An Intelligent Tree Planning Approach Using Location-based Social Networks Data

J. H. van Staalduinen[1]
LIACS – Leiden University
Leiden, The Netherlands
j.h.van.staalduinen@umail.leidenuniv.nl

J. Tetteroo[1]
LIACS – Leiden University
Leiden, The Netherlands
j.tetteroo@umail.leidenuniv.nl

D. Gawehns
LIACS – Leiden University
Leiden, The Netherlands
d.gawehns@liacs.leidenuniv.nl

M. Baratchi
LIACS – Leiden University
Leiden, The Netherlands
m.baratchi@liacs.leidenuniv.nl

***Abstract*—How do we make sure that all citizens in a city can enjoy the necessary amount of green space? While an increasing part of the world's population lives in urban areas, contact with nature remains important for the human well-being. As optional tree planting sites and resources are limited, the best site to plant must be determined. Can we locate these sites based on the popularity of nearby venues? How can we detect groups of people who tend to spend time in tree deprived areas?**

Currently, tree location sites are chosen based on criteria from spatial-visual, physical and biological, and functional categories. As these criteria do not give any insights into the amount of people benefiting from the tree placement, we propose a new criterion taking socio-cultural aspects into account. We combine the Foursquare mobility data set with a tree location data set, both of New York as a case study. Using the Foursquare data set we create a venue interaction network from which we extract venue communities. These communities are then scored based on the amount of trees in the vicinity of their venues. By combining the popularity of venues with the tree density of venue communities we can identify locations where planting a tree can benefit the most number of people and make the largest impact.

***Index Terms*—Urban computing, tree planning, social networks**

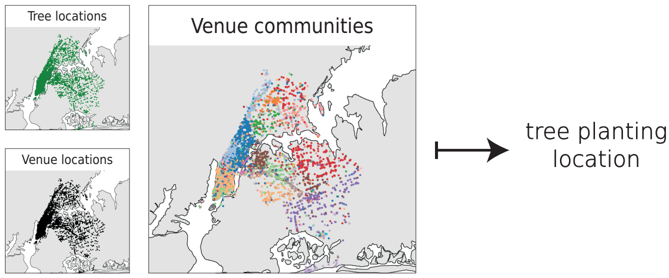


Fig. 1: We combine three types of data (tree locations, venue locations, venue communities) to determine a new criterion which can be used in selecting potential tree planting sites.

I. Introduction

As of 2018, 55% of the world's population lives in urban areas, a number which is projected to grow to 68% by 2050 [4]. The North-American continent stands out in particular, where this number is already at 82%. While it is easy to point out the economical reasons for moving to the city – at least at the first sight [7] – there are certainly downsides attached to urban life. One of them is the inescapable fact that cities, by definition [6], have a higher population density, leading to more built-up areas and thus a scarcer supply of nature than in rural areas. However, as Rohde and Kendle put it, "it is obvious from any casual observation that many human beings do not like to be dissociated from the natural world; as a nation we spend millions of pounds every year on garden and household plants" [15]. Indeed, contact with nature does seem to be linked to human well-being and positive emotional effects and is even said to strengthen urban communities [10, 13]. Apart from socio-cultural benefits, urban greenery can help to mitigate two characteristically urban problems: air pollution due to traffic [11] and (extreme) warmth due to the urban heat island effect [12]. The inclusion of parks and street trees in city landscapes is, therefore, an important aspect of the urban planning process.

To date, socio-cultural arguments play a marginal if not non-existent role in formal frameworks describing criteria for selecting potential tree planting sites. The criteria in these frameworks do not account for the amount of people that are accommodated by the newly planted trees. When following the established criteria, trees may end up in places where they are beneficial to some people, but its effects may not serve the majority of people, or may never reach the people yearning for them most.

To tackle this problem, we propose an additional tree planning criterion. Prioritization should be given to sites visited by many people and specifically people who tend to move between areas lacking trees.

We identify such locations by combining two ways of analyzing the structure of a venue interaction network. By combining the knowledge about venue popularity and venue communities with a low tree density, we can detect popular venues within tree deprived communities and thus provide a prioritization that can be used for site selection in the tree planning process, as schematically shown in Fig. 1. This prioritization can be embedded within the criteria of established tree planning frameworks that currently lack this socio-cultural value and insight.

