Supplementary: Leveraging Network Representation Learning and Community Detection for Analyzing the Activity Profiles of Adolescents

Full list of author information is available at the end of the article

Map of Columbus, Ohio

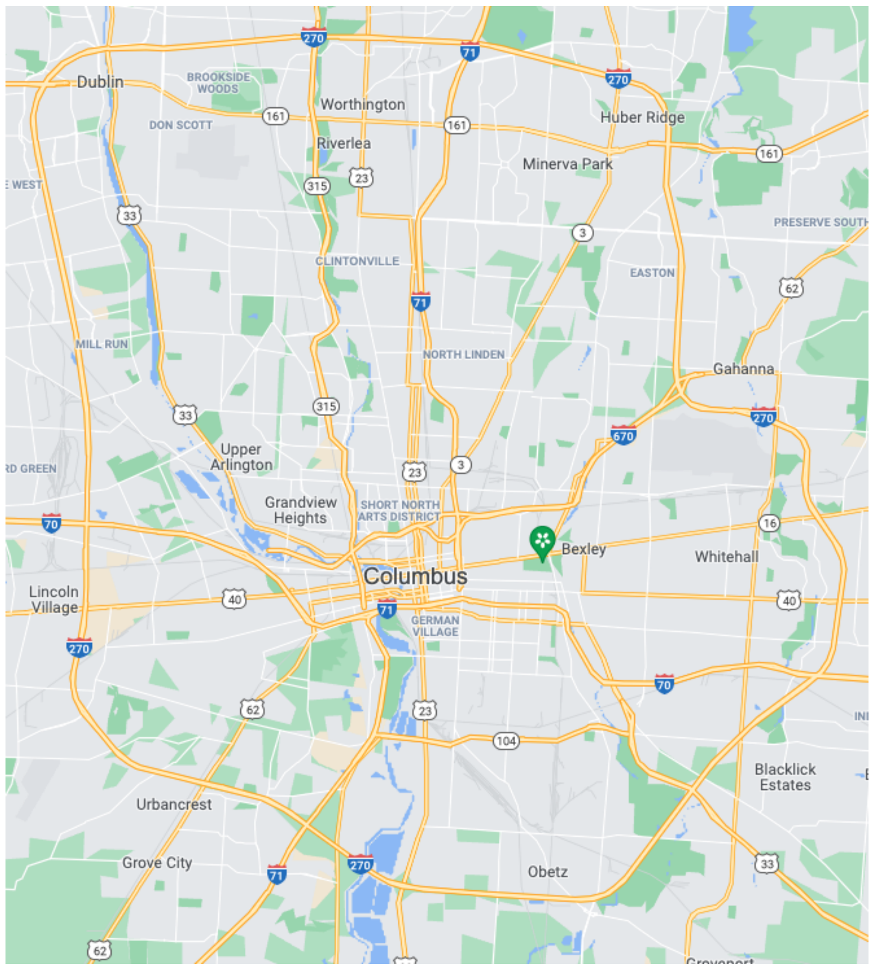

Figure 1: Columbus Map. Snapshot taken from Google Maps.

1 Methods summary

Deepwalk

The popular and competitive [1] method Deepwalk [2] performs truncated random walks on the graph. The method accepts parameters: walk length of the random walk

