Philadelphia Bikeshare Network Analysis

Craig Fox tuk88485@temple.edu

Abstract

Biking provides an environmentally-friendly, economically-affordable, physically beneficially alternative to driving. Within cities, bike shares provide quick and convenient options to access bikes. Studies have been performed analyzing these bikes including some network-based analysis on foreign bikeshares, but no network-based analysis has been done on the Philadelphia bikeshare. By developing a network of stations and rides, the bike network can be viewed and filtered. Additionally, this analysis can be used to judge Indego's decision to focus on expanding their bikeshare to South and West Philly. I found that the bike network was incredibly dense and no single route was overly popular. Day and Walkup Passholders displayed similar behaviors which included a focus on tourist locations and a larger amount of self loops. Meanwhile Month and Year Passholders had a more varied location distribution. Philadelphia's bike network matches most bikeshares' emphasis on "last mile" routes but does not have the higher amount of traffic between the residential and central business district typical in bikeshares. In light of passholder behavior, Indego's decision to expand into South and West Philadelphia without an increase in current area capacity makes sense for tourist and scenic areas but not for the central business district.

Introduction

Biking has steadily risen in popularity as a form of transportation especially in Philadelphia during the COVID pandemic [1]. It offers a number of benefits over cars including decreased congestion, decreased environmental impact, and increased physical movement. Bikeshare programs such as those located in New York City, Philadelphia, and Washington, D.C. provide a cheap and convenient way for both locals and tourists to utilize bikes for transportation. Additionally by sharing bikes they have a smaller environmental impact on the environment. The downside of bikeshares are that the bikes must be retrieved from specific stations. This leads to the need to “balance“ the network where bikes are moved to different stations to decrease filled stations supply and increase empty stations supply. Typically this is done by employees moving the bikes which leads to higher costs and lost revenue. Analysis of these bikeshare networks can identify how the bikeshares are used and where they can improve in order to increase ridership.

Bikeshare programs are a relatively new phenomena in North America. Washington D.C. was the first city in North America to launch a bikeshare program in 2008 [2]. In Philadelphia, the Indego bikeshare program began operations in 2015 [3]. At the end of 2020, Indego was awarded a 10 year contract to continue their bikeshare operations in Philadelphia. They decided to expand their operations to further sections of Philadelphia, starting with South and West Philly as seen in Figure 1 [4]. I hope to identify their rationale for this plan and identify how it will impact their network specifically what type of users it will attract and how it will impact bike location balancing.