Our paper has the following contributions:

- We describe a novel criterion for potential tree planting

[1]Both authors contributed equally to this work.

site selection based on network communities within a venue interaction network;
- We combine this criterion with venue popularity, based on network analysis of venue interaction data from the social media platform Foursquare;
- We apply this method to prioritize venues as potential tree planting sites in New York City.

II. Related work

Most of the work in the field of tree planning revolves around selecting appropriate tree species for predetermined planting sites [17, 18]. This reflects the observations by Spellerberg [18] and Pauleit [14] that tree planning is often – or at least has been for some time – an afterthought in the urban design process and characterised by pragmatism. According to an Australian survey, while visual aesthetic of trees and socio-cultural function of green spaces in the city seem to be important motives for planting trees, the first motive only plays a small role in the tree planning process [16] and the second motive is not reflected in the sparse body of site selection criteria that we could find. The work by Amir and Misgav [2], in which they aim to describe a complete tree planning decision framework, does incorporate criteria on site selection. They define three useful criterion categories, which are *spatial-visual*, *physical and biological* and *functional*. Criteria relating to the socio-cultural function of green spaces however, are missing. We observed several works describing site selection criteria [8, 14], but those fall within the category of *physical and biological* criteria that are essential for the survival of the tree. Moriani [11] did use population density in their planting priority index, but as they focused on the air pollution-reducing quality of trees, this still falls within the category of *functional* criteria. We believe then, that the body of site selection criteria is still incomplete and that we can contribute to this framework by introducing a new socio-cultural criterion which takes people movement into account.

III. Methods

A. Venue popularity

A naive approach to maximize the impact of planting a tree, is planting it near a place where many people go. To find this place we compute the degree of all nodes in the undirected network graph $G = (V, E, W)$, where nodes $v \in V$ are venues and edges $e = (v_1, v_2), e \in E$ movements of people between two venues v_1 and v_2, with weight $w_e \in W$ as the number of movements between the pair of venues. The degree of a node v is then defined as the sum of the weights of the edges that are connected to it:

$$\deg(v) = \sum_{e \in \{(u,v) | u \in \mathrm{adj}(v)\}} w_e \tag{1}$$

B. Venue community tree density

Although trees near popular venues may reach many people, they may not reach groups of people who tend to visit other venues. It may be the case that some people never come across arboreal areas. To deal with this shortcoming of the naive approach, we introduce a measure we call the *tree density coefficient*, which is based on communities in the network. A community is a group of nodes which is densely connected with each other, but much less with the rest of the network [5]. By looking at these communities, we use the fact that it is not necessarily bad for a venue not to be covered in trees, if people often move from that venue to a venue that is covered. To detect the communities, we use the Louvain community detection algorithm [3]: a fast algorithm able to find communities with high quality. It is based on the optimization of modularity, a measure that compares the density of connections within a community with the density between communities.

As it is computationally heavy to compute the modularity of a community, the Louvain algorithm uses heuristics to approximate it. Therefore, it does not necessarily return the best community layout. In order to gain confidence in the robustness of our communities we choose to run the algorithm many times to create a large number of community layouts.

To compute the tree density coefficient for a venue, we first count trees in the vicinity of the venues. We approximate this vicinity by creating a grid of the city, where each grid cell is 50 by 50 meters, calculated using Universal Transverse Mercator coordinate system [9]. Each venue v_i is mapped to a cell in the grid and is assigned the number of trees in the cell as its *venue tree density* vtd_i.

We compute the *community tree density* ctd_i for a venue v_i by averaging the vtd_i with the venue tree densities of all the other venues in its community C_i, over multiple iterations k of the community detection algorithm:

$$\mathrm{ctd}_i^k = \frac{1}{|C_i|} \sum_{v_j \in C_i} \mathrm{vtd}_j, \quad 0 < k \leq k_{\max}. \tag{2}$$

In the end, the *tree density coefficient* c_i for a venue v_i is its average community tree density value over all iterations of the community detection algorithm:

$$c_i = \frac{1}{k_{\max}} \sum_{t=1}^{k_{\max}} \mathrm{ctd}_i^k. \tag{3}$$

C. Combined method

A venue with a low tree density coefficient could have only one visitor, whereas other venues in the same community that have a similarly low coefficient could have many visitors. In this case, the latter venue(s) would be more appropriate as a tree planting site.

