MCS UIUC CS 416 Data Visualization: Narrative Data Visualization Project

1. Messaging. What is the message you are trying to communicate with the narrative visualization?

Message: This visualization is designed to help Gowalla APP understand their user activities and identify key influencers, providing valuable insights to inform business initiatives.

Data origin and data process: I am using a sample dataset processed from the original open-source dataset. This dataset represents Gowalla APP's friendship network, which includes:

- 196,591 nodes
- 950,327 edges
- 6,442,890 check-ins
- 2,273,138 triangles
- An average clustering coefficient of 0.2367, indicating a 23.67% chance that two neighbors of a node are also neighbors of each other.

I processed the original edges.csv and user_total_checkin.csv files in data/process.py, reformatted the data structure into one, and extracted a 100-sample dataset into sampled_combined_user_data.json. This dataset includes:

- User ID
- List of check-in times
- List of checked-in locations
- List of location IDs
- List of connected nodes of the user

With this sampled data, the resulting graph has an average clustering coefficient of 0.4891. This higher clustering coefficient indicates that the nodes in my force-directed graph tend to form **moderate to high-density clusters**, with nearly **48.91%** of potential connections between neighbors realized.

Using this sampled data, I aim to visualize the behaviors of the **top-connected and highly connected users.** Specifically, I will analyze:

- The number of connections each user has → help identify influential users who can be targeted for marketing campaigns or rewards benefit.
- Their app usage duration → help identify the app's stickiness and potential areas for improving user engagement and retention.
- Their check-in frequency → indicate user activity levels and help tailor marketing messages to frequent users.

- The times they check-in and stay on the app → help optimize the timing of notifications, promotions, and new feature rollouts to maximize user engagement.
- Which users are clustered by these categories → allows for personalized marketing and tailored user experiences, enhancing satisfaction and loyalty

2. **Narrative Structure.** Which structure was your narrative visualization designed to follow (martini glass, interactive slide show, or drop-down story)? How does your narrative visualization follow that structure? (All of these structures can include the opportunity to "drill-down" and explore. The difference is where that opportunity happens in the structure.)

   I use a **drill-down(drop-down) story narrative structure**. It provides a large overview of Gowalla APP's top 100 socially connected users in a force-directed graph on the main page. Additionally, there are many options allowing users to select any path to explore the data, no guidance. This creates a more free-form interface designed by me.

   For example, while exploring the main graph, users can filter by connections, durations, or frequency on the same page. Additionally, users can click on options such as "Cluster by Connection," "Cluster by Duration," or "Cluster by Frequency" to see groups of users clustered based on the selected parameter, either by resetting the main graph or on top of filtered graph. Users can freely access individual user node in the graph to view specific user check-in activities, in scene1, scene2, or scene3.

   There are multiple of ways user can manipulate the visualization. Users can interchangeably filter node(s) and then cluster them, or cluster nodes first and then apply filters. They can also view individual node check-in activities based on the filtered or clustered nodes. The graph can be reset to the original force-directed view at any time, allowing users to explore different aspects of the data flexibly.

3. **Visual Structure.** What visual structure is used for each scene? How does it ensure the viewer can understand the data and navigate the scene? How does it highlight to urge the viewer to focus on the important parts of the data in each scene? How does it help the viewer transition to other scenes, to understand how the data connects to the data in other scenes?

   Scene 1:

   - **Visual Structure:** This scene starts with text includes objective, data information, and general instruction. Below the text is the main force-directed graph occupies 80% of the page height and 80% of the screen width within the SVG container providing an immediate overview with incorporated instructions at the top. Users can drag any node, reposition it, making it easy to see relationships with other nodes. This visualization offers an intuitive and interactive way to understand the social structure.

- Understanding and **Navigation**: The force-directed layout makes it easy to see the overall Gowalla APP's user connections and density of the network. Nodes are color-coded and sized based on the number of connections, providing immediate visual cues about user connectivity.
- **Highlighting** Important Parts: The top 10 socially connected user nodes are marked with their connection number on the screen and highlighted with color and drop shadow styling. Filters for connections, duration, and frequency allow users to highlight specific subsets of the data. Hover effects and tooltips provide additional information about individual nodes, making key details accessible on demand.
- **Transition** to Other Scenes: The scene includes buttons to trigger clustering views (Scene 2) and click interactions to reveal detailed user activities (Scene 3). These transitions are visually smooth and maintain the user's focus by gradually shifting the view within the same place rather than abrupt changes.

Scene 2:

- **Visual Structure:** Clustering views displays Gowalla APP's user nodes group by three categories. This view replaces the original force directed graph, groups users by parameters (connection, duration, or frequency) occupies 80% width and height of the screen. The graph's social links are removed to focus on clustered groups.
- Understanding and **Navigation**: The clusters are visually distinct and labeled, making it easy for users to see and compare different groups. Each cluster is color-coded and spatially separated by density.
- **Highlighting** Important Parts: The clustering method emphasizes group patterns and similarities, guiding the viewer's attention to how users are grouped based on different criteria. The top ten connected users are highlighted, and users can see how these socially connected users are clustered based on connections, duration, and frequency. Total distance between each node with neighbor node are counted for each cluster and annotated.
- **Transition** to Other Scenes: Users can switch between different clustering criteria using buttons, and hovering over nodes in clusters shows a tooltip with key details. Clicking on a node transition smoothly to Scene 3, providing individual user details while maintaining the context of the cluster in Scene 3.

