Dynamic exploration of recording sessions between jazz musicians over time

Darya Filippova
Computer Science Department
University of Maryland
College Park, MD 20742
dfilippo@cs.umd.edu

Michael Fitzgerald
Learning Resources Division
University of the District of Columbia
Washington, DC 20008
mfitzgerald@udc.edu

Carl Kingsford
Computer Science Department
University of Maryland
College Park, MD 20742
carlk@cs.umd.edu

Fernando Benadon
Department of Performing Arts
American University
Washington, DC 20016
fernando@american.edu

***Abstract*—We present a new system for exploring, in an intuitive and interactive way, a large compendium of data about collaborations between jazz musicians. The system consists of an easy-to-use web application that marries a ego-network view of collaborations with an interactive timeline. We develop a new measure of collaborative influence that is used to highlight strong and weak collaborations in the network view. The ego-network is arranged using a novel algorithm for ordering nodes that avoids occlusion even when the network is frequently changing. Finally, the system is applied to a large, unique, hand-curated dataset of recorded jazz collaborations. The system can be accessed at http://mapofjazz.com/socialcom.**

I. Introduction

The evolving community of jazz musicians is an example of a social network where personal connections are essential [1]. In jazz, a highly collaborative art form, one person's individual style is a result of constant experimentation and exchange of techniques and ideas with fellow musicians [2]. Every musician is part of a dense network of collaborators with many transient connections: a composer may arrange music for multiple bands simultaneously, band leaders may recruit new members to their bands and lose them to competition, musicians' skills may improve to the point where they are featured as soloists and have a prominent place in the band. Study of recorded jazz collaborations can help identify influences on style, explain career success, and lead to a richer understanding of the progression of the jazz art form.

The traditional means by which these collaborations are explored is via the compilation and study of discographies presented as lists and tables in either hard-copy books or in computerized databases [3], [4]. These discographies list recording sessions, the roles each musician played in them, often songs and albums that were produced as a result, along with other information. They provide an extremely rich source of information from which to trace the collaborations of musicians. However, such a static and textual presentation is difficult to comb through and does not easily allow the user to comprehend the dynamics of changing collaborations. While changes in band membership are easily traceable, it is hard to assess the overall contribution of a band member over time unless the historians are intimately familiar with the band's history. A system for exploring jazz collaborations that makes large discography data more approachable is needed.

Social networks such as those between collaborators or friends are an object of intense study. Often, such connections are assumed to be immutable and the networks are considered static. However, there are many social interactions that violate this assumption: work colleagues, neighbors, acquaintances, friends, and family may all change over the course of a person's lifetime. This is particularly true of jazz collaborations where band members frequently come together for only a single recording session, and where some musicians have played with over a thousand other artists over their decades long careers. Understanding changes in connectivity on a scale of the whole network can provide an insight about the global change within the network, but has proven to be a difficult task for both algorithmic [5], [6], [7], [8] and visualization [9], [10], [11] approaches. The dynamism of jazz collaborations requires new approaches to visualize frequently changing networks.

Apart from the topological changes, there may be other, more subtle variations in the characteristics of relationships over time. Node or edge attributes may change, e.g. the relationship between A and B may gradually change from an acquaintance to close collaborators, or the "importance" of A's immediate collaborations may grow, thus indirectly increasing the importance of A itself. Collectively, these changes may affect a life of an individual in a significant way, but these observations are lost in the sea of data when analysing the network as a whole. This, along with a traditional focus on the lives of individual musicians, leads to the desire to have a visualization that can be focused on subregions of the entire space of collaborations. It also leads to the need for techniques to quantify the strength of the relationships encoded in an evolving network and to show this information effectively through a visualization.

In this paper, we propose a way to quantify the pairwise influence between two actors in the network based on the frequency and timing of the events in which they have both participated. We provide a visualization system that displays much of the data available in large discographies. We visualize collaborative influences with an interactive egocentric network view that allows users to focus on an individual and observe large and small scale changes in collaborations. We couple this network view with an interactive timeline that allows the user to see how influences have changed over time. We also introduce a novel algorithm that arranges collaborator nodes around the central musician in a way that minimizes node occlusion and variation in node positions as the network view changes over time. This helps the users to maintain their mental map of evolving collaborations. We demonstrate the utility of this approach on an extensive hand-curated collection of jazz collaborations spanning almost a hundred years.

II. Related work

A. Dynamic network visualization

There are two main approaches to visualizing time-varying networks: to show animations by constantly recomputing layouts at every time step [12] and to compare static snapshots of a network at several distinct time points. For either approach, the objective is to highlight the differences between the network views at different time steps.

Brandes and Corman [13] stack network snapshots on top of each other in 3D where the nodes with the same labels are connected by vertical columns. The graph is laid out using a spring-embedded algorithm, and each slice shows only the nodes and edges present at that time point. Diehl and Görg [14] place the two network snapshots side by side and propose three algorithms that minimize dissimilarities between the two representations. A recent study by Khurana et al. [15] presents an extension to NodeXL [16] that aggregates the snapshots of a network at two different time points into a single view. The edges are colored based on the time interval at which they existed. Additionally, NodeXL plots the values of several common graph properties such as node and edge counts as they change over time between the two time points. An approach by Yi et al. [11] combines a small multiples display (a histogram showing node degree over time) and a matrix representation of a network into a single view.

