Automated Visualization of Public Transportation Time Schedules

Alan Joyce
Stanford University
P.O. Box 12500, Stanford, CA 94309
ajoyce@stanford.edu

ABSTRACT

We introduce a time-centric method for visualizing a public transportation network from the perspective of a single location at a given point in time. We design for the specific case of a stationary display at a transit station, optimizing our visual encodings to aid passengers in quickly and easily identifying the fastest routes to their destinations. At the same time, we develop a system that is general enough to support a variety of public transportation networks, and we evaluate the usability and effectiveness of this system across a range of such networks.

INTRODUCTION

Navigating a public transportation network can be a challenging task, especially for someone who is unfamiliar with the network's structure. This challenge is typically addressed by the creation of system maps, the publication of time schedules, and the development of web or smartphone-based trip planning applications. However, each of these methods carries its own disadvantages for the task of helping a passenger move efficiently through a network, so we use these disadvantages to motivate a new approach to a long-standing problem.

A traditional system map often provides a good overview of a network's structure, and can be distorted to prioritize readability over geographic accuracy (see Figure 1). However, the basic design of such a map still focuses primarily on depicting geographic position and ignores the temporal elements that become crucially important while trying to navigate the system. There is no guaranteed mapping between physical distance on the map and ride time, and no information is provided with regard to stop times or service periods.

Figure 1. The London Underground metro system map, designed by H.C. Beck in 1933. Dense urban routes are expanded, while sparser external routes are compacted. [3]

To alleviate this deficiency, a system map can be combined with a time schedule. This often takes the form of a large table, wherein each cell represents the time at which a single vehicle reaches a single station. Such a schedule can even be plotted visually using a Marey Graph, where diagonal lines represent the movement of vehicles over time and a route can be plotted by simply tracing a path from left to right (see Figure 2). However, the issue with time schedules is twofold: first, they include a large amount of information that is irrelevant to a given passenger at a given point in time, and second, they remove the ability to directly include any relevant geographic information.

In recent years, transportation agencies and independent developers alike have turned to smartphone applications to overcome the issues posed by traditional transit aids. These applications typically allow the user to specify their destination before calculating an optimal route and leading the user through it. This type of system works well for an individual, but transit networks need to help large numbers of passengers find their way on a daily basis, many of whom do not have smartphones or have not downloaded the necessary application.

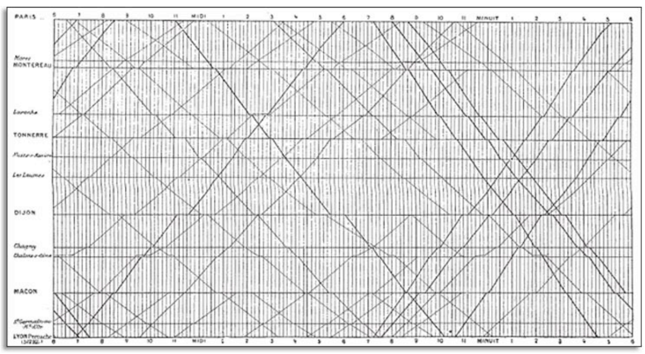

Figure 2. A Marey Graph of train service between Paris and Lyon, created by E.J. Marey in 1885. Stations are organized on the vertical axis, while time is depicted horizontally. [6]

From this situation arises a need for a visualization system that aids transportation network navigation while incorporating elements of both geography and time, only displaying information relevant to the passengers using it, and not requiring the use of any personal electronic devices. It is this sort of system which we aim to develop here, grounded in the real-world scenario of a display monitor at a transit station. For the purposes of approaching this task, we simplify the problem as follows:

- We have a passenger at location A at time T
- We want to help this passenger reach another location B as quickly as possible
- We know A and T
- We do not know B

This formulation of the problem allows us to craft a solution that functions without knowledge of the user's route, but still utilizes the contextual information that is available to it by accounting for location and time.

RELATED WORK

There has been a significant amount of work over the past century in the area of transportation visualization, from the hand-crafted creations of Beck and Marey to a host of much more recent models for algorithmically generating various types of transit maps and graphs. While it has only been in the past few years that automatically designed transportation maps have begun to approach the quality of hand-drawn ones, the need for such digitally rendered equivalents has been well-recognized for at least the past decade.

