. The majority of the bands in the network produced few number of albums and share few number of artists with other bands.

Due to the collaborative nature of music, the artists have a tendency to keep more connections. As a result, the degree distributions have a peak between 4-7, rather than a monotonic decline as in the band and album networks. The solo artists are outnumbered by the ones playing in bands.

The strength of a node is defined as the sum of the weights of the edges connected to it [23]:

$$s(i) = \sum_j w(i,j) \quad (j \in V_i)$$

where w(i,j) is the weight of the edge between the nodes i and j, and V_i is the set of node i's neighbors.

In our case, it is also what the degree of the nodes would be had we kept the initial multi-graph model, due to the definition of weights in Section III-D. For completeness, we included a comparison of degree and strength distributions in Figure 4c. The figures imply that the values of degree and strength share the same underlying distribution. The distributions retain their similarity with other values of t_{start} (see Appendix A 1). Since the minimum edge weight is 1, the strength of a node is always equal or greater than its degree. As a result, the strength distributions have a longer tail than the degree distributions.

C. Distance

In order to study the network structure in more detail, we investigate the distance measures in addition to the degree and strength. In graphs, the distance between two nodes is defined as the number of edges in the shortest path that connect them together. Analysis of distance related measures yield valuable information about the network, such as its 'small world' properties and its layout.

The diameter of a network is the largest distance between any two nodes in it. We used diameter and the average shortest path length as the distance-based measures. Since the network is not fully connected in the most time windows, we only considered the GCC. Figures 6 and 7 show the change in diameter and average distance in different time window sizes and positions. There is a strong correlation between the diameter and the average distance (see Figure 8).

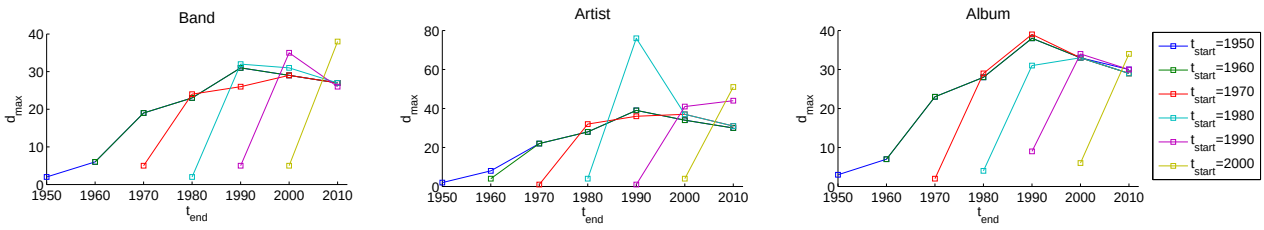

FIG. 6: The change in the diameter with various time window size and positions.

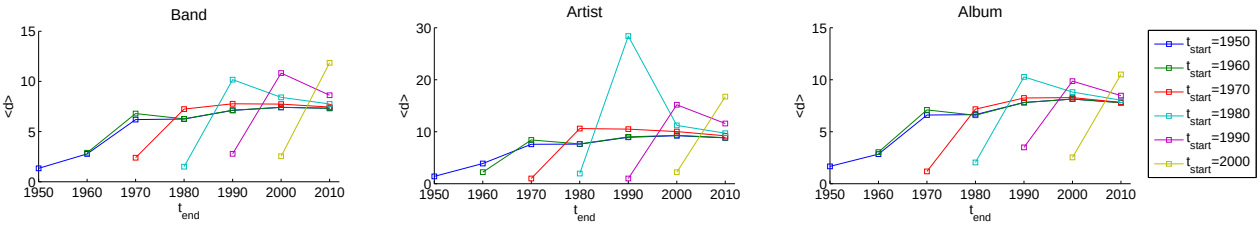

FIG. 7: Average shortest path length in the networks with various time window size and positions.

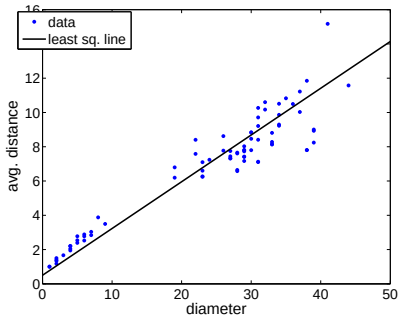

FIG. 8: Diameter values compared to the average distance, for all networks and time windows.

When we only consider the diameter with $t_{start} = 1950$, we can see that it reaches its peak in 1990 in all three networks, then has a slight decrease in the next 20 years. A study by Leskovec et al. finds shrinking diameters over time in various networks [15]. In the case of musician networks, the fall is not significant enough to claim that the diameter is shrinking. However, the peak in 1990 and the fall after that time is significant, as they show that there is a sufficient number of shortcuts formed in the network to counteract the effect of the network's growth on the diameter.

When we consider different values of t_{start}, we observe that the diameter and average distance is considerably low when the $\Delta t = 0$ and ($t_{start} = t_{end}$). When the time window is minimal, the aggregated network is rather small (see Figure 2), and the size of the GCC is even smaller due to the lack of edges.

In Figure 6, the plots with different values of t_{start} tend have closer values to each other and overlap at some points, while the average distance plots show less inclination to overlap. Removal of the nodes that lie before t_{start} obviously affects the distances, but the distance between the peripheral nodes of the network is less affected by that.

By checking the values of different plots in the same value of t_{end}, we can see how the removal of the past nodes affect the network structure. The 'jump' in diameter and average distance in $t_{start} = 1980$ and $t_{end} = 1990$ is noteworthy. While the most drastic increase is in the artist network, it is also noticable in the band and album networks. This implies the emergence of central individuals between 1970-1980, serving as 'hubs' for the individuals that arrived to the scene in the next decade. Removal of these shortcuts decreases the small world effect in the network. We put this conjecture to the test in the following section as we investigate the node centrality values.

D. Centrality

Analysis of centrality values help us determine a node's position in the network and reveal the influential nodes and hubs acting as shortcuts. There are various methods to determine how central a node is, based on different interpretations of what centrality means.

We considered the following measures of centrality in this work:

Closeness: The closeness centrality of a node is the multiplicative inverse of its average distance. While it is a commonly used and intuitive measure of a node's position in the network, it has certain disadvantages: the values tends to be 'cramped up' due to the relatively small size of shortest path lengths in the network. As a result, central nodes and the less important nodes are not separated well enough,

and the ranking is sensitive to slight changes in the structure.

Betweenness: The betweenness of a node is the number of shortest paths passing through it. Unlike closeness centrality, the values are well-separated. We normalized the betweenness by dividing it by the number of all shortest paths in the network, to make the values comparable among different time window sizes.

k-core: A k-core of a network is its largest sub-network that contains only the nodes with the degree k or more [24]. By determining k-cores, we can assign a core number to a node as the maximum k of a k-core that contains it. Having a large core number not only requires having a high degree, but also neighbors with a high degree as well.

We only consider the giant connected component for the measures involving shortest paths.

Figure 9 shows the normalized betweennness distributions, right-skewed as seen in previous studies [1], and with a decreasing slope as the time window size is increased. The normalization causes the individual betweenness values to decrease quadratically as the number of nodes is increased.

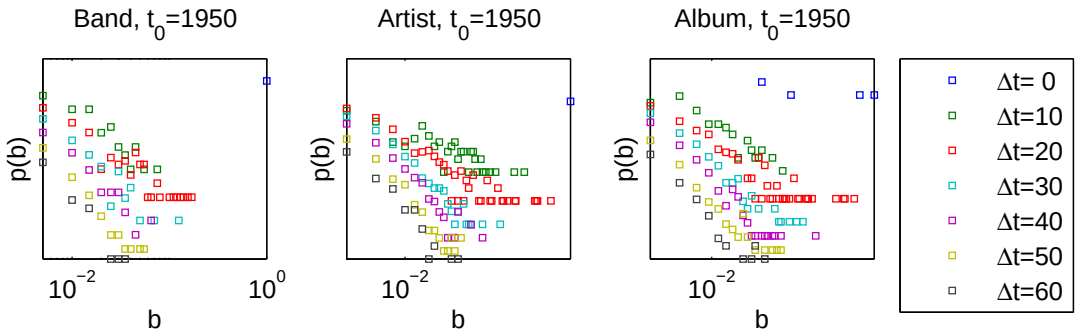

FIG. 9: Betweenness distribution for band, artist and album networks

In earlier sections, we revealed the emergence of central individuals in certain periods with high betweenness values, and that the removal of those nodes causes the network structure to change. This suggests that the network has a layered structure, with one generation growing around certain central nodes in the previous generation. In order to test this, we investigate the distribution of betweenness values of the nodes by their year of joining. Figure 10 a and b shows the distribution of betweenness values in the networks aggregated between 1950-2010, for bands and artists. In both networks, the maximum value of betweenness peaks in a certain year (1987 and 1974, respectively) and starts decreasing.