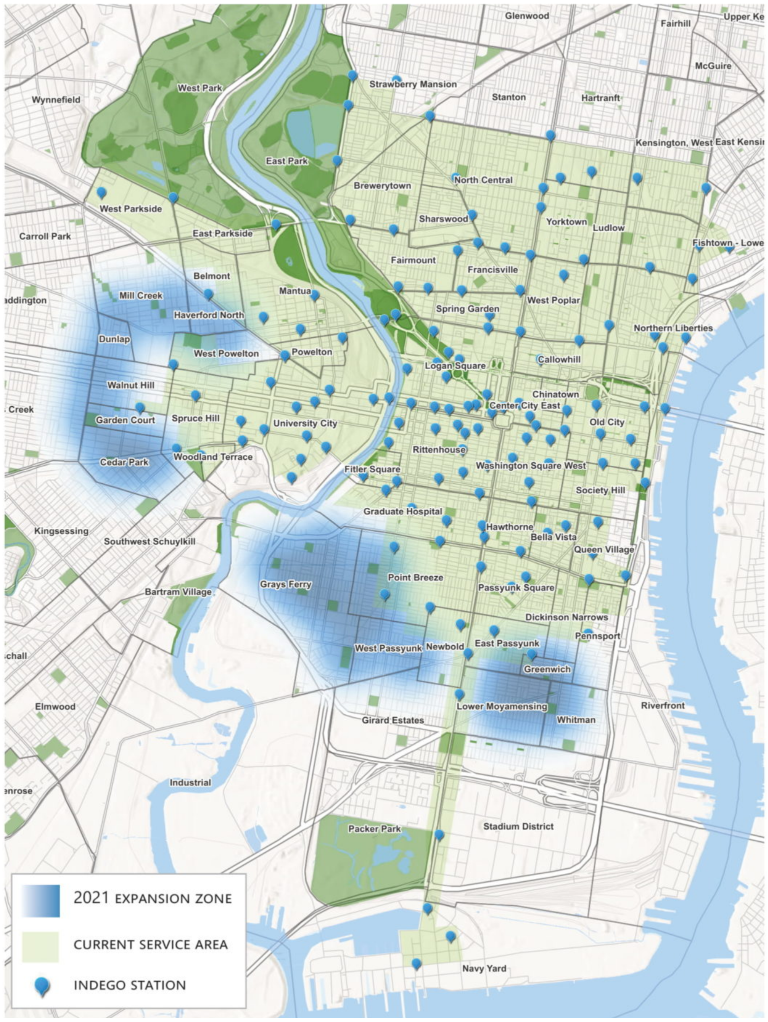

Figure 1. Indego's current service area and 2021 expansion plans

Related Work

Numerous studies have been undertaken to analyze bike sharing. Many focus on how time, weather, land use, and bike infrastructure impact ridership [5]. Analysis has also been done on these networks using predominantly visual analysis [6] or by analyzing a specific attribute of the trips . Some have even taken a network focused approach. Typically these studies either focus on clusters of station types [7], a case study of a small subset of stations [8], the flows of bikes [9], or other limited aspects of the network. The most comprehensive information and social network analysis I could find of bikeshare networks was performed on the bikeshare program in Nanjing, China [10].

Because of the comparative novelty of Philadelphia's bikeshare program, not much study has been done on it and none of these studies have focused on the network nature of its system. My work expands on previous research by applying it to Philadelphia and investigating it through a broader range of network tools on a large sample of stations than most others. Additionally I will be providing a qualitative prediction on how growth will impact balancing issues. Typically researchers behave reactively in determining how to improve balancing efforts instead of proactively in predicting when balancing issues arise [11].

Methodology

Using data provided by Indego [12], Philadelphia's bikeshare provider, I created a network composed of nodes of stations and edges of rides between these stations. The stations were cross referenced with a list of Indego Stations' geographic location to connect nodes with spatial information like Philadelphia region and proximity to other transportation methods. It is a direct network with the weight of each edge being equal to the fraction of total rides that are taken along that route. I then analyzed this network using the connected components, degree distribution, average path length, clustering coefficient, node centrality, and self loops of the graph. I generated a distribution of the net out-degree of each node by subtracting the out-degree of the node by its in-degree to find the difference. Afterwards I repeated these processes for filtered versions of the graph where only specific passholder types were used. The passholder type was used to serve as a proxy for whether the user was a local or a tourist. Based on these results, I analysed Indego's decision to expand their stations in South and West Philadelphia and predicted how their network balance would be impacted.

Results

The most basic results were exactly what was to be expected. The network is incredibly dense and decentralized. There are 144 nodes and 19353 edges. This gives the graph a density of 0.933. The directed graph was composed of only one strongly connected component. Even being a directed graph, the average shortest path length is 1.067. Unsurprising the mean clustering coefficient was 0.961. The mean combined number of in and out degrees was 268.792.

All of these distributions contained a few outliers. These outliers exist because they are geographically separated from the central cluster of stations.