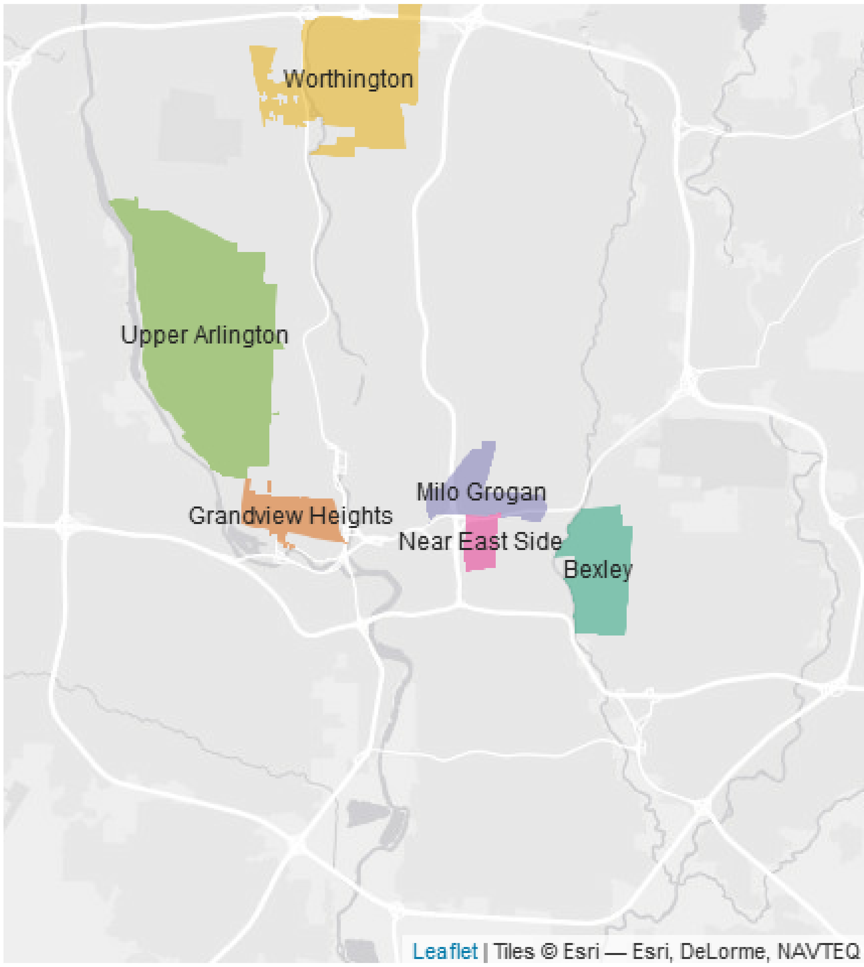

Figure 2: Frequently mentioned regions.

and number of random walks from each node to perform truncated random walks. Deepwalk then identifies the source and context nodes from the computed random walks and learns the node embeddings using the skip-gram objective function [3]. The objective function minimizes the following negative log-likelihood function

$$minimize_{\phi} \quad -\log Pr(\{v_{i-w}, \cdots, v_{i+w}\} \backslash v_i \mid \phi(v_i)) \tag{1}$$

where Pr is the probability function, ϕ represents the embedding/representation of the node, v_i is the source node and v_{i-k} is the context node and w is the size of the context window.

The computation of the probability term in Equation 1 is not tractable due to the normalization factor being computationally expensive. Hence, Deepwalk factorizes the above conditional probability using Hierarchical Softmax [4]. Hierarchical Softmax constructs a binary tree where the leaf nodes are the nodes of the graph (denoted by V), and the probability of observing a context node v_j given the representation of source node $\phi(v_i)$ is computed by the probability of reaching the node

v_j from the root of the binary tree given $\phi(v_i)$. Mathematically,

$$Pr(\{v_j|\ \phi(v_i)) = \prod_{l=1}^{*log\ |V|} Pr(b_l|\phi(v_i)) \tag{2}$$

and

$$Pr(b_l|\phi(v_i) = 1/(1 + \exp^{-\phi(v_i)\times\Psi(b_l)}) \tag{3}$$

where b_l is the node on the path from the root of the binary tree to v_j and $\Psi(b_l)$ is the representation of node b_l.

LINE

LINE [5] learns node representations of the networks by preserving the first-order and second-order proximity of the nodes. The first-order proximity constraint ensures that directly connected nodes' representation is close to each other in the embedding space. The second-order proximity constraint ensures that nodes with similar contexts (or neighbors) have similar representations.

Formally, the first-order proximity is defined as follows

$$O_1 = -\sum_{(u_i,l_j)\in E} w_{(u_i,l_j)}\ log\hat{P}(u_i,l_j)\ \text{ where } \hat{P}(u_i,l_j) = \frac{1}{1 + exp(-\phi(u_i)\phi(l_j))} \tag{4}$$

The second-order proximity is defined as follows

$$O_2 = -\sum_{(u_i,l_j)\in E} w_{(u_i,l_j)}\ log\ \frac{exp(\phi(u_i).\psi(l_j))}{\sum_{l'_j\in V,\ l'_j\neq l_j} exp(\phi(u_i).\psi(l'_j))} \tag{5}$$

where E represents edges in the co-location network. $\phi(u)$ and $\psi(u)$ are the node and context embeddings of node u, respectively. The objectives O_1 and O_2 are optimized with negative sampling approach (Equation 14) through edge sampling.

BiNE

The constructed co-location network is essentially a bipartite network with two types of nodes (users and locations). To better capture the properties of the bipartite network in the node representations, Gao et al.[6] proposed BiNE. It performs biased random walks designed for bipartite networks to obtain vertex sequences and then applies the proposed optimization framework on those vertex sequences.

The biased random-walk procedure can be summarized as follows: The random walk starts from a node u and moves to one of the nodes u's neighbors at each step. This random-walk process is stopped with probability p at each step. For each node, the random-walk process is repeated n number of times where n is dependent on the importance of node u. BiNE sets the importance of the node equal to the authority score computed from the Hubs and Authorities Algorithm [7].