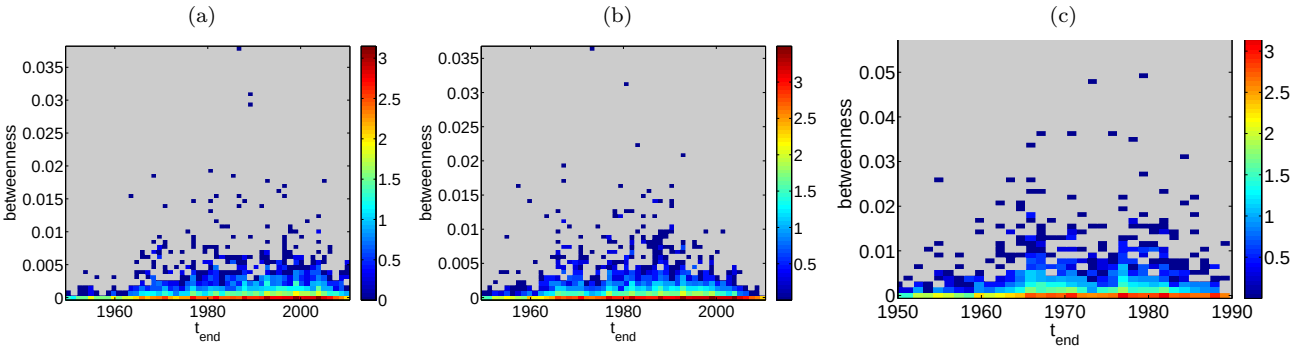

FIG. 10: The density of the betweenness values in the network aggregated between 1950 and various values of t_{end}. The y axis denotes the betweenness score, the x axis denotes the year the node joined the network, and the color denotes the number of nodes that fall into a particular bucket, scaled to log 10. (a) Band network $t_{end} = 2010$, (b) Artist network $t_{end} = 2010$, (c) Artist network $t_{end} = 1990$

The position of the peaks is noteworthy: while there are individuals with high betweenness from the early years of the network, the most central nodes are in the late 80s (bands) and the 70s (artists). This implies that the central positions were taken over by the nodes in the newer generations that start in the 70s and the 80s. Figure 11 shows the ranks of the artists with the highest betweenness values

in $t_{end} = 1960$ and $t_{end} = 2010$, with $t_{start} = 1950$ in both cases. The figure shows the process of the older central players falling out of the ranking, the new central nodes entering the giant component and eventually acquiring the top positions. In the band network, all central players joined the giant component in $t_{start} = 1990$. The artist network, however, has central nodes that date back as far as to $t_{start} = 1970$.

Figure 10c shows the same distribution for bands between 1950-1990, in order to explain the jump in diameter mentioned in section IV C. The results confirm the emergence of central individuals with high betweenness values between 1970 and 1980, which are essential to the network structure in $t_{end} = 1990$. It also explains the jump in diameter, as the removal of individuals with high betweenness is analogous to removing the shortest paths from the network.

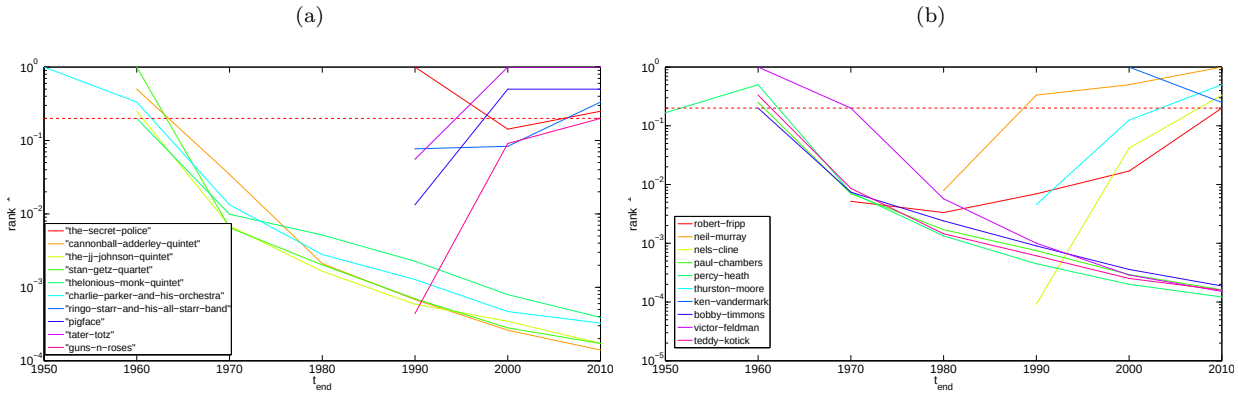

FIG. 11: The betweenness centrality rank for the top 5 nodes in 1960 and 2010, with the network aggregated from $t_{start} = 1950$ and changing values of t_{end}. The top 5 positions are above the dashed line. (a) Band network (b) Artist network

By comparing the change in k-core versus the change in average distance, we can get a better insight on how a node's position in the network evolves over time. In order to do that, we consider the values of these two attributes from $t_{start} = 1950$ with $\Delta t = 0, 10..., 60$. We note the change in the value each time we increase the time window size, and record these changes as vectors of $(kcore, |d|)$ from the previous point to the next one. After that, we divide up the area into a rectangular lattice, and relocate the starting position of each vector to the nearest point in the lattice. By averaging over all vectors directed from a point in the lattice we created a mean vector field.

The results are shown in figure 12. The nodes that appear earlier in the network start with a lower average distance, and the distance value tends to rise up over the years. The k-core values either go up or stay the same, since we never remove any nodes or edges from the network. A notable difference between band and artist networks is that the k-core values in the artist network have less inclination to change, whereas the k-core of the bands have a tendency to increase. This can be explained by the way bands and artists form their connections: the connections formed by a band represents the collective actions of all its members and former members, while the connections formed by the artists reflect the individual efforts. Since artists tend to play with the same people throughout their career and make very few new connections.

Investigating centrality measures allows examination of the changes in micro scale. This gave us the possibility to explain and verify the results we obtained about the network structure.

V. CASE STUDY

We collected the list for the best-selling bands and artists from RIAA's website [25]. We selected the top ten artists that exist in our network and tracked their attributes over time as we increased the aggregation window size. The results are shown in Table III.

The solo artists tend to be in the periphery of the network, regardless of their success. Michael Jackson and Elvis Presley, for example, are not even part of the giant connected component until the last century. Billy Joel and Elton John, while a part of the giant component from the start, keep a betweenness value of 0 and a degree of 2. We can safely conclude that collaborating with other artists on albums is not a

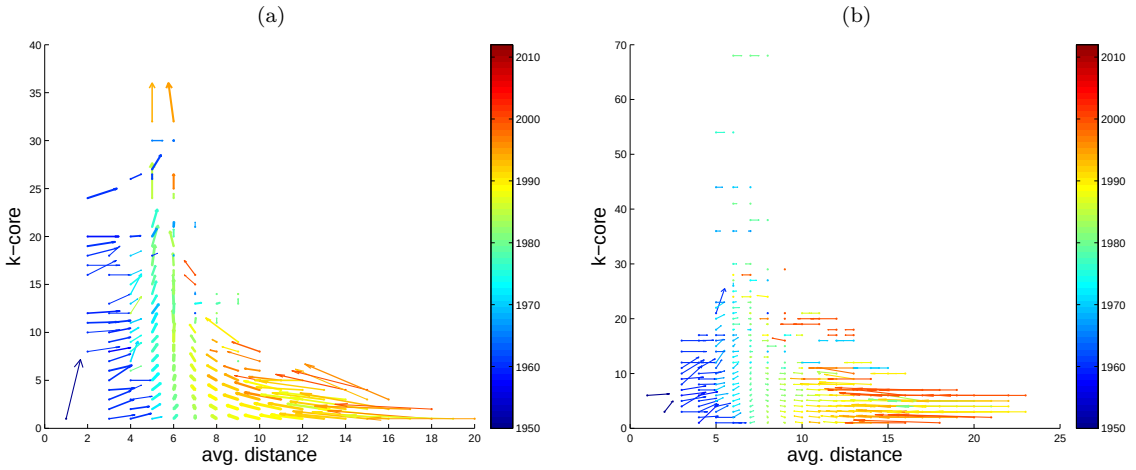

FIG. 12: The change in k-core number of a node with comparison to the change in the average distance. The color denotes the average creation year of the nodes in the area. (a) Band network, (b) Artist network

prequisite for a solo artist to become a star. Although, it should be noted that there exists other mediums of collaboration which are not captured by our network, such as TV or concerts. Lack of collaboration on albums must not be mistaken with lack of collaboration in other ways.