In a 2001 publication, Avelar and Huber [2] observe this need and take a logical first step toward addressing it. They reason that before we can automatically generate competent transportation maps, we require a solid, unified data model from which to work. Avelar and Huber combine the geographical and topological aspects of a transportation network into a representation that accurately captures a great deal of information about it. However, absent from their model is a fine-grained description of route and stop timing information, meaning that while such a model is valuable for depicting a geographical overview of the network, we require a more specific dataset if we are to provide the sort of timed routing information that could be found in a traditional transit timetable.

Fortunately, the state of the art in transportation network modeling has evolved over the past decade, and in 2006, a collaboration between Google and the TriMet public transportation system in Portland, Oregon [8] yielded the General Transit Feed Specification (GTFS) [4]. Expanding beyond the model proposed by Avelar and Huber, GTFS incorporates detailed temporal data, geographic context, and ride fare information. Since its introduction in 2006, GTFS has been adopted by nearly every major transportation agency in the United States, as well as many others around the world. It is now the most widely recognized format for transit data, and is, as a result, the format we choose to be the source for our visualization.

In addition to evolutions in the underlying data format, recent years have yielded a significant amount of development in generating useful computer-designed representations of transit data. Avelar's 2008 study uses transport networks in Zurich [1] as a basis for presenting a number of principles useful in the design of such representations. The study gathers a series of recommendations about lines, backgrounds, routes, stops, and labels for a variety of transportation network types. The resulting principles are a combination of simplifications ("one route per service or coincident services together") and generalizations ("generalized streets, rivers, lakes..."), and inform a pair of alternate designs for Zurich's transportation system.

While Avelar's work provides some important background in transportation visualization aesthetics, it does not yet manage to make a crucially important leap: the transition to a fully automated design process. For this, we turn to a series of recent studies, beginning with the work of Stott et al. [9]. Focusing on metro maps, the authors use multicriteria optimization to craft an ideal layout for the Mexico City and Sydney CityRail systems, along with several others. The work combines design elements from traditional paper maps with a series of optimization criteria and clustering routines to produce results that are similar in nature to the traditional maps, but are created by a fully automated process. Encouragingly, the authors are able to show that their automatically drawn maps result in faster route finding and more favorable user opinions than the equivalent traditional metro maps.

In a later study, Nöllenburg and Wolff [7] break the process of automatically rendering metro maps into two pieces: the "layout problem" (finding a good placement for map objects) and the "labeling problem" (preventing station labels from overlapping). Again, the authors draw from traditional real-world map examples, crafting a set of design rules which are applied as either hard or soft constraints and inform the automated design of metro maps. As with Stott et al., an evaluation of the resulting map designs suggests that the automated system is competitive with a hand-design process, although the authors were unable to produce effective labeled visualizations of highly complex networks (only unlabeled visualizations of these networks could be competently produced).

Finally, a recent study by Wang and Chi [10] combines an automated approach to transit map layout with the design principle of focus+context to develop a set of visual techniques that optimize for viewing on mobile devices. Unlike in other studies, this process is designed for interactive individual use, as the map layout is dependent upon the user's destination. The shortest path through the system to this destination is highlighted and the irrelevant pieces of the network are de-emphasized. The authors also manage to achieve real-time performance, while maintaining an effective layout and labeling scheme that is used to produce maps of the Stockholm, Mexico City, and Sydney metro systems.

Overall, related work in the generation of public transportation visualizations has focused heavily on adapting the desirable traits of traditional paper maps to fit within an automated generation process. The techniques associated with this process have evolved significantly, and the results have been shown repeatedly to be competitive with existing, hand-drawn alternatives. However, much of the existing work still centers around recreating or extending the paradigms dictated by paper maps. In this study, we endeavor to break from this trend by developing a new approach to visualizing the current state of a transportation system, while still grounding this approach in the basic ideals of both traditional and automated transit network visualizations.

METHODS

The system we have developed uses a continually updating radial timeline, an octilinear layout of route lines and stop labels, and a set of simple interaction mechanisms in order to display an entire transit system from the point of view of a single stop at a single point in time.