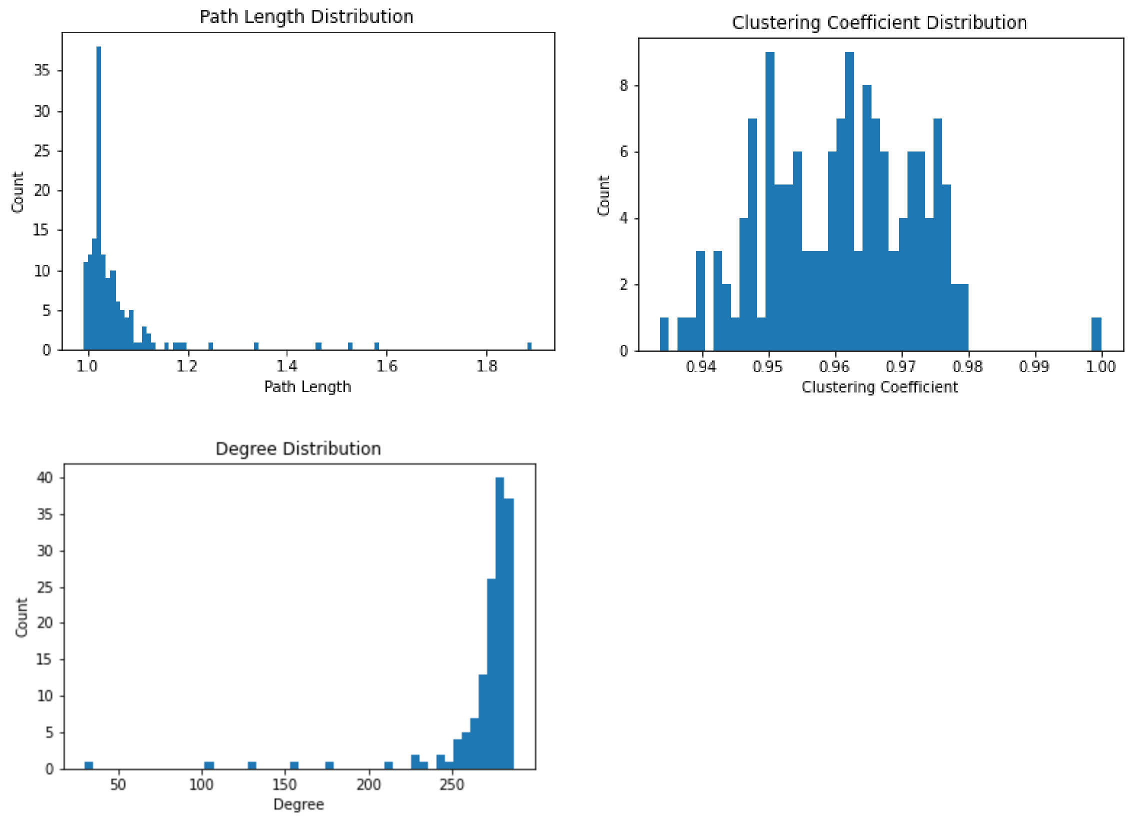

Figure 2. 2A (top left) shows the distribution of mean path length. 2B (top right) shows the distribution of mean clustering coefficients. 2C (bottom left) shows the distribution of the mean total degree

The edge weights display an extreme long-tail distribution. This shows that most of the routes are balanced in usage with only a few high weights.

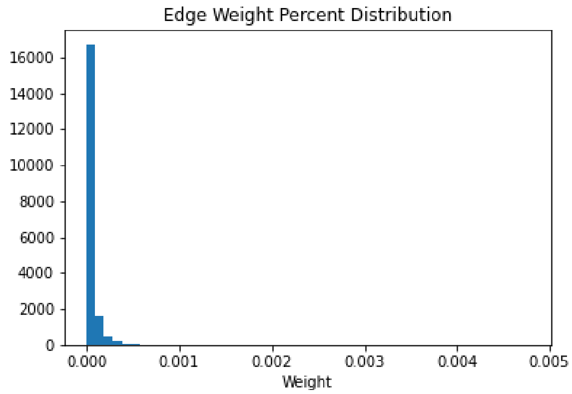

Figure 3. The distribution of the weight of each edge. The weight is the fraction of rides that edge makes up out of total number of rides

When each node's out degree was subtracted by their in degree, the Net Out Degree Percent Distribution was created. Most nodes are nearly balanced but a few display variations up to .4%. These nodes would require occasional general rebalancing but would not be a huge undertaking because they are on the scale of months and years to make a difference.