Bands, on the other hand, can acquire relatively more central positions. The betweenness values of Led Zeppelin in the last two decades are a good example of that. The band's betweenness rank jumps from 4339 to 1065, after a consistently decline until 2000. Conversely, bands can also start in central positions and lose their position gradually. Aerosmith's betweenness starts at the rank 683 and eventually falls down to 3637.

(a)

		t_{end}					
		1960	1970	1980	1990	2000	2010
1. The Beatles	degree	1 (278.5)	17 (139.5)	20 (255.5)	22 (380.5)	25 (631)	27 (945)
	strength	3 (222)	38 (84)	50 (139.5)	53 (201.5)	76 (191)	80 (278)
	closeness cent.	-	0.1729 (445)	0.197 (513.5)	0.1862 (413)	0.1743 (804)	0.1771 (861)
	betweenness	-	0.0027 (291)	0.0011 (865)	0.002 (756)	0.0014 (1130)	0.0006 (2358)
2. Elvis Presley	degree	2 (232.5)	2 (895)	2 (2473)	2 (4877)	2 (9834)	2 (15707)
	strength	2 (250.5)	2 (1023.5)	2 (2721)	2 (5390)	2 (10638)	2 (16524)
	closeness cent.	-	-	-	-	-	0.085 (18614.5)
	betweenness	-	-	-	-	-	0 (15617)
3. Led Zeppelin	degree	-	4 (572)	4 (1645.5)	8 (1590.5)	9 (2849.5)	11 (4038.5)
	strength	-	5 (662.5)	5 (1812.5)	11 (1809)	12 (3422)	14 (5027)
	closeness cent.	-	0.1729 (444)	0.1806 (1070)	0.1745 (1033)	0.165 (1885)	0.1763 (941)
	betweenness	-	0.0109 (107)	0.0018 (632)	0.0004 (2002)	0.0002 (4339)	0.0013 (1065)
4. Eagles	degree	-	-	14 (440.5)	20 (459)	21 (840.5)	23 (1256.5)
	strength	-	-	34 (244.5)	41 (316)	44 (539.5)	59 (532.5)
	closeness cent.	-	-	0.1759 (1262)	0.1747 (1020)	0.1634 (2174.5)	0.1678 (2314)
	betweenness	-	-	0.005 (262)	0.0068 (175)	0.0032 (381)	0.0028 (388)
5. Billy Joel	degree	-	-	2 (2473)	2 (4877)	2 (9834)	2 (15707)
	strength	-	-	2 (2721)	2 (5390)	2 (10638)	2 (16524)
	closeness cent.	-	-	0.099 (2750.5)	0.0934 (5466.5)	0.0937 (11151.5)	0.0996 (17949.5)
	betweenness	-	-	0 (2280)	0 (4629)	0 (9504.5)	0 (15617)

(b)

		t_{end}					
		1960	1970	1980	1990	2000	2010
6. Pink Floyd	degree	-	6 (391)	6 (1145.5)	9 (1386.5)	11 (2193.5)	11 (4038.5)
	strength	-	6 (576)	6 (1633.5)	23 (781.5)	25 (1333.5)	25 (2351.5)
	closeness cent.	-	-	-	0.1651 (1755.5)	0.1621 (2366)	0.1584 (4605)
	betweenness	-	-	-	0.0024 (606)	0.0013 (1168)	0.0008 (1832)
7. Elton John	degree	-	1 (1174.5)	2 (2473)	2 (4877)	2 (9834)	2 (15707)
	strength	-	1 (1221)	2 (2721)	2 (5390)	2 (10638)	2 (16524)
	closeness cent.	-	0.1161 (843.5)	0.1663 (1569)	0.1534 (2623)	0.145 (5220)	0.1408 (9705)
	betweenness	-	0 (735.5)	0 (2280)	0 (4629)	0 (9504.5)	0 (15617)
8. AC/DC	degree	-	-	9 (773.5)	19 (497.5)	21 (840.5)	24 (1170)
	strength	-	-	15 (745)	31 (482.5)	38 (690.5)	43 (954.5)
	closeness cent.	-	-	0.1619 (1716)	0.1859 (427)	0.1726 (954)	0.1719 (1569.5)
	betweenness	-	-	0.0055 (243)	0.005 (266)	0.0024 (575)	0.0018 (708)
9. Michael Jackson	degree	-	-	2 (2473)	2 (4877)	2 (9834)	2 (15707)
	strength	-	-	2 (2721)	3 (4655)	3 (9371.5)	3 (14804.5)
	closeness cent.	-	-	-	-	-	0.102 (17783.5)
	betweenness	-	-	-	-	-	0 (15617)
10. Aerosmith	degree	-	-	3 (2001.5)	6 (2151)	7 (3847.5)	8 (5989)
	strength	-	-	4 (2046)	15 (1316.5)	16 (2428)	17 (3982.5)
	closeness cent.	-	-	-	0.1525 (2686)	0.1456 (5101)	0.145 (8458)
	betweenness	-	-	-	0.0021 (683)	0.0006 (2207)	0.0004 (3637)

TABLE III: The position of the best selling bands and solo artists (6-10) in the network over time, aggregated from t_{start}=1950. The numbers in parantheses denote the ranking among the other nodes in the network. In the case multiple nodes with the same value, the average rank for the value is given. If a node has an existing degree and strength, but the betweenness and closeness values are marked as '-', it is not a part to the giant component at that time.

VI. DISCUSSION

A. Results

We demonstrated the formation of the giant connected component over time. We showed the change in the network structure by aggregating the network in different time windows and studying the measures of degree, strength and distance. We studied the changes in centrality of the nodes in a micro scale and revealed the movements of the nodes in the network over the years. We also investigated the distribution of the central individuals by generations and revealed how the central positions in the network are taken over in time. A case study of the successful bands and artists showed that the success in real life does not appear to be correlated with the position in the network.

Our methods show the importance of temporal analysis of social networks. The static network is presented as a single giant connected component. However, investigating its size over time reveal its formation as a gradual process. By investigating this formation, we also found the time periods where many independent bands without prior connections join the network, such as the late 60s.

Investigating the degree and strength is essential to understanding the periods with high levels of collaboration. We found the period between 1950-1960 to be one of these cases for bands, as seen from the peak in the average degree. The high level of interconnectedness is not reflected as strongly in the artists network, suggesting that same people played under the name of different bands (which seems to be the case with many jazz musicians that belong to our network in that period). By changing the time window size and removing the nodes that are left outside the time window, we showed their importance in the formation of the network structure.

After studying the degrees, we analyzed the changes in the distance related measures, which helped us examine the small world effects in the network and obtain further information about the network structure. We found that removing the nodes that first appear between 1970-1980 in the artist network causes the diameter to increase significantly. We suggested that the increase with the removal of central players in the artist network, which have a high betweenness value.

Analysis of centrality measures supported our conjecture about removal of central players. Moreover, we found that the central positions in the networks were taken over by bands and artists from new generations in 1970s and 1980s. We also investigated the changes in k-core and average distance values. The newly added peripheral nodes start with a high average distance value, which decreases over time. In constrast, the old nodes start out with a low average distance (since the network is small in its early days) which increases over time as new nodes are added to the network. The average distance values approach to a stable value from both sides. We also found that the k-core value of the artists is less likely to change, meaning that artists tend to play with the same people throughout their career. In contrast, it is easier for bands to build new connections over time, as the connections made by bands reflect the collective actions of its individual artists.

We also showed the importance of using different time window positions in aggregated networks. Shifting the value of t_{start} while keeping t_{end} constant allowed us to discover critical periods where the network structure changes.

By studying all the time-variant properties of the network, we revealed important details about a social network's formation and structure, which would be impossible to obtain by static network analysis. Finally, and most importantly, we demonstrated that the community of musicians is an evolving, constantly changing entity.

B. Future work

The networks we analyzed contain valuable information about the relationship between musicians and bands. We uncovered some interesting information; nevertheless, there is room for future work on the data. Combined with external data, for example, one could investigate possible correlations between the community structure and the attributes such as geographic location, nationality, race, generation and genre. In addition, different types of networks could be constructed, e.g. a bipartite network of bands and artists. The methods in this work could also be used in other networks, such as citations, scientific collaboration and web communities.