Graph layout for the animations may be fixed or constantly updating at each time step. Moody et al. [17] keep the node positions fixed and allow the edges to appear and disappear as the time progresses in their *flip-book* animations. Yee and colleagues [12] allow the nodes to move between the concentric circles along a smooth tangential trajectory.

While these approaches are able to highlight the topological changes, they all make an assumption that node and edge attributes (such as edge length) are either absent or remain static over time.

B. Circular and ego-network layouts

As early as 1990 circular layouts were used for organizing trees by placing the root in the center and assigning the child nodes to concentric circles with increasing radii [18] with nodes at the same level of a tree assigned to the same circle. Six et al. [19] extended the technique to work for more general graphs by connecting multiple circular structures. Yee and colleagues [12] have developed the technique further to support nodes of different sizes (where node size is proportional to some node attribute). They adapted their layout to handle dynamic graphs by showing animations of nodes traveling on a smooth trajectory from old locations to the new ones.

Wang, Shi, and Wen [20] experiment with a dynamic ego-network design where the central node and all of its connections are shown at the same time. The main node has several copies with each one representing the node at specific point in time and linked only to those nodes with which it was associated during that period. This approach is not feasible if the central node has many collaborators over his or her career.

Gansner and Koren [21] develop heuristics that order nodes on circle's periphery in a way to minimize the edge crossovers and to reroute some of the links to go outside the circle's circumference. These authors suggest edge bundling for the links inside the circle to reduce clutter further. Their work does not consider dynamically changing links.

C. Artist collaboration networks

Artist collaboration networks have received their fair share of attention in the wake of area of social network analysis. The data on interactions among artists is available in some databases [22], has been collected through surveys [23] or manually by processing the tapes of interviews with the artists [24]. Due to difficulties in data collection, these sources cover only a few artists and lack temporal information about their collaborations. For example, Gleiser and colleagues [22] base their analysis on 198 bands that were active in 1912-1940. Heckathorn and Jeffri's survey reached out to 110 musicians in New Orleans, 264 in New York, and 300 in San Franscisco [23], a small fraction of the estimated 33,000 jazz musicians living in the New York.

Examples of applications supporting exploration of such networks are few. An online Classical Music Navigator [25] helps users expand their musical interests by suggesting composers who influenced or were influenced by a composer the users initially searched for. The Navigator offers a simple text and link interface. A visualization of Last.fm data [26] allows one to compare two artists and their musicial associations, but does not provide any intuition about collaborations between the two artists.

III. Jazz Collaboration Data

The Map of Jazz uses data that have been collected over the course of more than twenty years of discographical research, with additional content subsequently added in targeted batches. The data come from myriad sources: from general and artist