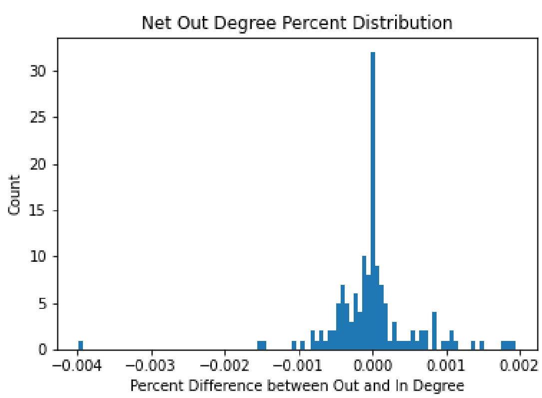

Figure 4. The distribution of net out degree. The net out degree is the out degree of the node minus the in degree of the node

While the whole network is too dense to visualize, if it is filtered to show only edges with an edge weight above a certain threshold it becomes manageable. I chose .001 and .0015 because they were large enough to have multiple edges but small enough to understand their connections visually. When using the stricter .0015 threshold, there are five routes remaining: University City to 23rd & South (0.00258), 23rd & South to

University City (0.00162), Rodin Museum to Perelman Building (0.00159), Perelman Building to Perelman Building (0.00158), and Philadelphia Museum of Art to Philadelphia Museum of Art (0.00477).

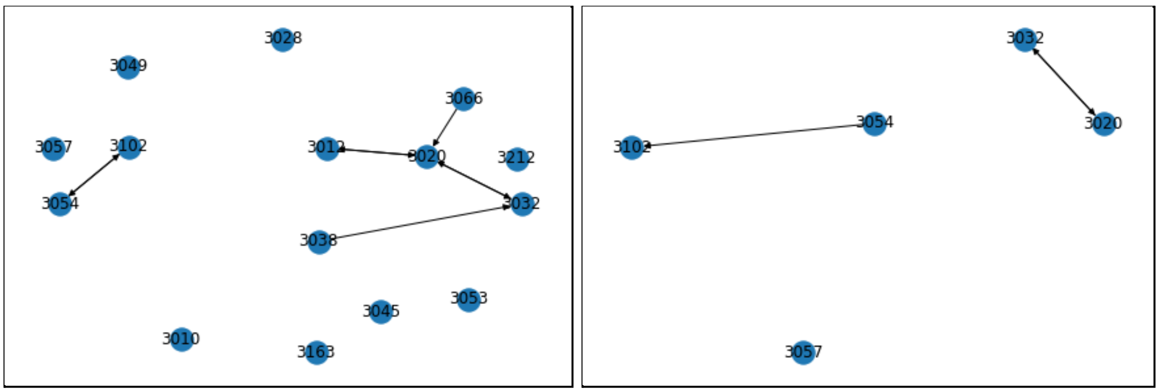

Figure 5. 5A (left) displays the bike network with only edge weights greater than .001 displayed. 5B (right) shows only edge weights greater than .0015. Nodes that appear isolated actually contain a self loop

Station Id	Station Name
University City	3020
23rd & South	3032
Rodin Museum	3054
Perelman Building	3102
Philadelphia Museum of Art	3057

Table 1. A list of common stations with their id and name

Using passholder type as a filter and the same thresholds, the results change significantly.