The edges in the network we created are permanent, two nodes never disconnect once an edge has been formed. Assigning a 'life-span' to edges, or causing the edge weight to 'decay' over time with lack of activity could bring out interesting results. Furthermore, our choice of t_{start} and Δt were limited by

time and computational capabilities. Using more frequent intervals of t_{start} and Δt could allow a more detailed study of network evolution.

An examination of the results from a historical and sociological perspective would also lead to new possibilities. The findings such as influx of independent bands in 60s, the periods of high connectivity in the band network, the emergence of central individuals would be better verified and explained with a detailed knowledge of rock'n roll geneology.

VII. ACKNOWLEDGEMENTS

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

The author is thankful to the staff and students of Warwick Complexity Centre for the constructive feedback he received during course of his work on this project.

---

[1] M. Newman. *Networks: an introduction*. Oxford University Press, Inc., 2010.
[2] J. L. Moreno. Who shall survive? *A Journal of Studies Towards the Integration of the Social Sciences*, 5(4), 1952.
[3] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.
[4] D.J. Watts and S.H. Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998.
[5] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
[6] P. Gleiser and L. Danon. Community structure in jazz. *Arxiv preprint cond-mat/0307434*, 2003.
[7] R.D. Smith. The network of collaboration among rappers and its community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2006:P02006, 2006.
[8] C. Lee and P. Cunningham. The geographic flow of music. *Arxiv preprint arXiv:1204.2677*, 2012.
[9] R. Lambiotte and M. Ausloos. Uncovering collective listening habits and music genres in bipartite networks. *Physical Review E*, 72(6):066107, 2005.
[10] R. Lambiotte and M. Ausloos. On the genre-fication of music: a percolation approach. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1):183–188, 2006.
[11] F.N. Stokman. *Evolution of social networks*, volume 1. Routledge, 1997.
[12] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.F. Pinton, and W. Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.
[13] G. Krings, M. Karsai, S. Bernharsson, V.D. Blondel, and J. Saramäki. Effects of time window size and placement on the structure of aggregated networks. *Arxiv preprint arXiv:1202.1145*, 2012.
[14] S. Redner. Citation statistics from more than a century of physical review. *Arxiv preprint physics/0407137*, 2004.
[15] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM, 2005.
[16] J.S. Katz. Scale-independent bibliometric indicators. *Measurement: Interdisciplinary Research and Perspectives*, 3(1):24–28, 2005.
[17] G. Kossinets and D.J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.
[18] Bandtoband.com : Mapping the rock'n roll genome. `http://www.bandtoband.com/`, April 2012.
[19] T. De Smedt and W. Daelemans. Pattern version 2.3. `http://www.clips.ua.ac.be/`, April 2012.
[20] lxml - xml and html with python. `http://www.lxml.de/`, April 2012.
[21] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.
[22] A.L. Barabasi. Linked: The new science of networks. 2002. *Cambridge, MA: Perseus*, 2002.
[23] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747, 2004.
[24] S.B. Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.
[25] Top selling artists. `http://www.riaa.com/goldandplatinum.php?content_selector=top-selling-artists`, May 2012.

APPENDIX A: NETWORK DATA

1. Degree, strength and weight distributions

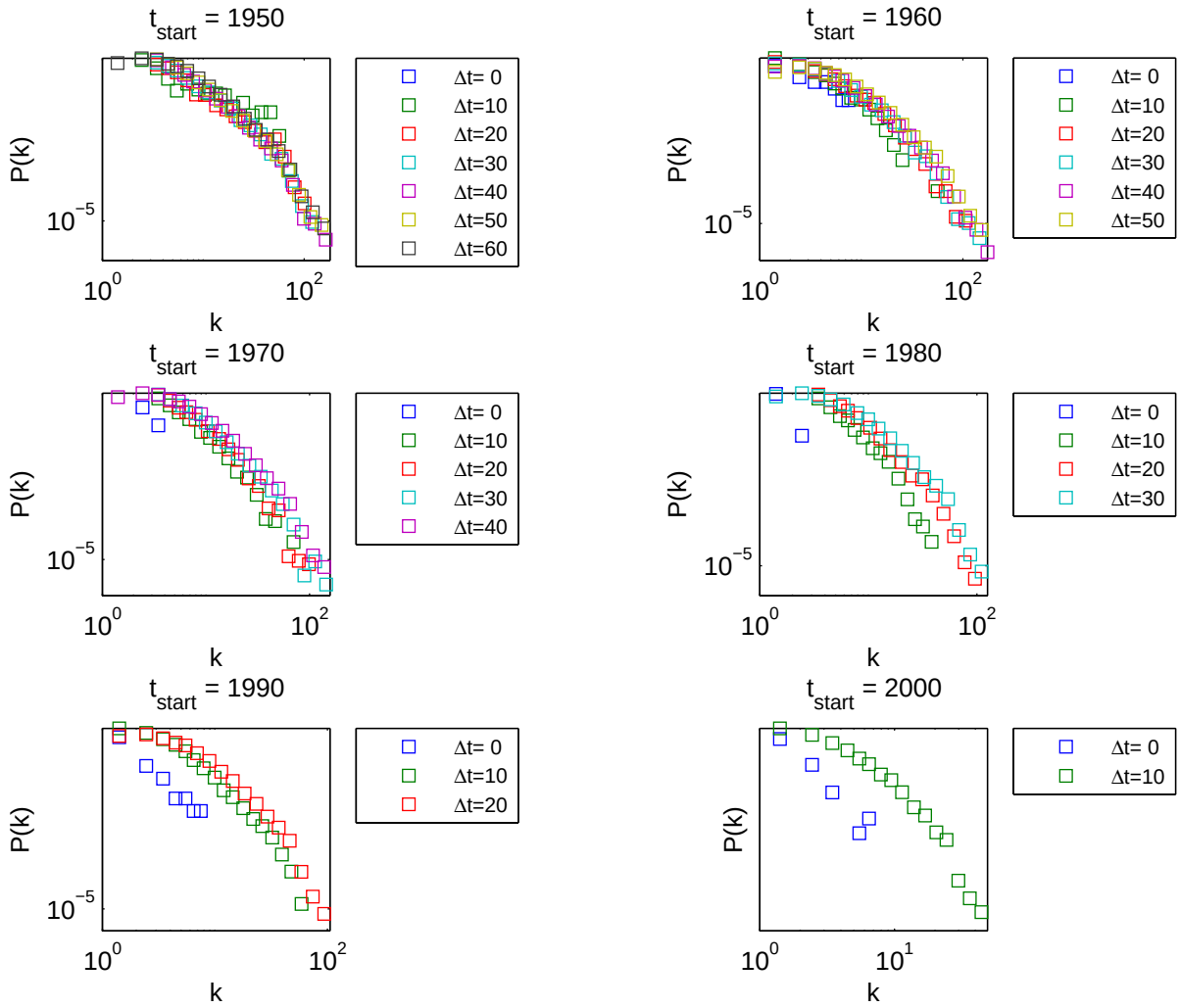

FIG. 13: Degree distribution for bands, log binned.