Scene 3:

- **Visual Structure:** The line chart displays check-in dates on the x-axis and check-in times on the y-axis with independent and continuous values, making trends and patterns in user activity over time easily discernible.
- Understanding and **Navigation:** Users can hover over **time** dots to see tooltips with detailed check-in information, including specific dates and times. This interactivity helps users gain precise insights into user behavior.

- **Highlighting** Important Parts: Key events, such as peak check-in hour(s), are annotated and highlighted in white. The tooltip displays the date and time on the mouse over the dot in the line chart. The chart updates dynamically based on the selected user node from Scene 1 or Scene 2, ensuring that users focus on relevant data.
- **Transition** to Other Scenes: The scene transition is smooth, with Scene 1 or Scene 2 adjusting to occupy 40% of the width and Scene 3 taking up 60% of the width in one screen. Clicked node user will be changed to a unique green color for users to recognize currently interacting data. Users can continue interacting with the clustering buttons from Scene 2 (when Scene 1 and Scene 3 are displayed next to each other) or Scene 1 (when Scene 2 and Scene 3 are display next to each other), allowing them to see how individual check-in activities relate to broader user graph or clusters.

4. Scenes. What are the scenes of your narrative visualization? How are the scenes ordered, and why?

Scene 1: The main force-directed graph for user nodes and user connections. It includes filter functionalities for connections, duration, and frequency.
Scene 2: Clustered node groups by connection, duration, and frequency. It includes filter functionalities for connections, duration, and frequency.
Scene 3: A check-in time and date line chart that shows each user's check-in time details.

Ordering:

All three scenes have a common instruction on the top with this narrative visualization's graph elements illustration (with node, link, and degree color), data background, and dynamic scene descriptions when scene changes.

Following the top instruction,

- **Scene 1** has a primary graph view, providing an immediate overview. Within the graph, there are three filters designed within the Scene 1 SVG container help **guide** users to interact with filter range and graph. On the right-top corner of the Scene 1 graph wrapper, it also visualizes **t**hree clustering buttons, designed to **attract** users to quickly click and switch to Scene 2. It also designed to enable users to **randomly access** any node without specific guidance to see each user node's check-in time details.
- **Scene 2** is ordered to support Scene 1 by providing more detailed group information. Users can trigger Scene 2 by clicking on the clustering buttons. It is a user-directed order designed by author (me) to help users selects the path and offer insights into user behaviors based on these clustering parameters (connections, duration, and frequency).

- **Scene 3** is triggered within Scene 1 or Scene 2 when users click on randomly accessed node. The line chart in Scene 3 with 60% width of the screen highlights the path author (me) wanted to share. The graph then adjusts to occupy 40% of the width on the left next to the line chart. This helps users identify which node they are examining and allows them to switch to another node to see different check-in trends. Users can interchangeably interact with the clustering buttons on the right-top designed path when Scene 3 is active. When a clustering button is clicked in Scene 3, the view adjusts: Scene 3 (line chart) remains on the right with 60% width, and the clustered groups (Scene 2) appear on the left with 40% width, guiding users to interact with nodes within the clusters.

Why This Order?

Scene 1 provides a general overview with multiple ways of author-designed paths interactions such as filtering, clustering, and randomly accessing user nodes. Scene 2 and Scene 3 are ordered more based on author design while still ensuring free interaction within it to understand user behaviors. This ordering helps Gowalla gain a comprehensive understanding of their users' overview, connection between them, user community, and individual activities.

5. Annotations. What template was followed for the annotations, and why that template? How are the annotations used to support the messaging? Do the annotations change within a single scene, and if so, how, and why?

For the annotations, I followed the **d3.annotationCallout** template. This template was chosen because it provides a clear and visually distinct callout that can effectively highlight message on each scene, point a node with annotation x, y, dx, and dy positions and specific node in the graph.

- **Scene 1**: Annotates the top 3 users each with the most connections, the longest duration of app usage, and the most frequent check-ins. When filtered by categories, it annotates the top 2 highly engaged users, whose check-in frequency and duration meet or exceed the 75th percentile thresholds among all users.
- **Scene 2**: The annotations explain additional information about density within each clustered group at four corners. This helps identify how users within the community group are actively connected to each other, providing insights into community structure and interactions.
- **Scene 3**: The annotations help understand the clicked user's peak check-in hour(s) by analyzing the frequency of check-ins by hour in the line chart within the graph or cluster on the left graph. This helps identify users' behavior patterns, which can be useful for business or marketing initiatives.