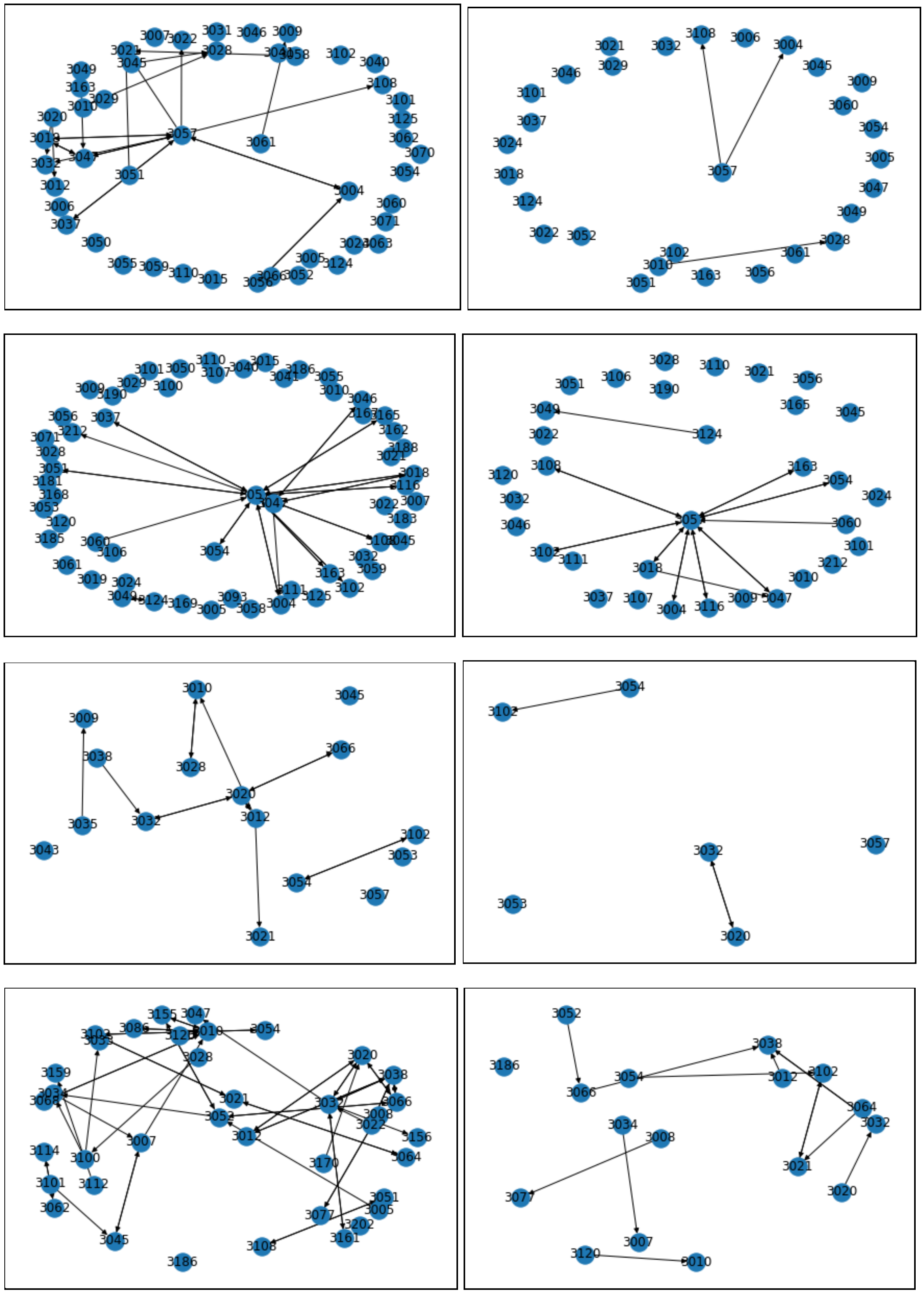
3007
3022
3031
3046
3009
3021
3028
3041
3058
3102
3040
3045
3049
3163
3108
3029
3101
3010
3020
3125
3057
3062
3019
3061
3070
3047
3032
3051
3054
3012
3004
3006
3060
3037
3071
3050
3024
3063
3005
3124
3055
3059
3110
3015
3066
3052
3056
3108
3006
3021
3032
3004
3029
3045
3046
3009
3101
3060
3037
3054
3024
3005
3057
3018
3047
3124
3049
3022
3052
3028
3102
3061
3010
3163
3056
3051
3110
3015
3101
3050
3107
3040
3186
3041
3055
3029
3009
3190
3100
3010
3046
3056
3037
3163
3165
3212
3162
3071
3188
3028
3021
3051
3018
3181
3116
3168
3057
3047
3007
3053
3022
3120
3183
3185
3060
3108
3045
3106
3054
3032
3061
3059
3163
3019
3024
3102
3093
3111
3049
3124
3169
3125
3005
3058
3004
3028
3110
3021
3106
3056
3051
3190
3165
3049
3045
3124
3022
3108
3163
3054
3120
3024
3032
3057
3060
3046
3101
3102
3111
3212
3018
3010
3037
3107
3004
3116
3009
3047
3010
3045
3009
3038
3028
3066
3020
3012
3035
3032
3102
3043
3053
3054
3057
3021
3054
3102
3032
3057
3053
3020
3155
3047
3102
3086
3010
3054
3033
3125
3020
3028
3159
3038
3034
3021
3066
3065
3032
3053
3008
3022
3012
3156
3007
3114
3100
3064
3170
3101
3112
3051
3062
3077
3005
3202
3045
3108
3161
3186
3052
3038
3186
3102
3054
3012
3066
3064
3032
3034
3008
3021
3077
3020
3007
3120
3010