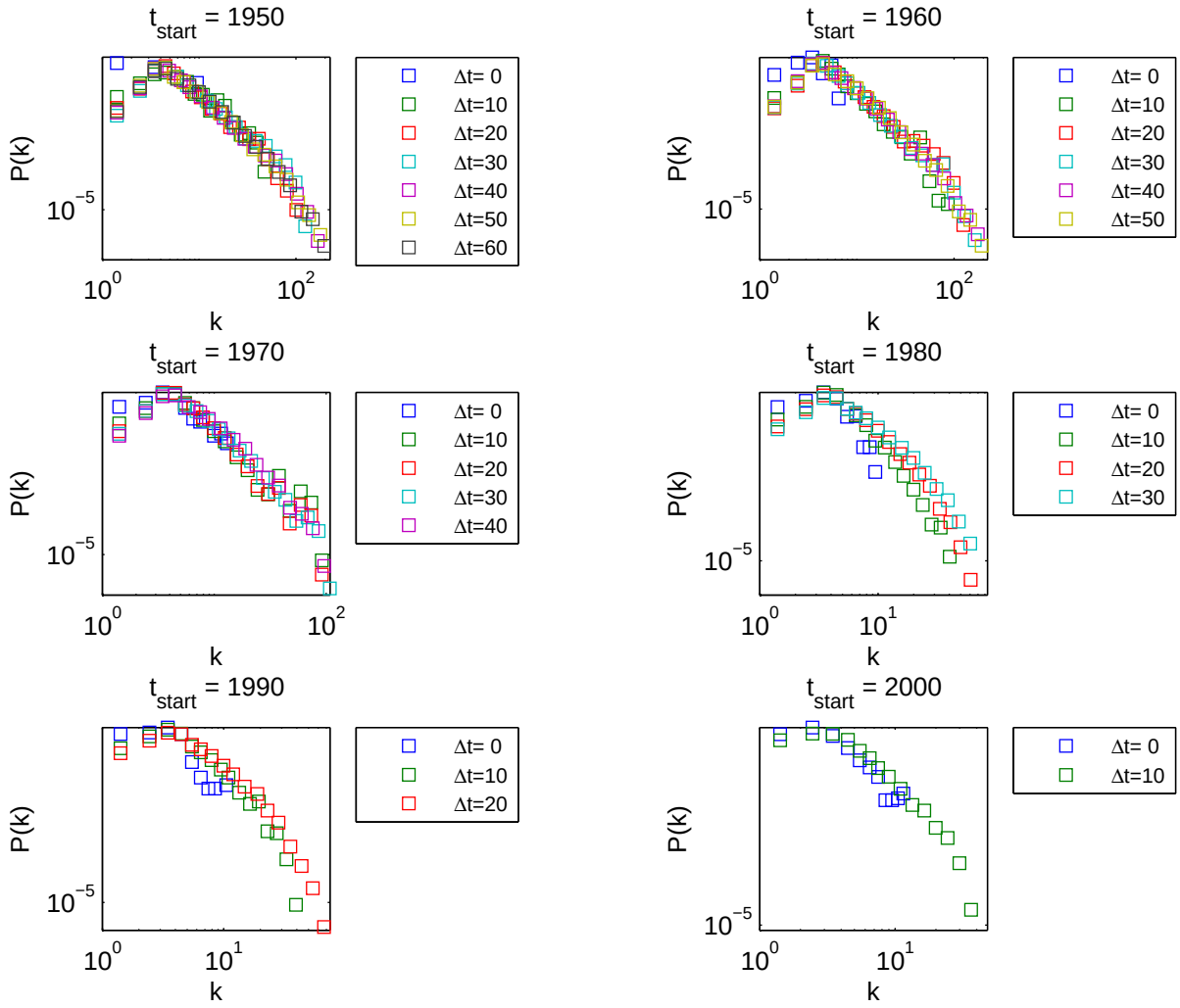

FIG. 14: Degree distribution for artists, log binned.

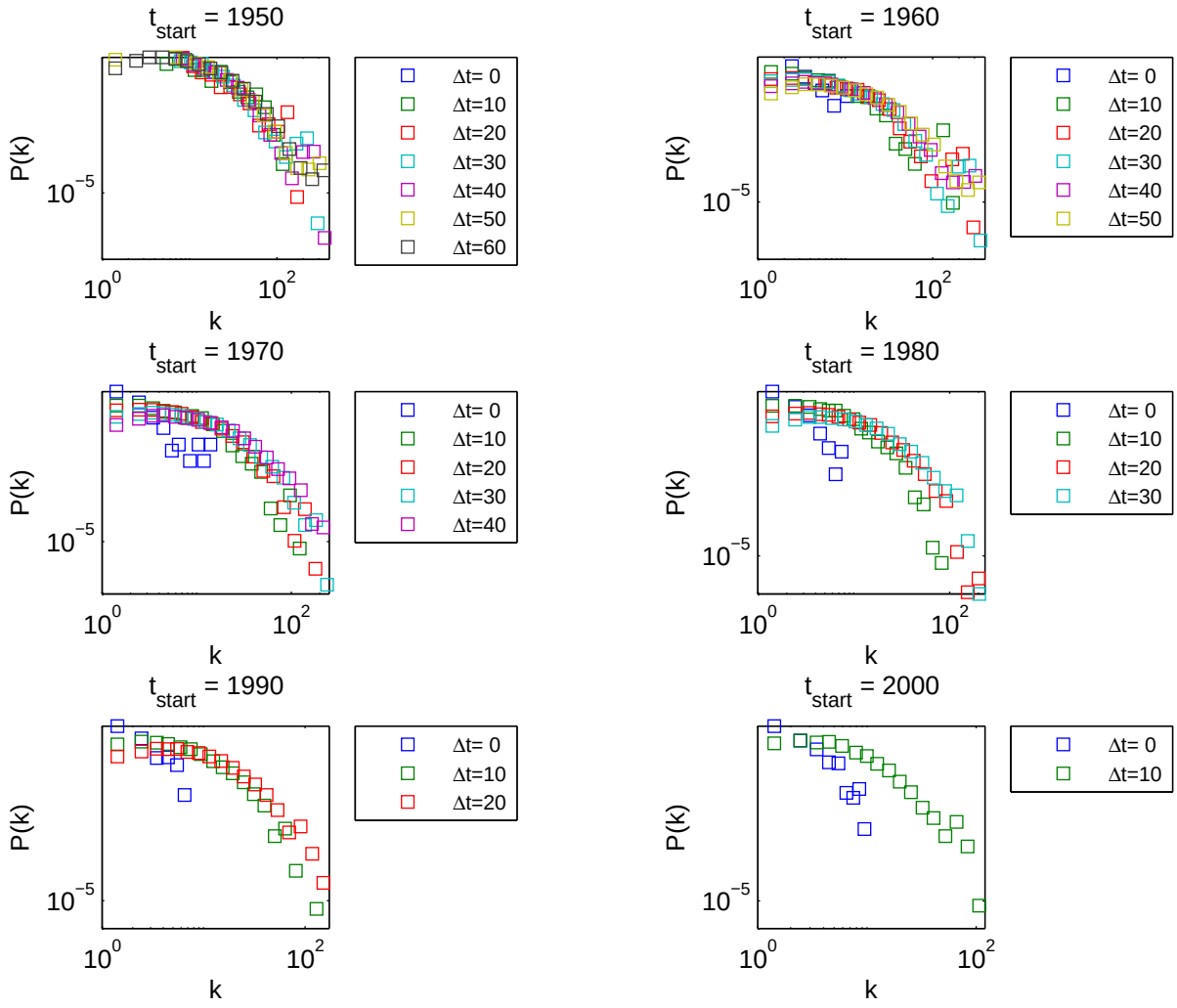

FIG. 15: Degree distribution for albums, log binned.