Timeline

Time is encoded as a function of radius, with points farther from the center of the graph representing points in time farther away from the current moment. A global scaling constant determines the relationship between radius and time, and is used to convert between the two when drawing features on the plot. For reference, a series of concentric circles marks intervals of 15 minutes, up to an hour after the current time, beyond which the circles are spaced increasingly far apart.

As time moves on, all items in the graph are gradually pulled into the center, where a thicker, white circle represents the current location in space and time. All points within this circle are considered to be a dead zone, wherein all time is equal to the current time (displayed in the top-left corner of the visualization).

Routes and Stops

The entirety of a transportation network's time schedule is parsed from GTFS format and broken down into a simple list of routes and their constituent stops. These routes are then filtered to include only those that are active on the current day and still accessible from the current station. After this is done, stops are removed from routes until the only stops that remain are the soonest possible arrivals at each of the other stations in the transit network. We then gather any routes that still contain useful stops and display the resulting graph, showing only the fastest methods of reaching every possible destination in the system.

Routes are only displayed if they are accessible within a cutoff horizon, calculated as the amount of time needed to reach a corner of the graph. If more routes fall within this

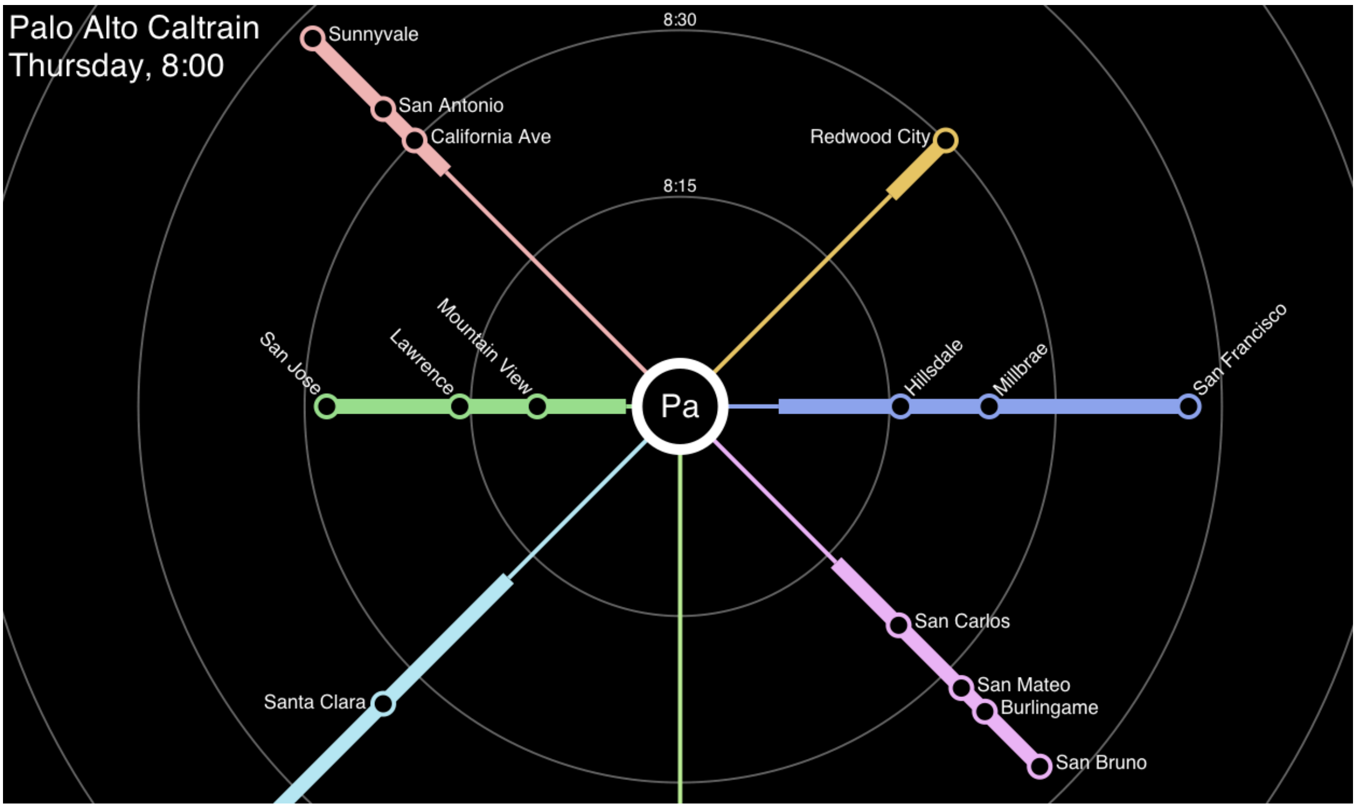

Figure 3. The Caltrain system as seen from the Palo Alto Caltrain station at 8:00 AM on a weekday.

horizon than can be cleanly plotted at once, routes with sooner access times (the time at which the vehicle arrives at the current station) are prioritized over later routes.

Before it is plotted, every route is associated with a list of location priorities. In most cases, these priorities are based on visual aesthetics, but for some systems, they can also be based on geographic information. For example, in the Caltrain graph, southbound routes prefer to be plotted on the left half of the screen while northbound routes prefer to be plotted on the right half. Routes are sorted by access time and assigned in turn to their highest priority location on the plot. If a route is already visible on the graph, it remains in its current position until it is no longer displayed. The goal is to maintain a sense of visual persistency with regard to the current set of active routes.

As with the concentric circles, routes and their stops are plotted according to a conversion between time and radius. A stop that could be reached in 30 minutes is plotted at a point timeToRadius(30 * 60 seconds) pixels from the center of the graph. Additionally, the line thickness of a route indicates whether a stretch of time would be spent waiting for (thin line) or riding (thick line) the route's vehicle. Thus, the entire system is shown in terms of what locations can be reached in a given amount of time, and encoded in such a way as to indicate how a passenger would most quickly move to any destination.