We extend the community based density coefficients with venue degrees by combining the two measures and detecting the set of venues that are Pareto efficient, i.e. the venues that are found by minimizing the tree density coefficient and maximizing the influence of the venue: the optimal trade-offs between the two measures. Also called the Pareto frontier, the venues in this set meet our criterion of helping most people needing trees. Tree planners could choose any of the venues along the Pareto frontier, depending on their preference towards either of the two measures.

IV. Case study

A. City of choice: New York

We conducted a case study to investigate the implementation and workings of our criterion using real data. For this we chose to focus on New York City as data on both venue interactions and tree locations was richly available.

We used two data sets to construct our criterion. We used venue interaction data of New York, provided by Foursquare as part of the Future Cities Challenge 2019, to create the venue interaction network. To assign tree density scores, we used a Street Tree Census data set [1]. This section describes the properties of both data sets and how we processed them to implement our methods.

B. Venue interaction data

The Foursquare venue interaction data set comprises of two parts: venues and movements between them. The data set contains information on ten different cities around the world. As we focused on New York in this case study, we used the New York data, but it should be noted this study is applicable to any of the other cities, provided we have access to a corresponding tree location data set.

As not all venues found in the movement data occur in the venue information data, we considered only the venues with known locations for the construction of the network. Additionally, we omitted all 86 venues not connected with the big component as the small components that are not connected never exceed a size of 3 nodes. In the end, we were able to use 15,803 venues in our analysis.

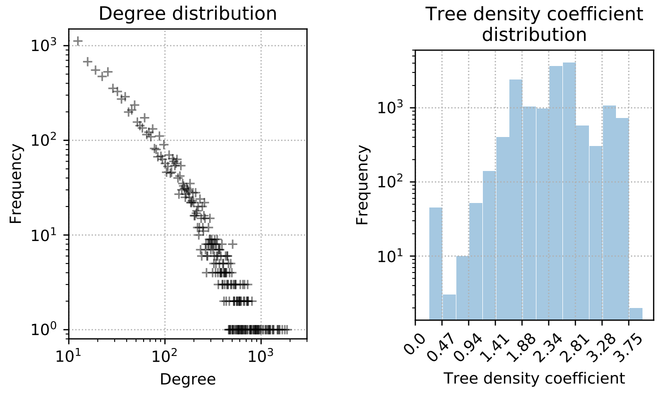
C. Street Tree Census

The Tree Census data set contains information on street trees in New York City and surrounding cities. It contains information on among others the *species*, *health*, as well as *longitude* and *latitude*. Only street trees were counted, which means that trees in parks were not taken into account and are not present in the data set.

V. Results

A. Venue popularity

We computed the venue popularity as the degree of each node and observed that the distribution follows a power law (see Fig. 2a), as is generally the case in scale-free networks modeling natural phenomena. To decide which venues would be interesting as a tree planting site according to this method, one should prioritize venues with higher degrees.



(a) (b)

Fig. 2: The power law distribution of venue degrees (a) and distribution of tree density coefficients (b).

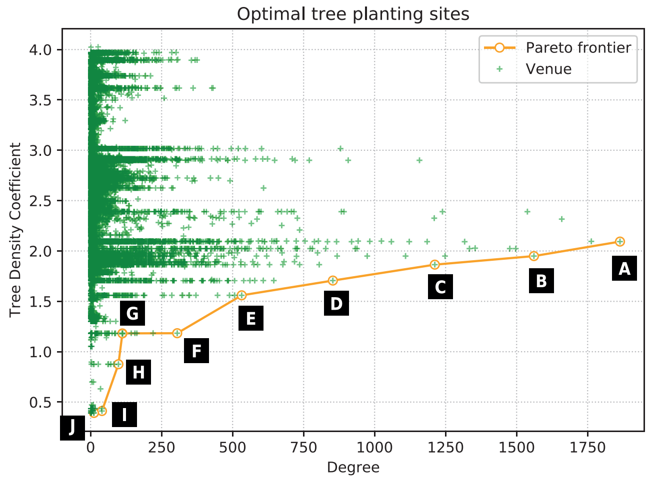
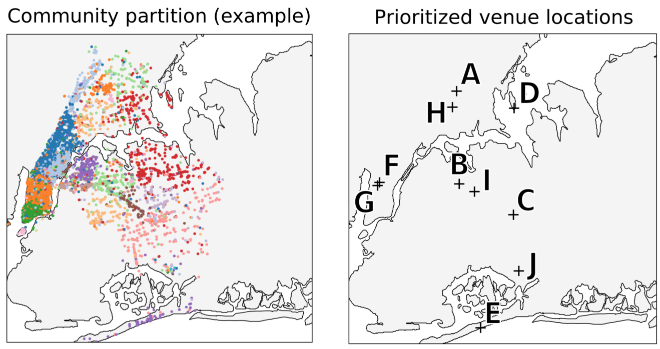