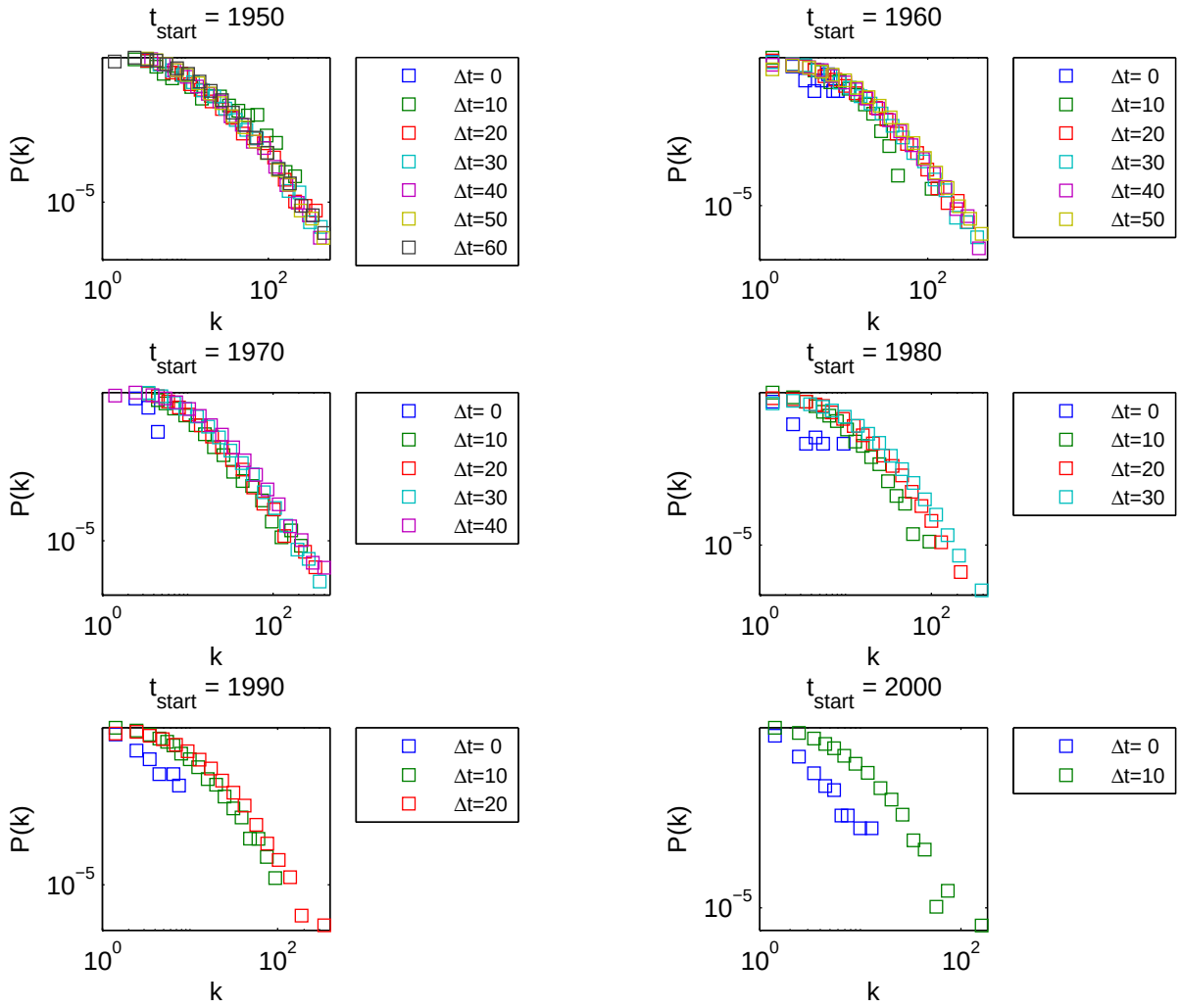

FIG. 16: Strength distribution for bands, log binned.

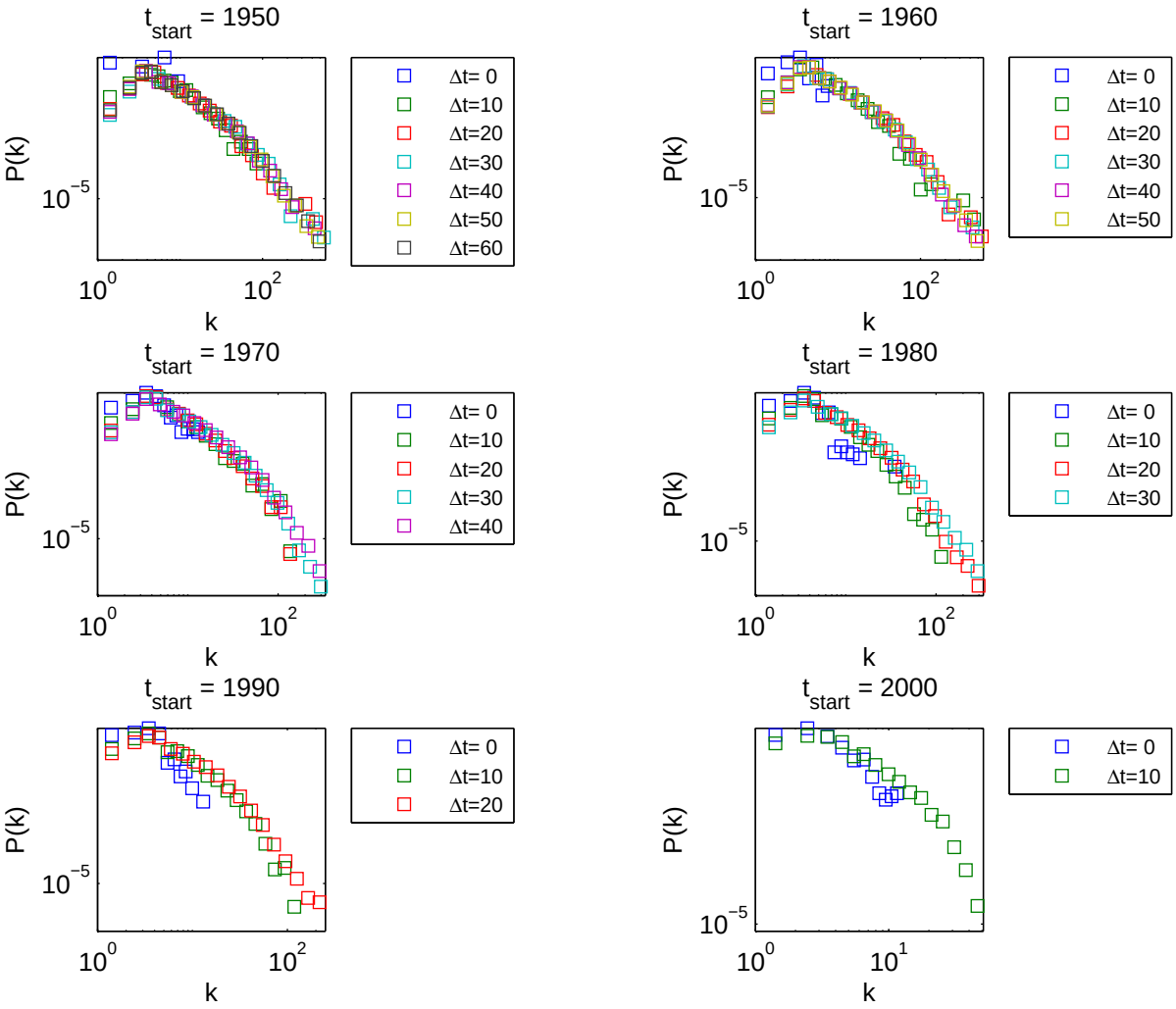

FIG. 17: Strength distribution for artists, log binned.

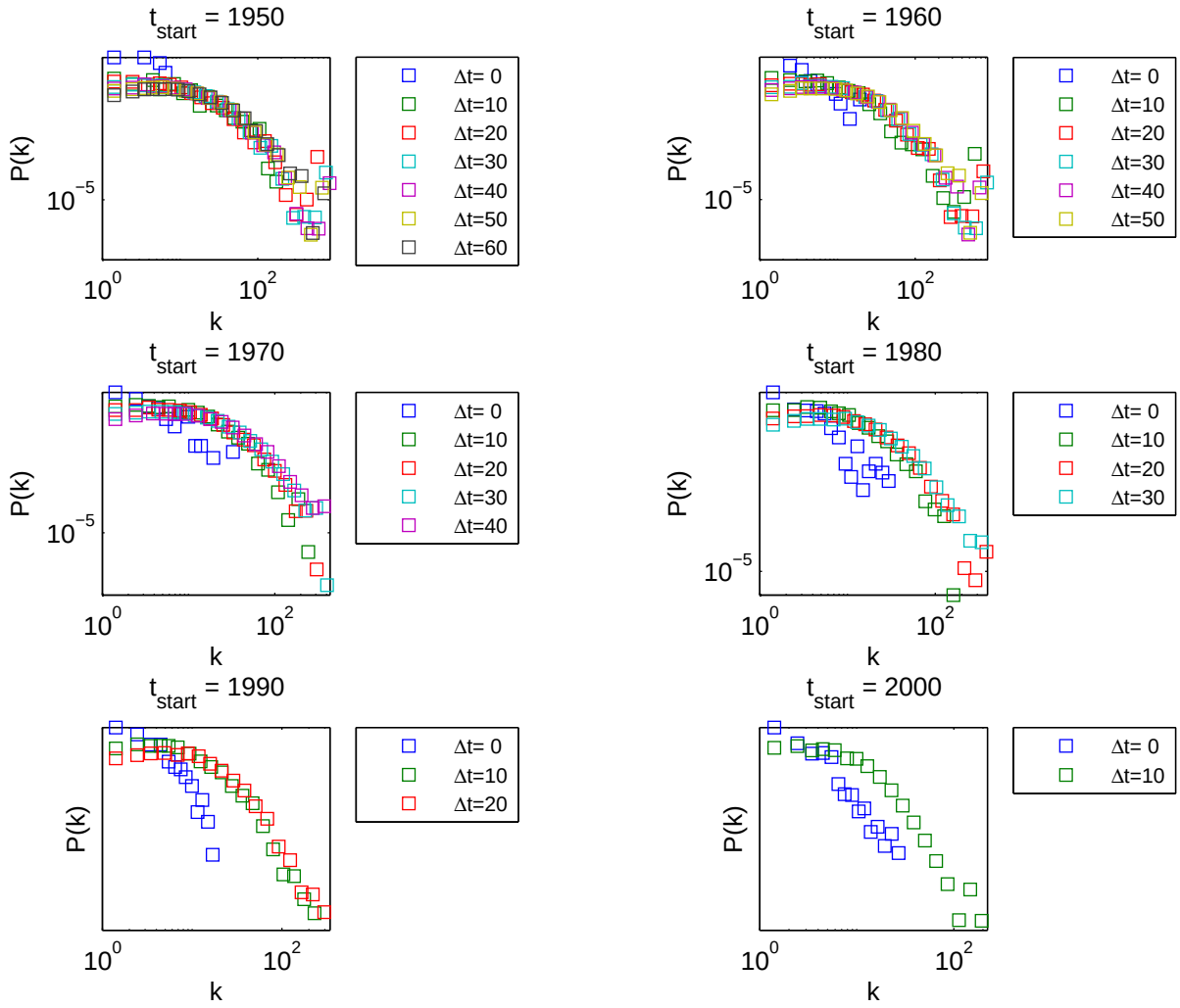

FIG. 18: Strength distribution for albums, log binned.