Labels

Label placement was influenced by Wang and Chi's description of an octilinear stop label model, but the complexities of the associated fitting process were reduced by also positioning all routes along fixed, octilinear angles. As a result, label orientation and position is simply a function of the associated route's orientation, with orientations chosen in a manner that attempts to minimize overlap.

However, it is not uncommon for labels along the same route to overlap with each other, a problem that is not resolved by simple orientation rules. To address this, we scan the stops in each visible route to find the smallest time

Figure 4. In a view of the BART transit system, station name labels are adjusted in size in an attempt to compensate for groups of tightly spaced stops along some routes.

gap between two stops (and thus, the shortest distance as shown on the graph). The label text is then scaled for each route based on this time gap, with upper and lower hard limits on font sizes to prevent unreadably small or unreasonably large text.

Animation and Interactivity

Because the deployment target for this visualization is a stationary display at a transit station, it is designed to continually update with a current representation of the transit system throughout the day. When a vehicle leaves the current station, the route associated with it disappears from the graph, as it is no longer accessible from the current location. As a result, all of the other stations reachable through that route are reallocated to other routes if possible, or shown on a new route if needed. The intention is for this visualization to constantly and automatically display the fastest method for reaching any point in the transit system given the current time without requiring any user input.

While a stationary billboard-like display was the primary target, it is not difficult to envision that such a device would be interactive, or that this visualization could prove useful on another platform. To account for this, we have included a set of simple interaction mechanisms that aid simple exploration of the data. Panning and zooming with a mouse are supported, and the user has the option to switch to another location at any time by clicking on a stop icon in the graph. Doing so moves the current location to the associated station but does not change the current time, allowing a user to shift their perspective on the system's current state.

RESULTS

We evaluate the visualization system's performance over two rail networks (Caltrain and Metrolink), one light rail network (BART), one bus system (SF Muni), and one ferry system (SF Bay Ferry Network). For each network, we begin by looking at a central stop at 8:00 AM on a weekday and use a debug mode to rapidly simulate the passing of several hours worth of time. While simulated time passes, we observe the system's ability to adapt to changes in the network's current state and consistently produce a view that is both readable and aesthetically sound.