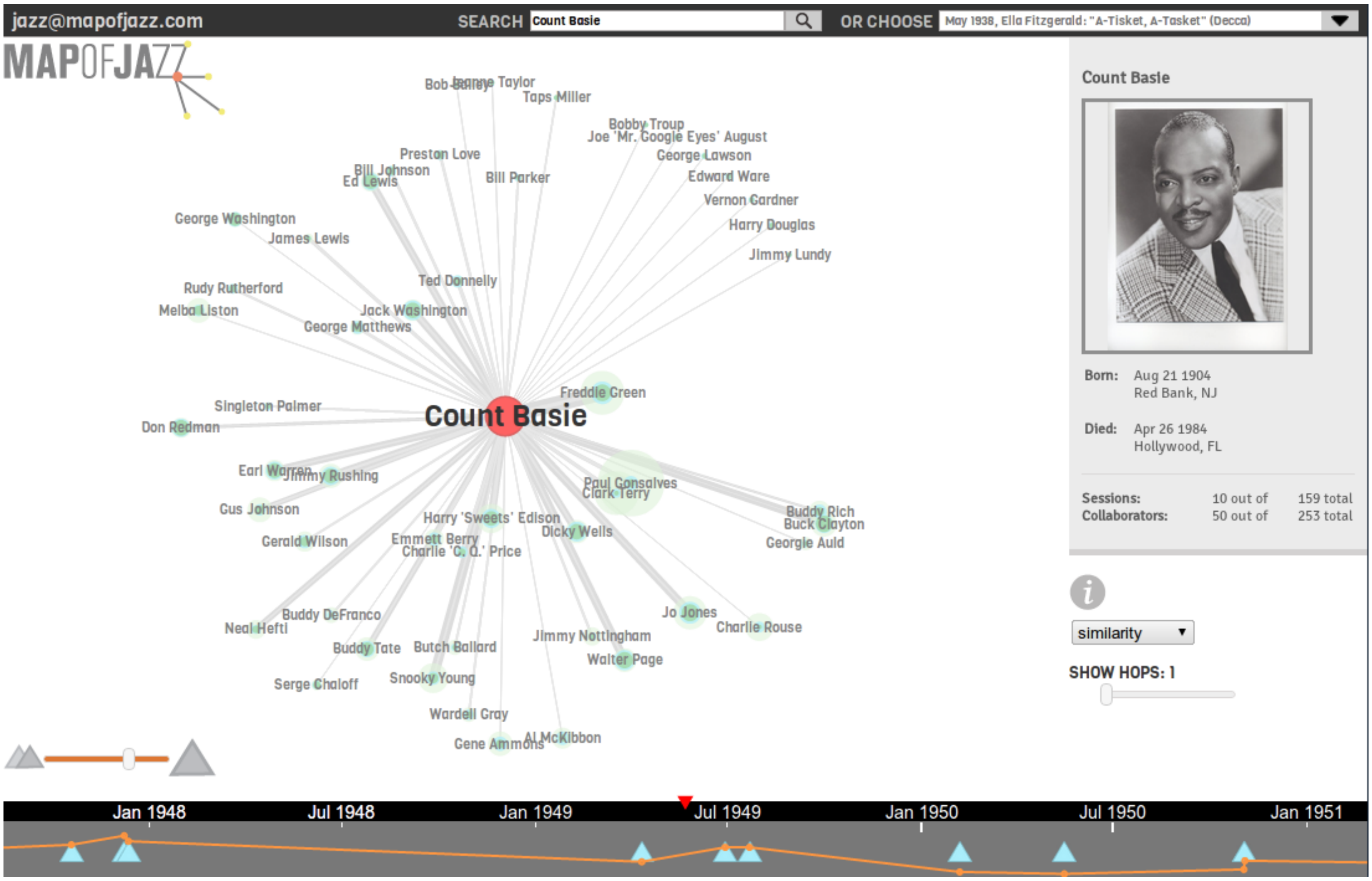

Fig. 1. View of the Map centered on Count Basie. Musicians closest to the center have played a more significant role in Basie's career at the end of 1940s (Clark Terry, Paul Gonsales). Musicians on the periphery have not recorded with Basie as often or not as recently as 1949. Only 10 out of 159 sessions are visible in the range currently displayed in the timeline. The node sizes indicate how active the musicians were throughout their career while the edge thickness corresponds to the strength of the relationship between Count Basie and his collaborators.

discographies to specialist journals, magazines, and newsletters to biographical and historical monographic literature and many more. In almost every case, a single entry is a collection of data from multiple sources because no one source covers every aspect of the recording session. The sessions cover a period of time from the early 1920s to the present day.

While there have been other attempts at digitizing discographical information, these used closed, proprietary systems that lacked both the ability to export and import or edit the information, and ultimately were unable to satisfy the information needs of serious researchers. The data used by the Map of Jazz was collected and stored using the open-source discographical software BRIAN [4] named so for the English discographer Brian Rust who perfected the session-based format for print discographies [27], [28]. BRIAN's support for data export and import allowed for multiple users to contribute to the project and amounted to a large number of catalogued recording sessions. The Map of Jazz is the first project to use large amounts of BRIAN data outside of the application itself.

At the heart of BRIAN is the conceptual idea that the session is the primary entity unlike many other databases designed to store sound recordings information. Sessions are events that have defined locations, both chronologically and geographically. Out of the many layers of details available in BRIAN, the Map of Jazz uses the top-level data on the sessions and musicians who performed during them, tus shifting the focus to the interpersonal relationships between the artists. The attribute for the main musical instrument helps to distinguish Bill Evans the saxophonist from Bill Evans the pianist; it also records what instrument each performer played.

At the time of publication, the database contained information on 11824 musicians and a total of 13873 recording

Name	Degree
Slide Hampton	1230
Kenny Barron	1090
Ron Carter	785
Michael P. Mossman	692
Freddie Hubbard	691

Fig. 2. Top 5 musicians with the highest number of collaborations. Slide Hampton was a prolific composer who has provided arrangements for multiple bands, hence, his interaction surpasses that of many famous band leaders.

sessions. The average number of people per recording session was 7.39 with the smallest sessions having just one performer and the largest session having 72 people (not including the members of orchestras). Network diameter (the longest among all shortest path between all pairs of vertices) was equal to 8, and the average (*characteristic*) shortest path was equal to 3.24. As with many social networks, very few musicians have a high number of connections with an overall average node degree of 33.68 (see Figure 2 for the top 5 artists by degree). However, the network does not pass the test for scale-freeness according to the test developed by Clauset et al. [29].

IV. Interface design

The Map of Jazz focuses on the dynamics within the collaboration networks of individual musicians. An egocentric network layout (Figure 1) is coupled with an interactive timeline that allows users to navigate through various time periods and observe gradual changes in the person's collaboration network. The timeline is augmented with aggregate statistics, such as the number of people per sessions, to aid in navigation through time. In the network view, nodes representing collaborators are arranged around the central node in a manner that preserves nodes' relative positions across time and shows the relative strength of a tie between the main musician and his or her collaborator. Nodes and edges used in the traditional node-link network representation are extended to display the change in attribute values over time. Finally, the interactions between the currently displayed main musician's collaborators can be explored on demand by highlighting a node of interest to which Map of Jazz would show the connections between the nodes in the neighborhood. The central artist can be selected by double-clicking on a performer's node or by entering a musician's name in a search box. To help users pick a starting point, Map of Jazz offers a dropdown with 21 hand-picked sessions that stand out in jazz history.

A. Timeline

Users may navigate the extensive Map of Jazz timeline by dragging it with a mouse. They may also zoom in on a particular period of the musician's life or zoom out to get an overview of the artist's career. Every time the users interact with the timeline, the ego-network is updated to present an accurate snapshot for the selected time period.

Sessions. The triangles on the timeline represent recording sessions (Figure 3). When users hover over the triangle with a mouse, the session and the collaborators who recorded for that session are highlighted in yellow. Users may click on a session to select all the participating collaborators and the session itself, and vice versa. A tooltip that appears above the triangle icon summarizes the information about the session: the date, location, and a list of participating musicians along with their primary skill (an instrument they played, e.g. alto saxophone, or a role they took, e.g. band leader, during a session).