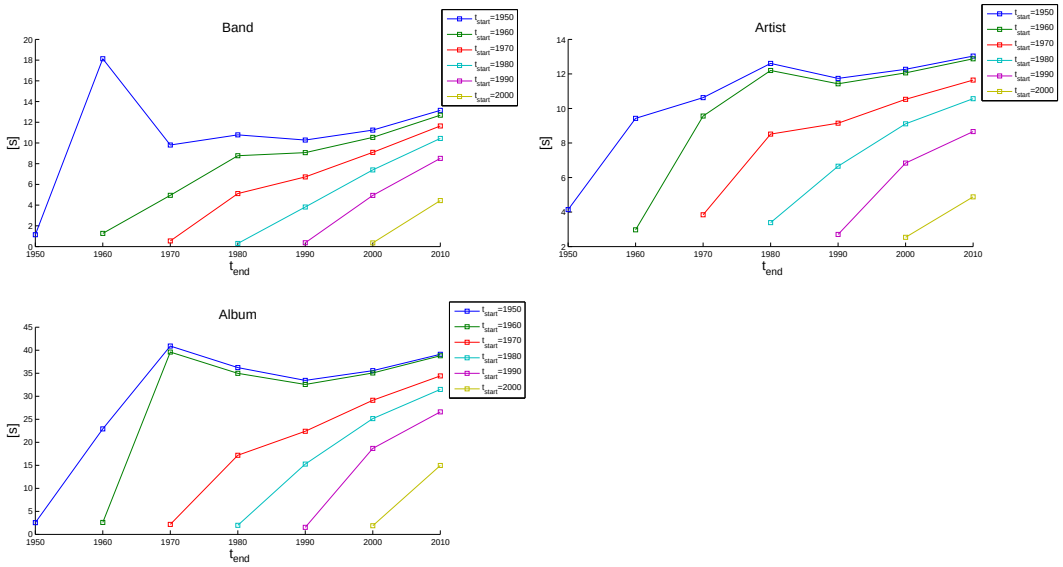

FIG. 19: Average strength in the networks compared to the time window.

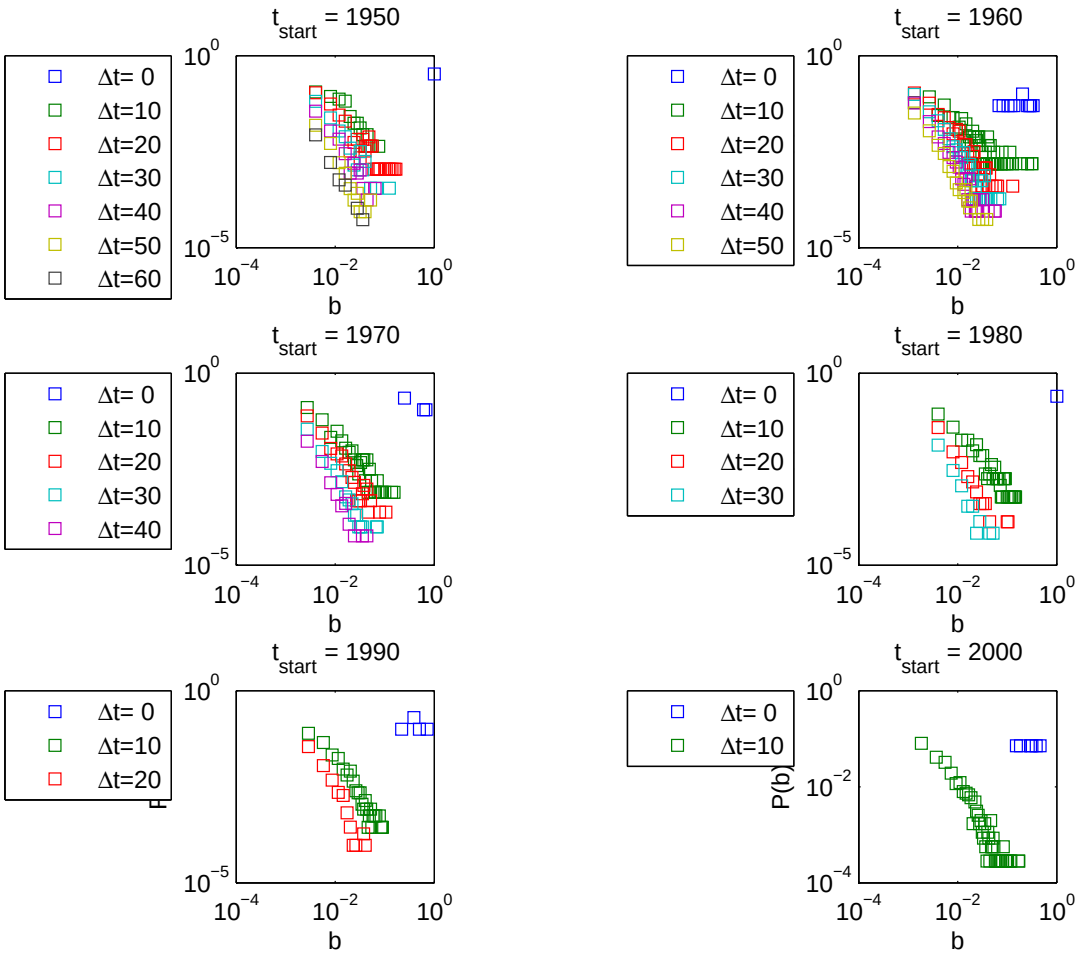

FIG. 20: Betweenness distribution for band networks.

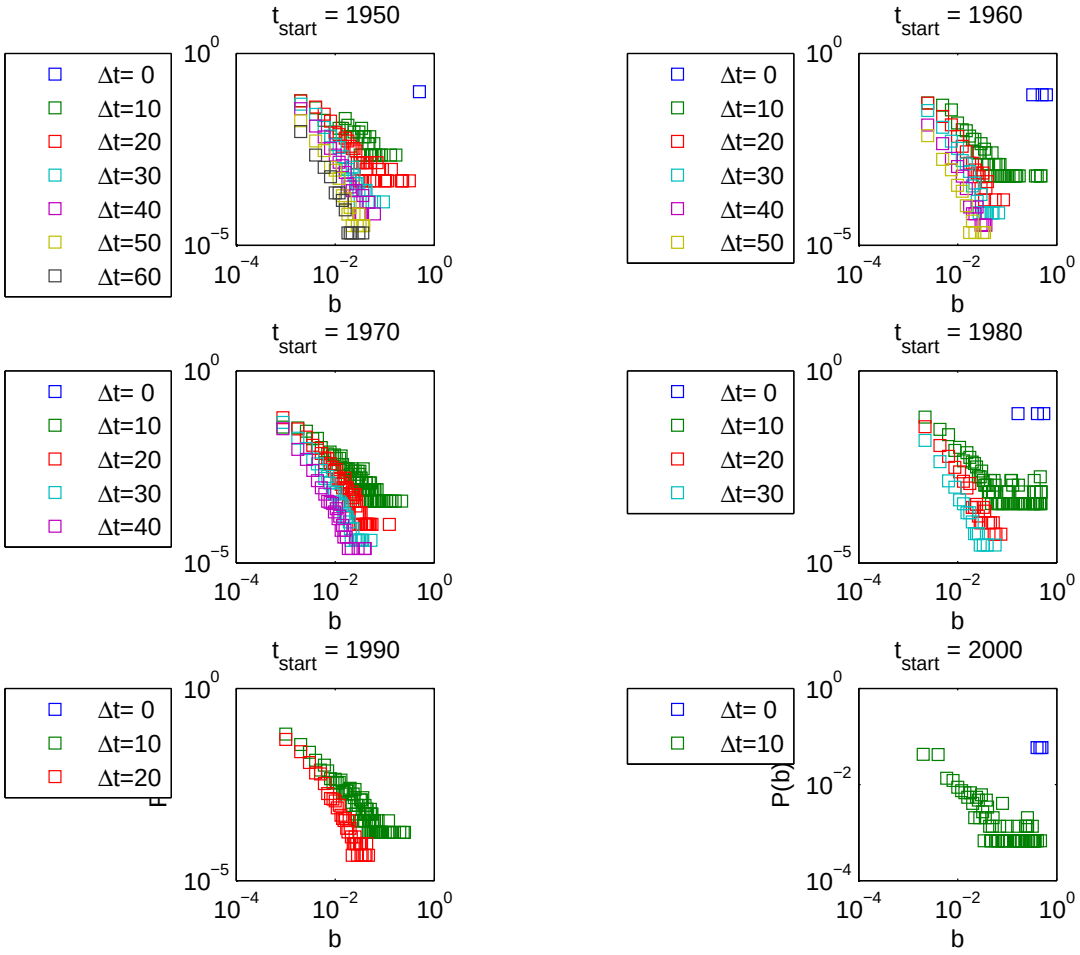

FIG. 21: Betweenness distribution for artist networks.