Fig. 3: The distribution of venues according to degree and tree density coefficient. The Pareto frontier shows the venues with the optimal tree planting location according to our criterion. (Venue labels correspond with Fig. 4b and Table Appendix-I)



(a) One of the 1,000 community partitions.

(b) Optimal tree planting locations (see Appendix–Table I).

Fig. 4: Map of New York City showing the optimal tree planting locations based on community structures.

B. Venue community tree density

We used the Louvain community detection algorithm as implemented in the Python `NetworkX` package. We set the resolution to 0.5 to find decently small communities. One such community lay-out is shown in Fig. 4a.

As the communities are detected using the heuristic Louvain algorithm, we averaged the community tree density of the venues over 1,000 runs of the algorithm, each time possibly detecting slightly different communities in the network, to obtain their tree density coefficients.

To find tree-deprived communities, we combined the locations of the venues within the communities with the tree locations in the street tree data set. First, we calculated the tree density for each venue. Then, the average tree density of the venues in the community was computed and returned to each of those venues as its community tree density.

We show the distribution of the tree density coefficient values in Fig. 2b. The distribution is slightly skewed to the right, which means most communities are filled with trees. Some, however, would still benefit from planting more. Prioritization for tree planting sites using this method should be given to the venues with the lowest coefficients.

C. Combined method

In order to select the most impactful planting locations, we combined both methods. This results in the distribution of venues and associated Pareto frontier as shown in Fig. 3. Here we minimize the tree density coefficient of the venues while maximizing their degree. These venues are highlighted by the Pareto frontier and should be prioritized according to our new criterion. To indicate the locations of the venues on the Pareto frontier, we show the venues on a map in Fig. 4b and provide additional insights in the data in Table I in the Appendix.

It is interesting to see that one of the selected venues (venue H) is a rose garden, amidst a park lush with trees. This is explained by the fact that the tree data set contains only street trees, and not trees in parks. Additionally, we found upon inspection using Google Street View that some of them (most notably venues A, B, D, G and H) do seem to be near a number of trees. When inspecting these locations in the tree data base,[2] we see that there are either only a few (venues B and G) or no trees (venues A, D, E and H) recorded in the immediate vicinity of the venues. We see that along with park trees, trees on private grounds are also not recorded.

VI. Conclusion

In this paper, we propose a novel criterion that can be used when selecting potential tree planting sites. The nature of the criterion is socio-cultural, capturing people movement along venues and tree-lacking (social) communities into one measure. Having implemented the measure for a case study on New York City, we show that the measure is applicable in the field and can be used to support decision makers by providing them with optional planting sites along a Pareto frontier.

We do see however, that some venues indicated by our criterion as tree lacking seem to actually be in a green area. We believe that the application of our method can be improved with a more detailed tree location data set. Then, the criterion proposed in this paper can be a meaningful addition to the established site selection criteria.

[2]The tree database can be explored on a map at https://tree-map.nycgovparks.org/, last visited 21 May 2019.

Acknowledgments

We thank Foursquare and the organisation of Netmob for organising the Future Cities Challenge and providing us with access to Foursquare's venue interaction data set.

References

[1] NYC Parks Recreation - TreesCount! 2015 Street Tree Census. https://data.cityofnewyork.us/Environment/2015-Street-Tree-Census-Tree-Data/uvpi-gqnh. Accessed: 15 May 2019.

[2] S Amir and A Misgav. A framework for street tree planning in urban areas in israel. *Landscape and Urban Planning*, 19(3):203–212, 1990.

[3] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008, 04 2008.