Caltrain

The Caltrain system was the initial test case for prototypes of this visualizations system, giving it a natural inclination toward working well under the visual encodings presented here. For the most part, this inclination is reflected during the testing process, as Caltrain's system is suitably presented (see Figure 3). During busy commuting hours, riding the Caltrain involves navigating a complex array of express and local routes, which are shown simply and effectively by the visualization. As time moves toward the middle of the day or the late night, Caltrain's routes converge into two horizontal lines representing the next

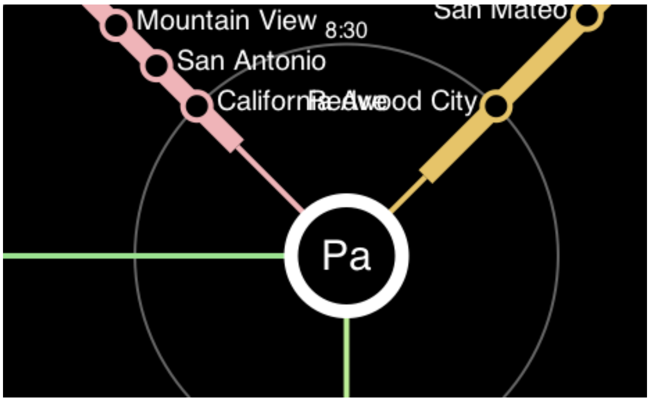

Figure 5. At one point in the weekday Caltrain schedule, two equidistant stops display overlapping labels.

arriving northbound and southbound trains. While this view does not use the available space to its fullest, it still manages to accurately present the necessary information.

At the same time, there are a handful of situations in which this view of Caltrain does not perform optimally. In Figure 5, we observe a point in time where two stop labels overlap each other in a visually unappealing and confusing manner. This sort of problem can occur, although it is relatively uncommon over the entirety of the schedule. Additionally, the opposite sort of problem can be observed after the last Caltrain vehicle of the day leaves the current station. In this case, the last plotted route disappears and the graph is left devoid of any routes or stops. While this accurately reflects the current state of the system, there is no visual reference to describe what has happened.

Metrolink

As another rail system, Los Angeles' Metrolink is displayed in a similar manner as Caltrain, although it is a network with more branching paths than the linearly-tracked Caltrain. As seen with Caltrain, Metrolink's network sometimes converges to a few long routes during the day, but overall the trend is toward many divergent routes moving passengers in different directions. The visualization is able to effectively show relevant routes throughout the day, and stops are almost always spaced far enough apart that labels become easily readable and do not overlap with each other.

One observed issue with the Metrolink visualization is the display of what are often lengthy wait times before the next train arrives. Because the visualization renders all routes on a linear scale, a long wait time for a train often means that the stops along the associated route are not initially shown unless the view is panned or zoomed, resulting in an inefficient use of space on more than one occasion throughout the day.

BART

Presenting the challenge of added complexity, the Bay Area Rapid Transit (BART) system incorporates a variety of branching light rail lines to serve a large number of stops. Aesthetically sound visualization of the BART system with encodings presented here requires the use of route-specific text rescaling, as seen in Figure 4. This process results in a readable depiction of BART that largely avoids label overlap, but does result in a few routes with significantly smaller stop labels.

Additionally, due to the sheer number of routes and stops within the BART system, as well as the relative frequency with which vehicles arrive at each station, the resulting views are more visually complex than with previously tested systems, potentially contributing to confusion and clutter. However, for the most part our visualization performs competently in depicting the state of the BART network over the course of a day, adjusting stop labels and arranging routes in such a way as to produce a simple, readable representation.

SF Muni

While rail and light rail networks produce networks that were generally aesthetically sound and readable when depicted using our system, the jump to visualizing a bus network is a significant one, so we evaluate the performance of the San Francisco Municipal Transporation Agency system (SF Muni). A network such as this incorporates significantly more routes with significantly closer stop times, meaning that source files can become much larger (over 50 MB for the SF Muni) and the network itself becomes more difficult to effectively visualize as a whole.

The initial process of loading the network into memory, which is perceptually instantaneous for all other networks tested, requires 8-10 seconds for the SF Muni system. Beyond this, we observe in Figure 6 that even a depiction of a single upcoming route can cause a staggering degree of visual clutter given the current visualization parameters. The lowest allowable font size still does not prevent label overlap and circular stop indicators blend together into an array of stacked crescent shapes. Unlike previously evaluated systems, the SF Muni transit data does not produce an aesthetically sound or efficiently readable visualization.