When the date of birth and/or death are available, the span of time from birth to death (or to the present day) is colored in a lighter shade of gray to indicate the span of the musician's life.

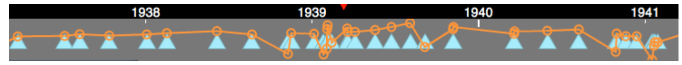

Fig. 3. A timeline augmented with a session similarity graph. Triangles represent the individual recording sessions. Most sessions on the timeline have a high pairwise similarity indicating that session memberships changed only slightly.

Augmented timeline. To aid users in focusing on a specific time period of interest, the Map of Jazz offers several basic metrics to be overlaid on top of the timeline. These include:

- the number of collaborators per session,
- the number of unique musicians the person collaborated with up to this date,
- the Jaccard similarity between the members of the current session and the previous session.

The number of collaborators and the speed at which a person attracts new collaborators have been shown to be significant in scientific [30] and jazz [1] collaborations, while the session size may explain the mechanism of acquiring new collaborators (i.e. switching between the bands versus playing with the same band).

B. Measure of collaboration strength

The notion of mutual influence between any two musicians lies at the heart of Map of Jazz. To quantify the strength of the relationship between any pair of musicians, we make an assumption that the recording sessions are not spontaneous events, but rather a result of previous undocumented collaboration (e.g. concerts and rehearsals). Consequently, the professional relationship between the musicians is not likely to start right before a session nor to stop immediately after recording it, but rather to grow before the session and wane gradually over time. The collaboration is strongest around the time of a recording session and is increased even more if the musicians record several sessions over a short period of time. With these assumptions in mind, we define a function of *collaboration strength* that takes into account the frequency and proximity in time of the collaborations between the two musicians:

$$g(t; S, \sigma) = \alpha_\sigma \sum_{s \in S} e^{-\beta_\sigma (t_s - t)^2}, \tag{1}$$

where S is a set of all sessions the two musicians shared, t is a point in time for which we want to evaluate the collaboraiton strength (i.e. the center of the timeline), and t_s is the date for a specific session s. To make the decay of the collaboration strength smooth, we model it as a normal distribution with $\alpha_\sigma = 4/(\sigma^2\sqrt{2\pi})$ and $\beta_\sigma = 1/(2\sigma^2)$. The function (1) is similar to kernel density estimator for normally distributed values, where choosing the bandwidth for the kernel is a known hard problem. To make the function smooth, we take an affine comination of two functions:

$$f(t; S) = \delta g(t; S, \sigma_1) + (1 - \delta) g(t; S, \sigma_2), \tag{2}$$

with $\sigma_1 = 1$, $\sigma_2 = 3$, and $\delta = 0.9$. The resulting function assigns more impact to the interactions that were closer to t in

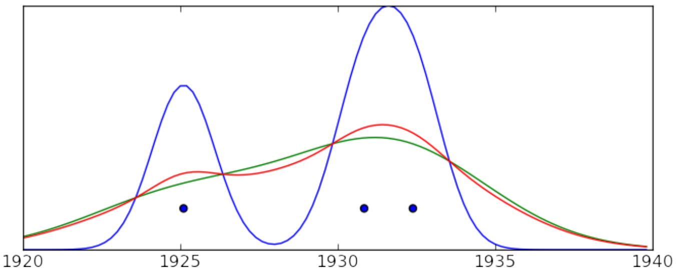

Fig. 4. The collaboration strength function (red) is a combination of two functions: one for which the value grows fast as t gets closer to the session (blue) and another one for which the change in value is much smoother (green). Here, two hypothetical musicians played together in 1925, 1930, and 1932 (blue dots). Their relationship is strongest when several sessions occur within a short period of time such as in 1930–1932. The collaboration strength drops slightly between 1925 and 1930 and declines rapidly after the last session in 1932.

time rather than assume an equal influence for all interactions no matter how long ago they occurred (Figure 4).

C. Egocentric network view

An egocentric view allows the user to track the fluctuating interactions between a given musician and his or her peers over time. Using the timeline, users may select a time range of interest and focus on the collaborations between the main musician and his or her collaborators within that period. Only the collaborators that were involved in sessions that occurred during that time range will be shown. For every collaborator v, we compute the collaboration strength between the center musician u and v where S, the set of sessions considered in the collaboration strength function, are the sessions that are currently visible on the timeline. The length of an edge connecting u and v is inversely proportional to the strength of a tie the two artists share, as measured by the collaboration strength function (1). This causes the collaborators who record with the main musician u often or have recorded with him or her recently to be placed closer to the center, while musicians who record with u rarely or have only recorded with u a long time ago are placed on the periphery (Figure 1).

As users drag the timeline, the collaborators' nodes may move closer to the center or may drift away to the periphery as the set of sessions that is displayed changes. Nodes disappear from the view when they are no longer involved in any session that is visible on the timeline, and new nodes appear when sessions containing them come into view on the timeline. To minimize the cognitive load on the users, the Map of Jazz preserves the angular positions for every collaborator node within the ego-network formed by a particular main musician. For every musician u, we assign a unique angular sector to every collaborator v who has ever recorded with u. Their trajectory, a straight line connected to the center, can be easily traced (Figure 5) as collaborator nodes move to or from the center.