The annotations in my visualization maintain consistency through the use of the d3.annotationCallout template, ensuring a uniform appearance with clear callouts. These annotations are mainly gray color, but white when cluttered with the background graph, and

are positioned dynamically based on data attributes, ensuring relevance and accurate placement next to the corresponding nodes or clusters.

In scenes where static positioning is needed, such as highlighting top nodes in Scene1, the annotations remain fixed to consistently communicate key messages. Dynamic updates occur in response to filter changes in Scene 1, cluster density in Scene 2 and peak hour(s) in Scene 3 for a user node details, allowing the annotations to reflect the most pertinent data while maintaining a cohesive and visually distinct style throughout the visualization.

6. **Parameters.** What are the parameters of the narrative visualization? What are the states of the narrative visualization? How are the parameters used to define the state and each scene?

Parameters: The narrative visualization uses several key parameters to control the visualization and interaction states:

1. **Connections**: The number of connections each user has.
2. **Duration**: The total duration of app usage by each user.
3. **Frequency**: The frequency of check-ins by each user.
4. **Cluster Type**: The criteria used to cluster nodes (e.g., connections, duration, frequency).
5. **Selected Node**: The specific node (user) selected for detailed information.

States of the Narrative Visualization:

1. **Initial State**: Displays the main force-directed graph with all nodes and connections.
2. **Filtered State**: Displays a subset of nodes based on user-defined filters for connections, duration, and frequency.
3. **Clustered State**: Displays nodes grouped into clusters based on the selected clustering criteria.
4. **Detailed View State**: Displays detailed check-in activities and peak times for a selected node.

Use of Parameters to Define State and Each Scene:

Scene 1: Main Force-Directed Graph

- **Parameters**: Connections, Duration, Frequency
- **State Definition**: In the initial state, the graph shows all nodes and connections. When filters are applied, the graph transitions to the filtered state, highlighting nodes that meet the selected criteria.
- **Annotations**: Initially highlight the top three users with the most connections, longest duration, and most frequent check-ins. When filtered, annotations update to highlight the top two highly engaged users based on the new criteria.

Scene 2: Clustered Nodes

- **Parameters**: Cluster Type
- **State Definition**: Based on the selected clustering criteria (connections, duration, frequency), the graph transitions to the clustered state. Nodes are grouped into clusters to reveal patterns and relationships.
- **Annotations**: Provide additional information about the size and characteristics of each cluster.

Scene 3: Detailed Node View

- **Parameters**: Selected Node
- **State Definition**: When a node is selected, the graph transitions to the detailed view state, showing a line chart with the user's check-in times and durations.
- **Annotations**: Highlight the user's peak check-in times, dynamically updating based on the selected node.

7. **Triggers.** What are the triggers that connect user actions to changes of state in the narrative visualization? What affordances are provided to the user to communicate to them, what options are available to them in the narrative visualization?

1. **Filter Sliders**: Users can adjust sliders for connections, duration, and frequency.
   Action: Moving the sliders triggers the graph to update and highlight nodes that meet the selected criteria.
   State Change: Transitions the graph from the initial state to the filtered state, displaying only nodes within the specified ranges.
   Affordances: Labels and numeric values next to the sliders help users understand the impact of their adjustments.

2. **Cluster Buttons**: Users can click buttons to cluster nodes based on connections, duration, or frequency.
   Action: Clicking a cluster button triggers the graph to group nodes into clusters according to the selected criterion.
   State Change: Transitions the graph to the clustered state, showing nodes grouped into relevant clusters by removing links.
   Affordance: The button labeled with clustering options provide a clear indication.

3. **Node Click**: Users can click on a specific node to see detailed information.
   Action: Clicking a node triggers the display of detailed check-in activities and peak times for the selected user.
   State Change: Transitions the graph to the detailed view state, focusing on the selected node's data and showing a line chart with the user's peak check-in times.
   Affordances: The cursor changing to a pointer signals they are clickable.

4. **Reset Button**: Users can click a button to reset the graph to its initial state.
   Action: Clicking the reset button triggers the graph to revert to showing all nodes and connections.
   State Change: Transitions the graph back to the initial state.
5. **Hover Effects**: Display tooltips when hovered over nodes and line chart time dots, indicating that they are interactive elements.
   Affordances: Tooltips appearing when hovering over nodes or line chart dots provide an immediate visual cue that additional information is available.
6. **Annotations**: Initial annotations guide users by highlighting key nodes and explaining their significance.
   Visual Cues: Annotations with explanatory text and update dynamically based on user actions, providing continuous feedback, and reinforcing the available options.
7. **Instructional Text**: Text instructions or help icons are included to guide users on how to interact with the visualization.
   Action: Most of the text are static, but the scene description texts are dynamic by scenes.
   Visual Cues: Scene description title is bolded, and colored to emphasize the visual cue modifications.

By providing these triggers and affordances, the narrative visualization ensures that users can easily understand and interact with the data, seamlessly transitioning between different states to explore and uncover insights.