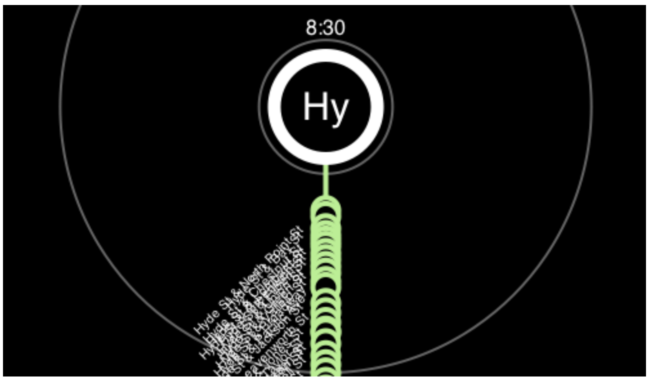

Figure 6. The Hyde St. and Beach St. station on the SF Muni displays one upcoming route with a dense sequence of stops.

SF Bay Ferry Network

The San Francisco Bay Ferry Network presents something of a counterpoint to the cluttered SF Muni system in that it is a sparse network that connects temporally distant stops via a scattering of predominately point-to-point routes. The result is a transit network that is effectively visualized by our model with a minimal amount of confusion or visual clutter. Each route has, at most, two or three stops along its path, and due to the small number of total possible destinations within the system, our visualization easily presents the shortest path to all destinations at any given point in time.

One issue with the SF Bay Ferry Network's visualization is an issue that was seen with the Metrolink system, wherein upcoming routes have long delays before the associated vehicle actually arrives, resulting in lengthy route lines that often stretch past the edges of the current view.

Summary

Overall, rail, light rail, and ferry networks are all shown to produce effective and readable visualizations when depicted using our model. Route and stop information is generally complete, understandable, and consistent throughout a day of passing time. By contrast, the SF Muni bus network induces a breakdown in aesthetic integrity and readability, representing the potential complications involved in representing more complex systems with the model presented here.

DISCUSSION

An important gauge for the usefulness of our approach is the relationship between the added complexity of introducing a new visual system for encoding transit schedule information and the simplicity gained by utilizing this new system. While a user study framed by the intended use case (a stationary display at a transit station) would be the ideal measurement of this relationship, such a study is not feasible within the scope of this project. As a result, we present here a series of informal observations about user interaction with the system that provide a general picture of its usefulness and viability.

The initial hurdle in usage of the system primarily concerns user comprehension of the temporal encoding mechanisms. It is unusual and unfamiliar for many people to view time in terms of animated concentric circles, and as a result, many users express initial confusion over the direction in which routes should be moving. However, once some amount of animation takes place over time, this representation becomes much clearer, and when paired with contextual information (e.g. "the train to San Francisco arrives in 15 minutes), the mapping of time schedule to visual display is more easily understood. Thus, there is an opportunity here for further clarification and explanation of the interface, especially if it were to be deployed in a real-world scenario. Overall, once users comprehended the encodings being used, they were receptive to and intrigued by the visual layout of the system.

Another major initial reaction gathered from users is the desire to search through the transportation schedule or plan out a specific route through the system. This is likely to some degree a result of demonstrating the visualization on a laptop rather than the less interactive sort of stationary display it is designed for, but it is a valuable observation that many users expect search and planning tools to be included with an automated transportation visualization.

Once users become acquainted with the interface, along with its capabilities and limitations, the majority are then intrigued by the unique view of a transportation system being presented to them. The interest of these users in watching and interacting with the prototype visualization signifies the potential of such a system to be an eye-catching, well-utilized visual aid if it were to be installed at a station of a real-world transportation network.

While some work could still be done to improve the visualization's immediate usability and comprehensibility, as a whole the system succeeds at rendering simple transportation networks in such a way as to both generate interest in navigating the network and ease the cognitive burden required for a passenger to visualize the temporal and spatial path to their destination.