If care is not taken with the assignments of nodes to angular sectors, collaborators may crowd near the center if the main musician consistently played with the same set of people (i.e. their band). To avoid this, we propose an ordering algorithm in Section IV-D to arrange the collaborator nodes in a way that assigns dissimilar angles to collaborators with which the center node has had similar patterns of collaboration.

D. Ordering collaborators in a circle

The Map of Jazz maintains the same angular node positions across every time period, allowing easier comparison of the network at different time points. As users navigate the timeline, new nodes may show up on the map, but they will never change the angular position of the nodes already on screen thus preserving the users' mental map of the ego-network.

The permanent angular node positions also help alleviate another problem common to graph drawing: node occlusion. More often than not, jazz musicians have a band, or several bands, with which they play and record on the regular basis. In this case, every person in the band would share a strong tie with the musician in question. The Map's circular layout would try to place all such frequent collaborators near the center causing the nodes and labels to occlude each other.

To minimize such occlusions, we identify groups of people that are likely to be at the same distance from the central node at the same time. To find the groups, we first sample the collaboration strength function values between the main musician u and each one of its collaborators v:

$$x_{u,v} = (f(t_1; S), f(t_2; S), \ldots, f(t_{1000}; S)), \qquad (3)$$

where t_i are equally spaced time points in the range $[t_{\text{min}}, t_{\text{max}}]$ with t_{min} equal to the date of u's earliest record and t_{max} equal to the date of u's last recording session. In (3), S is the set of all the sessions in which central node u participated. Next, we construct a pairwise similarity matrix for all u's, $M_u = \left(m^u_{i,j}\right)$, where $m^u_{i,j}$ is the cosine similarity between vectors x_{u,v_i} and x_{u,v_j}. To make it easier for the clustering algorithm to find groups in this matrix, we set every $m^u_{i,j} < 0.9$ to 0. The resulting sparse matrix M'_u is taken as a weighted adjacency matrix of a graph G_u. We run the Louvaine clustering algorithm [31], which attempts to find clusters maximizing modularity [32], on G_u to identify groups of musicians for whom the x_{u,v_i} vectors were similar.

Performers in the same cluster interact with the main performer in a similar fashion. To spread them out, we assign them sectors of the circle that are far from each other. To assign all N collaborators to their angular positions, we iterate through the clusters in order of decreasing size. For each cluster C, we assign a node in C to the first empty sector and continue assigning $v \in C$ to empty sectors evenly around the circle at intervals of $\lfloor \frac{N}{|C|} \rfloor$ sectors. If at any time the target sector is not empty, we search linearly clockwise for the next available sector and continue assigning the remaining nodes in C starting from that sector (Figure 5). This heuristic — similar to linear reprobing in a hash table — attempts to ensure that the nodes belonging to the same cluster are spread out around the circle at equal intervals.

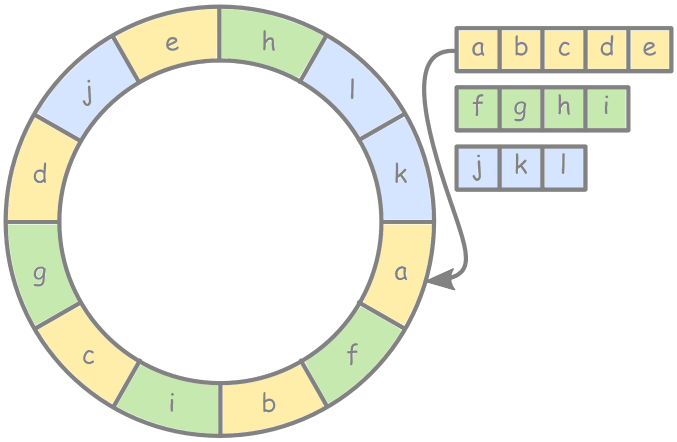

Fig. 5. Assigning angular positions for the nodes. Starting with the largest cluster, performers from the same cluster get assigned sectors of the circle that are far from each other. Here, a is assigned to the 1^{st} sector, b is $\lfloor \frac{12}{5} \rfloor = 2$ sectors away, and so on.

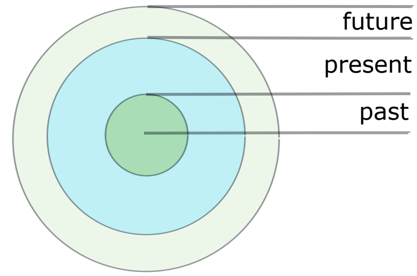

Fig. 6. The inner-most circle of the node glyph encodes the number of times the musician played in *some* session in the past. The middle circle (blue) represents the number of sessions the musician played with the main musician within the current time frame. The radius of the outer largest circle encodes the total number of sessions the musician played throughout his/her career.

E. Node and edge glyphs

A static snapshot of a collaboration ego-network may be misleading due to the fact that the past and future collaborations between the central and peripheral nodes are not visible. A collaboration that flourished previously may be represented by a single recording session in the selected time period and could be indistinguishable from many a sporadic one-time collaborations. Likewise, an ego-network does not provide enough detail about the level of activity for the nodes other than the central node based on the edge length alone. Questions such as: did the musicians represented by these nodes record often? How many sessions did they record overall within this time period? Will they continue recording actively? To answer these kinds of questions, we replace the standard graph nodes and links with more information-dense glyphs.

We use node glyphs to represent the collaborator's activity overall (i.e. sessions recorded, not necessarily with the main musician), and reserve the edge glyphs to represent collaborations between the main musician and the collaborator. The node glyphs, therefore, encode the musician's overall artistic output while the edges connecting it to the main musician quantify the strength of the tie between them.