Figure 6. 6A (top row left) displays the bike network of only walkup passholders with only edge weights greater than .001 displayed. 6B (top row right) shows only edge weights greater than .0015. 6C (second row left) displays day passholders with a .001 threshold. 6D (second row right) displays day passholders with a .0015 threshold. 6E (third row left) displays month passholders with a .001 threshold. 6F (third row right) displays month passholders with a .0015 threshold. 6G (bottom row left) displays year passholders with a .001 threshold. 6H (bottom row right) displays year passholders with a .0015 threshold. Nodes that appear isolated actually contain a self loop

As a whole the Day Passholders were very similar to the Walkup Passholders. For both Day and Walkup Passholders there are a larger number of edges that pass the threshold and these edges together make a larger percentage of the total amount of rides. Therefore the routes taken by Day and Walkup Passholders are more concentrated than the network as a whole. They also contain more highly weighted self loops. Popular locations for both Walkup and Day Passholders are primarily located around the Schuylkill River, Delaware River, and Fairmount Park. The Month Passholders are closest to the general graphs. Over 75% of rides are done by Month Passholders so this is to be expected. Year Passholders display the most varied interconnected graph. This suggests that Year Passholders are making the most varied trips with differing start and end stations. Both Month and Year Passholders had scattered popular locations with no clear geographic tendency.

Discussion

Philadelphia’s bike network displays some interesting similarities and differences compared to most major bike networks. Most bike networks have a large amount of traffic between residential and the central business district [9]. While Philadelphia definitely has this, it is not nearly as prevalent as seen by none of the most popular routes being between residential sections and the CBD. Instead tourism bikes in scenic sections make up a large fraction and travel inside the CBD is larger. This is perhaps due to how concentrated Philadelphia’s bike network is around the CBD and scenic sections with very few stations in the residential sections. The Philadelphia network does match most other networks' heavy concentration of “last mile” locations. These locations are the last stretch of transit a person needs to get to their destination. For example, a person might fly to a city, take the subway to a section of a city, and then need to get the last stretch to their specific location. These last stretches are a huge gap in transportation typically and bike networks have become a common method of filling this gap. Philadelphia’s bike network matches most others as seen from the popularity of the stations closest to 30th Street Station, Suburban Station, and Jefferson Station.

Indego’s decision to expand to South and West Philadelphia in the next year will have numerous effects on its bikeshare network. The most basic change will be the stations at FDR Park and the Navy Yard which currently make up the outliers in path length, clustering coefficient, and degree distribution will become more connected to the network because of more closer stations. On a whole, the clustering coefficient will decrease and the path lengths will increase because of more stations that are more spread out. Philadelphia’s bike network is dense compared to most bikeshare networks because it has a smaller number of stations in a smaller geographic region relative to its population density than most cities with bikeshares, so this change will bring it more in line with most similar cities. It will also lead to a growth in the number of Month and Year passholders. which is the primary group in the current service areas adjacent to the expansion areas. It will probably have no effect on Day and Walkup growth. It was a good decision for Indego not to expand bike stations in the scenic areas because most scenic routes are self loops, so new stations would probably not add much traffic. The expansion will most likely exacerbate balancing issues though. Month and Year Passholders are most likely commuters and expanding access to them means more bikes being located in the Central Business District during the day. Expanded capacity is needed in this area to prevent overfilled stations.