FUTURE WORK

The design of our visualization focuses on the use case of a stationary display, but a valuable next step would be to develop a set of adaptations that would allow the same visual encodings and techniques to function efficiently on a mobile device screen. As emphasized by the work of Wang and Chi, mobile devices are an important and useful tool for the navigation of public transportation networks, but the layout methodology established here does not necessarily translate directly to the small screen of a smartphone. By specifically considering this use case, we would introduce a new set of design constraints that could inform useful modifications and extensions not previously considered.

With a consideration of mobile devices, as with a transition to any medium that has the potential for extensive user interaction, we also foresee an increase in the ability of a passenger to search through and explore the transportation system using our visualization. It is clear that users have the desire to actively explore the system, as the most frequently requested feature has been some sort of route selection or planning mechanism. While this diverges from the core problem addressed in this work, it would be a valuable extension to support user-directed exploration of transit data through a similar visual interface as was developed here.

Additionally, there are a number of improvements that can be made in the process of interpreting the underlying transit network data. Most importantly, it would be immensely valuable to detect and display transfer points and connections between express and local routes. This was not fully attempted here due to the algorithmic complexity of the associated problem, but it would be a worthy avenue of further development. At the same time, there are data points present in the GTFS specification that do not currently have an encoding within the visualization, but which could be incorporated in the future. This includes fare information, along with the details of route numbers and colors (e.g. "KX Bus" or "Blue Line"), which could aid passengers in distinguishing between arriving vehicles.

In terms of the visual interface itself, an important next step would be the inclusion of additional geographic indicators and input parameters when arranging routes and stops on the graph. While we already make use of some directional context during the layout process, the source data associates stop and route information with latitude/logitude coordinates, potentially allowing for much greater utilization of geographic properties. Given a more sophisticated layout algorithm, we could use the angles of route lines to encode the real-world spatial relationships between stops, combining the existing time-centric view with a degree of valuable geographic context.

Finally, it would be informative to experiment with a force-directed layout for routes, stops, and stop labels. The current approach is limited to displaying a relatively small number of routes at any given time, which works well for simple systems but breaks down when a transit network becomes more complex. To account for these complex cases, a force-directed layout mechanism would support an arbitrary number of route lines and stop labels, while ideally producing a suitable overall design.

CONCLUSION

The visualization system presented here is a novel and useful approach to the problem of navigating a public transportation network. By focusing on building a system-wide view from the perspective of a known point in space and time, we are able to automatically generate informative and aesthetically conscious visualizations for a variety of transit system types.

While a number of visual design considerations and potential extensions remain to be addressed, we believe the current implementation represents at least a reasonable alternative to current visualization methods, while holding the potential to produce gains in the assistance of passenger navigation across existing transportation networks. Furthermore, we believe that this work supports the viability of automatic, time-centric transit mapping systems and indicates that such systems merit further investigation and real-world deployment.

REFERENCES

1. Avelar, S. Visualizing public transport networks: an experiment in Zurich. *Journal of Maps*. 2008.
2. Avelar, S. and Huber, R. Modeling a Public Transport Network for Generation of Schematic Maps and Location Queries. *Proc. 20th Int. Cartographic Conference*. 2001.
3. Exploring distance time graphs. OpenLearn. http://labspace.open.ac.uk/mod/resource/view.php?id=376811
4. General Transit Feed Specification. Google code. http://code.google.com/transit/spec/transit_feed_specification.html
5. GoogleTransitDataFeed. Google code. http://code.google.com/p/googletransitdatafeed/
6. London Underground Map. Victoria and Albert Museum. http://www.vam.ac.uk/vastatic/microsites/1331_modernism/highlights_19.html
7. Nöllenburg, M. and Wolff, A. Drawing and Labeling High-Quality Metro Maps by Mixed-Integer Programming. *IEEE Transactions on Visualization and Computer Graphics* 17, 5. 2011.
8. Roth, M. How Google and Portland's TriMet Set the Standard for Open Transit Data. *L.A. Streets Blog*. 2010.
9. Stott, J. et al. Automatic Metro Map Layout Using Multicriteria Optimization. *IEEE Transactions on Visualization and Computer Graphics* 17, 1. 2011.
10. Wang, Y. and Chi, M. Focus+Context Metro Maps. *IEEE Transactions on Visualization and Computer Graphics 17, 12*. 2011.