Node glyphs (Figure 6) consist of three concentric circles. The inner circle's radius is proportional to the square root of a count of sessions this musician played in the past, i.e. before the sessions currently visible on the timeline. The second circle's radius represents the square root of the number of sessions from both the past and currently visible on the timeline in which that collaborator recorded. The radius of the outer circle is proportional to the square root of the total number of sessions the musician has recorded over their life. As users navigate the timeline in the direction of the future, more sessions would transfer to the "past" increasing the radius of the inner-most circle. The radius of the circle representing the "past+present" may fluctuate depending on the number of sessions currently visible on the timeline. The radius of the outer circle, i.e. the square root of a total number of sessions recorded, would not change.

Edge glyphs encode information about the overall count of recording session in which both the main musician and the collaborators participated (Figure 7). Edges are colored in varying shades of gray with the inner edge representing collaborations in the "past" and its thickness proportional to the square root of the number of sessions he or she shared with the main musician in the past. The thickness of the middle part is proportional to the root of the number of sessions for that collaborator currently visible on the timeline. Finally, the total edge thickness represents the number of sessions the two musicians played together overall. As users interact with the timeline, the thickness of the inner edges may change depending on the number of shared sessions in the past or currently visible on the timeline; however, the total thickness of an edge would not change.

Fig. 7. Edge glyph. The thickness of the inner-most inset in the edge glyph is proportional to the number of sessions the main musician and their collaborator played in the past, and the thickness of the "present" inset corresponds to the count of sessions in the past and present. The total edge thickness represents the number of sessions the two musicians played together overall.

Various combinations of node and edge thicknesses should alert the users to different modes of collaboration. A combination of a large peripheral node and a narrow edge connecting it to the center indicates that while the collaborator has played many sessions overall, they played very few with the main musician. Further, if the node's inner "past" circle is large, such combination then indicates that the collaborator is an established musician who is sharing their skills with the up and coming musician at the center of the ego-network. On the contrary, if the inner "past" circle is small and the majority of sessions are in the "future", such behavior may indicate that the collaborator has joined the main musician briefly and later went on to form a successful career of their own. A

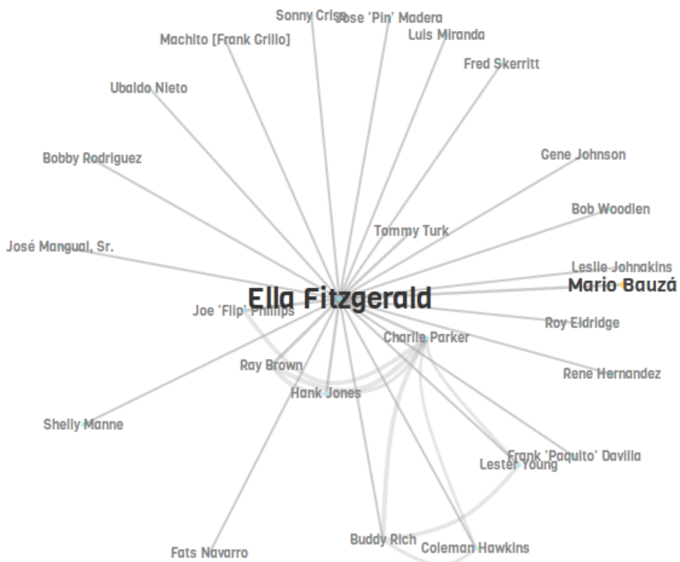

Fig. 8. Neighbohood connections among Ella Fitzgerald's collaborators

case where both the node and the edge are thick indicates that the two musicians have recorded many sessions together and, therefore, share a strong tie.

F. Exploring collaborators' connectivity

An ego-network allows users to focus on the dynamics of a few collaborations at the expense of hiding all other topological information. Knowing which, if any, peripheral nodes collaborate with each other helps to understand the communities that form in the immediate neighborhood of the main musician. The Map of Jazz provides such details on demand: when users hover over a collaborator's node v, the Map renders additional edges that connect v and its own collaborators (Figure 8). The thickness of the edges corresponds to the number of sessions the two musicians played in the past (including those they played with the main musician). When the edges connecting the main node and the performers on the periphery are thin indicating few collaborations, it would be noteworthy to see thick edges between those performers. Such a situation would imply that those artists form a strong community, or a band, outside of the main musician's neighborhood.

V. Example Explorations with the Map of Jazz

Duke Ellington's career spanned more than half a century from the early twenties to his death in 1974. The first record on the Map's timeline dates back to July of 1923 with Ellington playing on the piano and Elmer Snowden as the band leader — Ellington was yet to form his own band. Ellington's egonet for the period of 1923–1928 has several prominent performer nodes: Harry Carney, Otto Harwick, Sonny Greer, Fred Guy, Barney Bigard, and Wellman Braud. The size of Harry Carney's node, for example, indicates that he participated in a significant number of recording sessions throughout his career (1487 sessions), and the thickness of the edge connecting him to Duke Ellington reveals that most of those sessions were recorded with Ellington (1480 sessions). The same holds for others in the list above — their careers were tightly knit with Ellington's and helped him establish himself in the jazz world.