My research faced several limitations. First for privacy reasons, there is no publicly accessible data on riders. For example, are day or month pass users only buy a pass one time or are they repeat users? How many rides per user in each category is typical? This meant that my use of pass length as a proxy for local or tourist status is a rough approximation and does not reveal the whole picture of these groups behaviors. Additionally user data could be used to make a bipartite graph connecting users with stations to see how frequently different types of users use various stations.

Another limitation was the lack of data about IndeHero. IndeHero is one way Indego rebalances bikes without paying people to drive them. Users can register for IndeHero and earn points if they move a bike from a crowded station to an empty station. While this is intentional manipulation of the network by Indego, it appears like a regular ride in the data.

Because some stations are newer than others, they appear to be less frequently used. Future research could try to adjust the edges based on the amount of time those two stations have both been around. I do not think this would cause a major difference because the older stations are most likely in the more popular locations anyways and each year a progressively decreasing number of stations is added so most of stations have almost as much data.

Further research could also be comparing other bike networks and seeing what attributes suggest a successful network. Additionally temporal dynamics could be included in rebalancing analysis. Based on the results, further improvements to the Indego network could be suggested.

Acknowledgements

Thank you to Indego for their publicly available data.
Thank you to Jupyter Notebooks, Python, and NetworkX for providing the software used to perform this research.

References

[1] Tonya Russell, “Biking Was Big in Philly Even Before the Pandemic. Then the Past Year Sent it Skyrocketing”, March 15 2021, https://www.phillymag.com/be-well-philly/2021/03/15/philadelphia-biking-scene/
[2] Capital Bike Share, “About Capital Bikeshare”, https://www.capitalbikeshare.com/about
[3] City of Philadelphia Office of Transportation, “Indigo”, http://www.phillyotis.com/portfolio-item/indego/
[4] Bicycle Transit Systems, “Indego is Expanding!”, https://www.rideindego.com/blog/indego-is-expanding/
[5] Faghih-Imani, A.; Eluru, N.; El-Geneidy, A.M.; Rabbat, M.; Haq, U. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal.J. Transp. Geogr.2014,41,306–314.
[6] Oppermann, Michael & Möller, Torsten & Sedlmair, Michael. (2018). Bike Sharing Atlas: Visual Analysis of Bike-Sharing Networks. International Journal of Transportation. 6. 1-14. 10.14257/ijt.2018.6.1.01.
[7] Come, E.; Oukhellou, L. Model-based count series clustering for bike sharing system usage mining: A case study with the Velib’ system of Paris.ACM Trans. Intell. Syst. Technol.2014,5, 39–65
[8] Yao, Y.; Jiang, X.; Li, Z. Spatiotemporal characteristics of green travel: A classification study on a public bicycle system.J. Clean. Prod.2019,238, 117892
[9] Zhou, X. Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. PLoS ONE 2015, 10, 1–20
[10] Yao, Yi & Zhang, Yifang & Tian, Lixin & Zhou, Nianxing & Li, Zhilin & Wang, Minggang. (2019). Analysis of Network Structure of Urban Bike-Sharing System: A Case Study Based on Real-Time Data of a Public Bicycle System. Sustainability. 11. 5425. 10.3390/su11195425.
[11] Junming Liu, Leilei Sun, Weiwei Chen, and Hui Xiong. 2016. Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD '16*). Association for Computing Machinery, New York, NY, USA, 1005–1014. DOI:https://doi.org/10.1145/2939672.2939776

[12] Bicycle Transit Systems, “Indigo Trip Data”, https://www.rideindego.com/about/data/