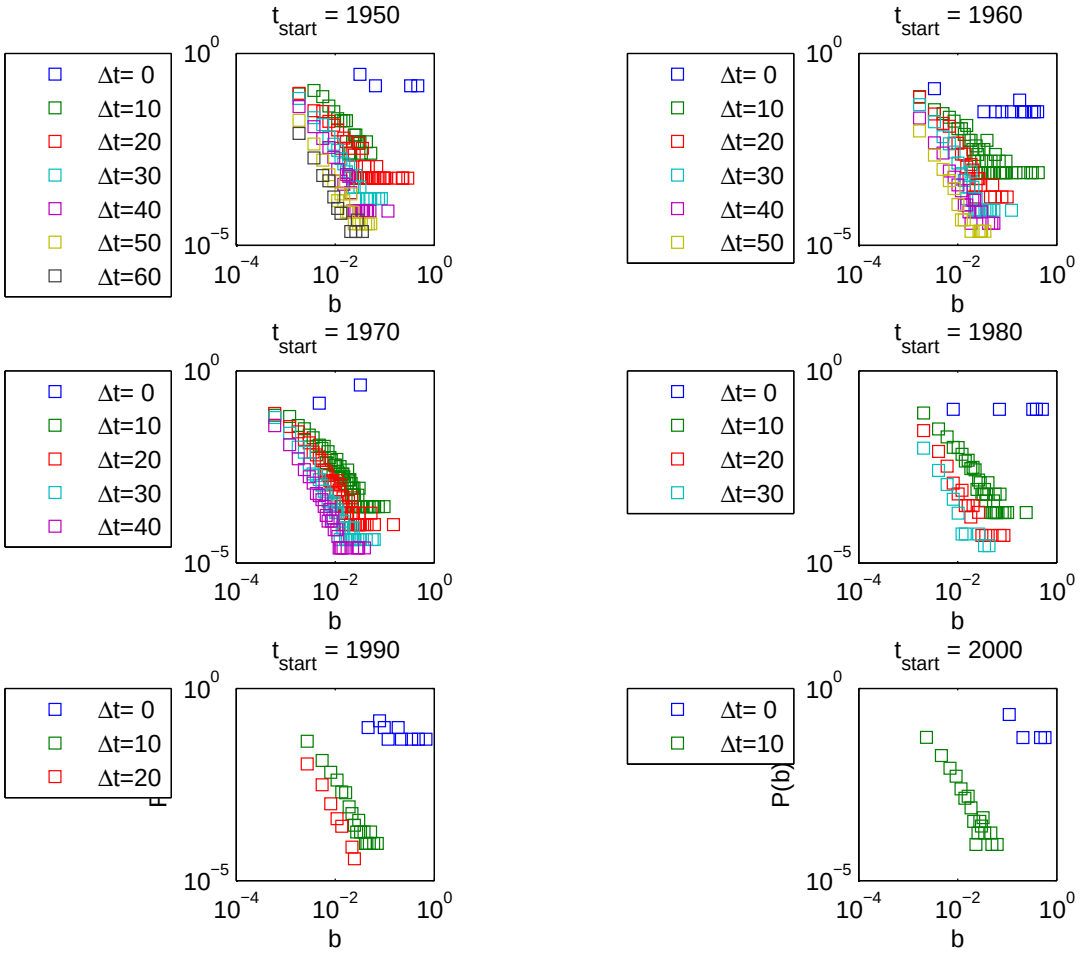

FIG. 22: Betweenness distribution for album networks.

3. Connected components

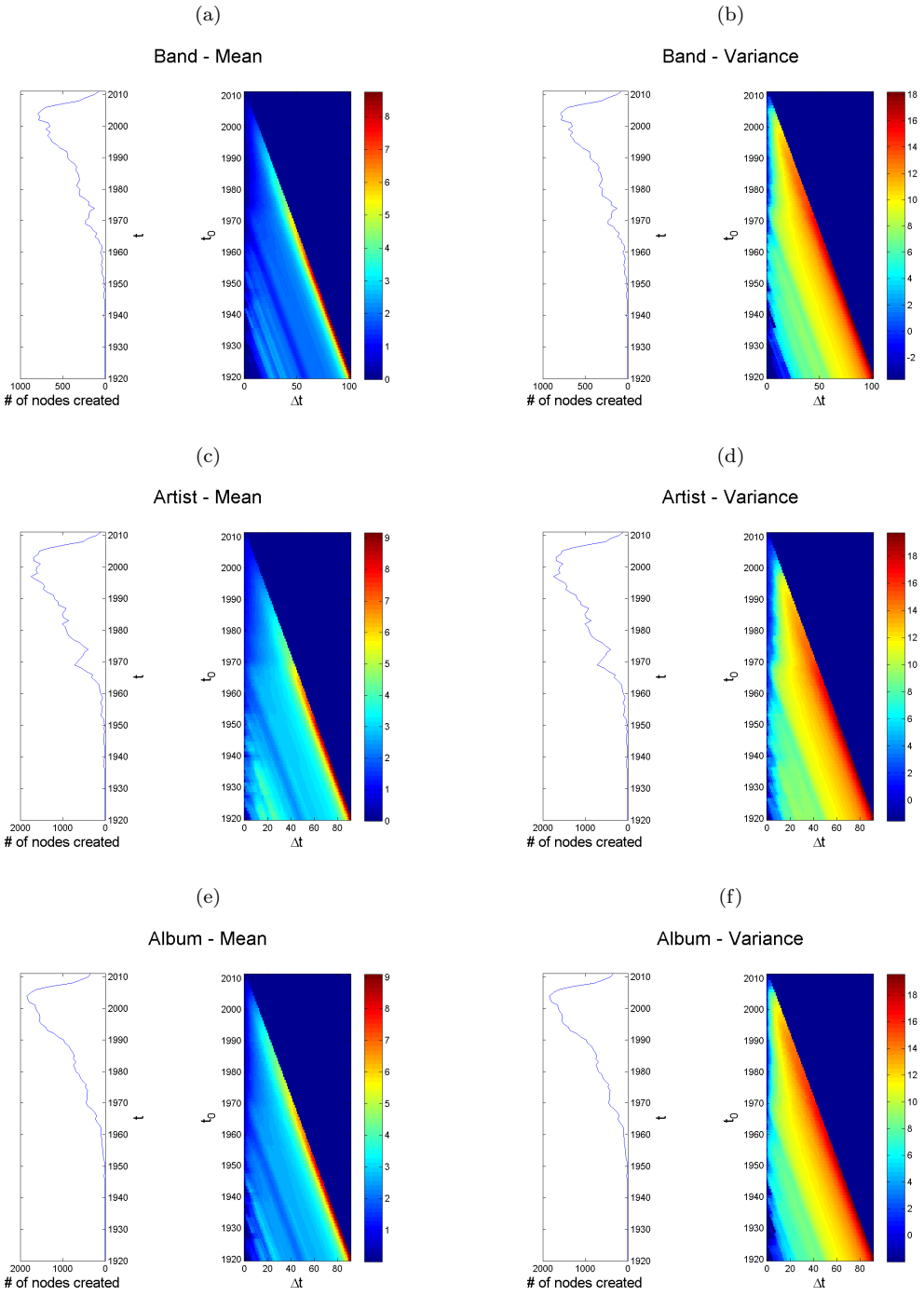

FIG. 23: The changes in the mean and the variance of the connected component size as the size and the position of the time window is varied.