From the details window, it is clear that Duke Ellington had a very productive career: over the course of his life he participated in more than 1740 sessions with 600 musicians. If the user zooms out on the timeline, it becomes evident how densely packed Ellington's recording sessions were, with the last record right before his death. The pairwise session similarities that are visible on the timeline and the average pairwise similarity of 0.60 suggest that Ellington played with a core group of close collaborators who would replace one another over the years, but never would the whole band membership change all at once. Switching to the session size graph, users can see that the average number of people per recording session was 14.19 with the largest session at 29 performers recorded in January 1968.

Among Ellington's closest collaborators, several stopped collaborating with him either permanently or for a significant period of time. Ellington and Greer recorded 588 sessions together from 1923 to 1951, but there are no sessions past that: their collaboration ended after Greers's propensity for drinking forced Ellington to hire a second drummer to replace Greer. Apart from Greer, Johnny Hodges and Lawrence Brown, two of his most prominent collaborators, have a gap in recording sessions starting in 1951 which correlates with both musicians leaving the band to pursue their ambitions elsewhere. The recording sessions that included Hodges restart 5 years later and span all the way until his death in 1970; Brown rejoined the band later in year 1960 to record 432 more sessions with Ellington. The closely coupled timeline and ego-network displays allow such events to be found with relative ease.

Slide Hampton has the most collaborators (1230) among all musicians represented in the Map of Jazz. He has composed and arranged music for many prominent musicians such as Kenny Barron, Chick Corea, Tommy Flanagan, Dizzie Gillespie, Clark Terry — their large nodes stand out in Hampton's egonet — as well as hundreds of lesser-known performers. Dragging the timeline across the length of his career shows that while at any given time period Hampton is connected to many artists, he rarely collaborated with them for prolonged periods of time: the collaborator's nodes do not move close to Hampton's central node, but rather stay at the periphery. The difference in collaboration style between Ellington and Hampton is especially pronounced when one compares their average sessions similarity (visible on the timeline): Hampton's average session similarity is at 0.19 compared to 0.60 for Duke Ellington. The average session size for Hampton is 12.09 — combined with the low session similarity we can conclude that Hampton accumulated the highest number of total collaborations by continuously recruiting new collaborators.

Count Basie's sessions account for 159 sessions available in the Map of Jazz, and his number of collaborators (253) may seem modest when compared to Slide Hampton or Duke Ellington. The first session available in the Map of Jazz dates back to 1936, when Basie acted as a band leader and pianist and Lester Young was on tenor sax (the roles of each musician in each session are available by mousing over the session in the timeline). The large size of Young's node in the ego-network foretells his successful career — indeed, later he became one of the most influential saxophone players.

By double-clicking on Young's node, we can start navigating through his personal timeline. Young recorded with Basie often in the 1930s and 1940s. The records are sparse for the period of 1941–1945, first due to an American Federation of Musicians' recording ban, then due to Young's being drafted into the army and serving a year-long jail term after being dishonorably discharged from service. Again, this gap in jazz productivity is clear from the display of sessions in the timeline. Among his later collaborators are Billie Holiday, Charlie Parker, Buck Clayton, and Coleman Hawkins with whom Young recorded several times in 1946. The "hops" edges reveal that Parker, Clayton, and Hawkins had numerous sessions together that did not include Young. Later, his recording sessions become sporadic, possibly due to Young's deteriorating technique and health.

VI. Discussion

The Map of Jazz approaches the problem of dynamic graph visualization from a new angle: instead of tackling the hard problem of visualizing large graphs and tracking temporal changes on a global scale, the Map focuses on individual nodes and local changes that would have an immediate and personal effect. The Map of Jazz arranges related nodes into an ego-network with a single individual in the center surrounded by his close neighbors that are placed according to the strength of their connection with the central node. Users can explore the dynamic properties of the network by dragging the time slider or zooming in on a particular era of interest. The ego-network adjusts the node positions according to the varying strength of their connection to the central node during that time. The appearance of nodes and edges updates as well to reflect the change in their attributes.

We tailor our visualization to assist the exploration of a novel jazz collaboration network. In social interactions, the frequency and recency of interactions determine the strength of the collaboration between individuals. We propose and implement an collaboration strength function that takes into account both past and future interactions and helps quantify the strength of the relationship between the two musicians at any point in time.

The concepts developed for the Map of Jazz can be applied to other social networks that record multiple interactions between individuals and to dynamic networks in general, especially those where the numerical attributes on nodes and edges change over time. One such example is the gene co-expression network where genes control the expression of other genes in the cell. The amount of one gene product may change over the natural cycle of a cell (cell division, growth, death) and affect the behavior of related genes. Applications to toher collaborations networks such as co-authorship data is straighforward.

Acknowledgment

This work was partially supported by a grant from the National Endowment for the Humanities [HD-51098-10]. D.F and C.K. were partially supported by grants from the National Science Foundation [CCF-1053918, EF-0849899, and IIS-0812111], the National Institutes of Health [1R21AI085376], and a University of Maryland Institute for Advanced Studies New Frontiers Award. C.K. received support as an Alfred P. Sloan Research Fellow.

The authors wish to thank Guillaume Marçais, Geet Duggal, and Rob Patro for insightful discussions. We are especially thankful to Yana Sakellion who helped with visual design.

References

[1] D. L. Pinheiro and T. J. Dowd, "All that jazz: The success of jazz musicians in three metropolitan areas," *Poetics*, vol. 37, no. 5-6, pp. 490–506, 2009.