[4] UN DESA. World urbanization prospects, 2018.

[5] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[6] Susan A Hall, Jay S Kaufman, and Thomas C Ricketts. Defining urban and rural areas in us epidemiologic studies. *Journal of urban health*, 83(2):162–175, 2006.

[7] John R Harris and Michael P Todaro. Migration, unemployment and development: a two-sector analysis. *The American economic review*, pages 126–142, 1970.

[8] Chi Yung Jim. A planning strategy to augment the diversity and biomass of roadside trees in urban hong kong. *Landscape and Urban Planning*, 44(1):13–32, 1999.

[9] Richard B Langley. The utm grid system. *GPS world*, 9(2):46–50, 1998.

[10] Cecily Maller, Mardie Townsend, Anita Pryor, Peter Brown, and Lawrence St Leger. Healthy nature healthy people:contact with natureas an upstream health promotion intervention for populations. *Health promotion international*, 21(1):45–54, 2006.

[11] Arianna Morani, David J Nowak, Satoshi Hirabayashi, and Carlo Calfapietra. How to select the best tree planting locations to enhance air pollution removal in the milliontreesnyc initiative. *Environmental Pollution*, 159(5):1040–1047, 2011.

[12] Briony A Norton, Andrew M Coutts, Stephen J Livesley, Richard J Harris, Annie M Hunter, and Nicholas SG Williams. Planning for cooler cities: A framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes. *Landscape and Urban Planning*, 134:127–138, 2015.

[13] William LI Parry-Jones. Natural landscape, psychological well-being and mental health. *Landscape research*, 15(2):7–11, 1990.

[14] Stephan Pauleit. Urban street tree plantings: identifying the key requirements. In *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, volume 156, pages 43–50. Thomas Telford Ltd, 2003.

[15] CLE Rohde and AD Kendle. Human well-being, natural landscapes and wildlife in urban areas a review. 1994.

[16] Sudipto Roy, Aidan Davison, and Johan Östberg. Pragmatic factors outweigh ecosystem service goals in street tree selection and planting in south-east queensland cities. *Urban Forestry & Urban Greening*, 21:166–174, 2017.

[17] Henrik Sjöman and Anders Busse Nielsen. Selecting trees for urban paved sites in scandinavia. *Urban Forestry & Urban Greening*, 9(4):281–293, 2010.

[18] Ian F Spellerberg and David R Given. Trees in urban and city environments: a review of the selection criteria with particular reference to nature conservation in new zealand cities. *Landscape review*, 12(2):19–31, 2008.

Appendix

TABLE I: The venues that are, according to the Pareto analysis, the most efficient to place trees next to.

	Degree	**Tree density coefficient**	**Venue ID**	**Venue Name**	**Latitude**	**Longitude**	**Venue Category**
A	1864	2.09323399	4b637f59f964a5207b7e2ae3	MTA Subway - West Farms Square/E Tremont Av (2/5)	40.8402	-73.8800	Metro Stations
B	1561	1.94936904	4f940fe7e4b059d7da88be53	Junction Blvd	40.7491	-73.8694	Miscellaneous Shops
C	1212	1.86431926	4e7647cffa76059701632021	MTA Subway - 179th St (F)	40.7125	-73.7846	Metro Stations
D	853	1.70625431	4bace08af964a520cf143be3	Sammy's Fish Box Restaurant	40.8390	-73.7836	Seafood Restaurants
E	532	1.55978734	4cc86db294e1a0933e6c978b	Rockaway Beach - 116th Street	40.5779	-73.8359	Beaches
F	305	1.18353191	4abcfe4bf964a520fa8720e3	Hulu Theater	40.7509	-73.9941	Music Venues
G	112	1.18192331	4c516433d2a7c9b6c4c61911	Bean & Bean Organic Coffee	40.7509	-73.9941	Coffee Shops
H	98	0.87556379	4debdb6b52b11677f060802e	Peggy Rockefeller Rose Garden	40.8592	-73.8735	Gardens
I	40	0.41121454	4d93a4489ef2721e6bffc3d2	I-495 / Grand Central Parkway Interchange	40.7400	-73.8455	Intersections
J	12	0.39198899	4e26fd0f1f6eb1ae139ad929	TSA Security Screening	40.6457	-73.7762	General Travel