[2] P. Berliner, *Thinking in jazz*. University of Chicago Press, 1994.

[3] W. E. Timner, *Ellingtonia: The Recorded Music of Duke Ellington and His Sidemen (Studies in Jazz)*, 5th ed. Scarecrow Press, 2007.

[4] S. Albin, "BRIAN." [Online]. Available: http://www.jazzdiscography.com/Brian/index.php

[5] J. Hopcroft, O. Khan, B. Kulis, and B. Selman, "Tracking evolving communities in large linked networks," *Proc. of Nat. Acad. of Sci.*, vol. 101, no. Suppl 1, pp. 5249–5253, 2004.

[6] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature Letters*, vol. 435, no. 9, pp. 814–818, 2005.

[7] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, "A framework for community identification in dynamic social networks," in *KDD*. New York, New York, USA: ACM Press, 2007, p. 717.

[8] L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Community evolution in dynamic multi-mode networks," *KDD*, pp. 677–685, 2008.

[9] S. Bender-DeMoll and D. A. McFarland, "The art and science of dynamic network visualization," *Journal of Social Structure*, vol. 7, no. 2, 2006.

[10] M. Rosvall and C. T. Bergstrom, "Mapping Change in Large Networks," *PLoS ONE*, vol. 5, no. 1, p. 7, 2010.

[11] J. S. Yi, N. Elmqvist, and S. Lee, "TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations," *International Journal of Human-Computer Interaction*, vol. 26, no. 11, pp. 1031–1051, Nov. 2010.

[12] K.-P. Yee, D. Fisher, R. Dhamija, and M. Hearst, "Animated exploration of dynamic graphs with radial layout," in *IEEE Symposium on Information Visualization 2001 INFOVIS 2001*, vol. 2001. Ieee, 2001, pp. 43–50.

[13] U. Brandes and S. R. Corman, "Visual unrolling of network evolution and the analysis of dynamic discourse," *Information Visualization*, vol. 2, no. 1, pp. 40–50, 2003.

[14] S. Diehl and C. Görg, "Graphs, They Are Changing. Dynamic Graph Drawing for a Sequence of Graphs," in *Graph Drawing*, ser. Lecture Notes in Computer Science, M. T. Goodrich and S. G. Kobourov, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, Nov. 2002, vol. 2528, pp. 23–31–31.

[15] U. Khurana, V.-A. Nguen, H.-C. Cheng, J.-w. Ahn, X. S. Chen, and B. Shneiderman, "Visual analysis of temporal trends in social networks using edge color coding and metric timelines," in *SocialCom*, 2011, pp. 549 – 554.

[16] E. M. Bonsignore, C. Dunne, D. Rotman, M. Smith, T. Capone, D. L. Hansen, and B. Shneiderman, "First Steps to Netviz Nirvana: Evaluating Social Network Analysis with NodeXL," *2009 International Conference on Computational Science and Engineering*, pp. 332–339, 2009.

[17] J. Moody, D. McFarland, and S. Bender-DeMoll, "Dynamic Network Visualization," *American Journal of Sociology*, vol. 110, no. 4, pp. 1206–1241, 2005.

[18] S. North, U. Doğrusöz, B. Madden, and P. Madden, "Circular layout in the Graph Layout toolkit," in *Graph Drawing - Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 1997, vol. 1190, pp. 92–100–100.

[19] J. Six and I. Tollis, "A Framework for Circular Drawings of Networks," in *Graph Drawing*, ser. Lecture Notes in Computer Science, J. Kratochvíyl, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, Mar. 1999, vol. 1731, pp. 107–116–116.

[20] C. Wang, L. Shi, and Z. Wen, "Dynamic Network Visualization in 1.5D," in *PacificVis*, Hon-Kong, China, 2011, pp. 179–186.

[21] E. Gansner and Y. Koren, "Improved Circular Layouts," in *Grapg Drawing*. Karlsruhe, Germany: Springer-Verlag, 2006, pp. 386–398.

[22] P. Gleiser and L. Danon, "Community structure in jazz," *Advances in Complex Systems*, vol. 6, no. 4, p. 12, 2003.

[23] D. D. Heckathorn and J. Jeffri, "Social networks of jazz musicians," in *Social Networks of Jazz Musicians*, 2001, ch. 4, pp. 48–61.

[24] C. Pattuelli, C. Weller, and G. Szablya, "Linked Jazz: an exploratory pilot," in *International Conference on Dublin Core and Metadata Applications*, Dublin, Ireland, 2011, pp. 158–164.

[25] C. H. Smith, B. Newhouse, and A. Wiedenbein, "The Classical Music Navigator," 1999. [Online]. Available: http://people.wku.edu/charles.smith/music/index2.htm

[26] O. Bieh-Zimmert, "Last.fm Vis," 2011. [Online]. Available: http://www.nextvis.com/lastfmvis/

[27] B. Rust, *Guide to Discography*. Westport, CT: Greenwood Press, 1980.

[28] ——, *Jazz and Ragtime Records (1897 - 1942)*. Denver, CO: Mainspring Press, 2002.

[29] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.

[30] A. M. Petersen, M. Riccaboni, H. E. Stanley, and F. Pammolli, "Persistence and uncertainty in the academic career," *Proc. of Nat. Acad. of Sci.*, vol. 109, no. 14, pp. 5213–5218, 2012.

[31] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[32] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences USA*, vol. 103, no. 23, pp. 8577–8582, 2006.