Given a bipartite network, BiNE computes two homogeneous networks – for co-location network the homogeneous networks would be individual-individual network and location-location network. The biased random walk is then independently performed on those two homogeneous networks. These random walks are then treated as sequences and the skip-gram model is applied on the target and context node pairs from these sequences. Mathematically, these two objective functions can be summarized as follows.

$$O_1 = minimize_{\phi} \; -\log Pr(\{u_{i-w}, \cdots, u_{i+w}\} \backslash u_i \mid \phi(u_i)) \tag{6}$$

and

$$O_2 = minimize_{\phi} \; -\log Pr(\{l_{j-w}, \cdots, l_{j+w}\} \backslash l_j \mid \phi(l_j)) \tag{7}$$

where u_i and l_i are the target individual and location nodes, respectively. The objective functions O_1 and O_2 are intractable, and hence negative sampling (Equation 14) is utilized where negative samples are taken from the corresponding homogeneous networks. An additional constraint is imposed on individual embedding, $\phi(u)$, and location embedding, $\phi(l)$, which can be summarized as follows

$$O_3 = minimize \sum_{(u_i,l_j)\in E} -w_{(u_i,l_j)} \; log\hat{P}(u_i,l_j) \;\; \text{where } \hat{P}(u_i,l_j) = \frac{1}{1+exp(-\phi(u_i)\phi(l_j))} \tag{8}$$

where E represents the set of edges in the co-location graph. The above objective function minimizes the KL-divergence between the empirical distribution of the co-occurring probability between vertices and the reconstructed distribution using node representations. BiNE performs joint optimization of the following objective functions

$$minimize \; L = \alpha O_1 + \beta O_2 + \gamma O_3 \tag{9}$$

where α, β, and γ are the hyper-parameters.

We next briefly review the work of Xi et al. [8]. Using the coarse-grained co-location network data from the AHDC Study, they adopt and leverage Latent Dirichlet Allocation [9] (LDA) to identify activity-space pattern profiles and an individual's community affiliation within the Columbus Metropolitan area. Xi et al. [8] provide initial evidence that such community structure offers meaningful dimensions of neighborhood functioning with respect to socioeconomic and racial composition.

The model can be summarized as follows: let $Y_{n_i,j}$ be the indicator function that the individual i $(1 \leq i \leq I)$ visits location j $(1 \leq j \leq J)$ where I and J are total number of individuals and total number of locations.

The LDA model assumes

$$Y_{n_i,j}|p_{n_i,j} \overset{iid}{\sim} Bernoulli(p_{n_i,j}) \tag{10}$$

where $p_{n_i,j} = W_{i.} \times H_{k.}$. Here, $W \in R^{I*K}$ refers to the community assignment matrix and $H \in R^{K*J}$ is the row stochastic matrix representing the probability of individual from community k visiting location j.

The following priors on W and H are as follows

$$W_{i.} \overset{iid}{\sim} Dirichlet(\alpha) \tag{11}$$

and

$$H_{k.} \overset{iid}{\sim} Dirichlet(\beta) \tag{12}$$

where α and β are known constant vectors of dimension K and J, respectively. Then, the LDA model identifies the parameters (W, H) such that the following probability is maximized

$$P(Y_{n_i,j} = 1) = \sum_{k=1}^{K} P(c_i = k) \times P(l_{n_i} = j|c_i = k) = \sum_{k=1}^{K} W_{ik} \times H_{jk} \tag{13}$$

where c_i is community of individual i and l_{n_i} be the n^{th} location visited by individual i. The parameters are estimated with Gibbs sampling method. LDA can identify the latent communities (topics) representation present in the co-location network through community assignment vector – a discrete probability distribution over communities – for the adolescents.

LocationTrails:

Like the work of Xi et. al. [8], LocationTrails [10] is designed to operate on co-location networks. However, it also takes inspiration from some of the neural models we have presented above. Instead of performing random walks on the co-location graphs, LocationTrails constructs sequences from the adolescents' actual location visits and then learns the node representations from these sequences. The constructed sequences can be abstractly viewed as performing constrained walks on the co-location network. LocationTrails then extract source and context nodes and apply the skip-gram model to learn the node embeddings. The intractable normalization factor in the skip-gram model (Equation 1) is approximated with the negative sampling in LocationTrails. Mathematically,

$$\begin{aligned} log\ Pr(v_j|\phi(v_i)) = log\ \sigma(\phi(v_j).\phi(v_i)) + \\ \sum_{p=1}^{m} E_{v_k \sim P_v(v_i)}[\ log\ \sigma(-\phi(v_k).\phi(v_i))\ ] \end{aligned} \tag{14}$$

	Deepwalk	LINE	LocationTrails	LDA	Metis	Graclus
Deepwalk	1.00	0.45	0.35	0.15	0.41	0.44
LINE	0.45	1.00	0.37	0.14	0.35	0.35
LocationTrails	0.35	0.37	1.00	0.14	0.28	0.30
LDA	0.15	0.14	0.14	1.00	0.13	0.13
Metis	0.41	0.35	0.28	0.13	1.00	0.39
Graclus	0.44	0.35	0.30	0.13	0.39	1.00

Table 1: Normalized Mutual Information between clusters.

where m is the number of negative samples, σ is the sigmoid function, v_k is the negative sampled node, and $P_v(v_i)$ is the noise distribution of all the nodes. The noise distribution $P_v(v_i)$ is set as the unigram distribution of all the nodes raised to power 0.75.

Experiments on real-world human mobility (traffic) datasets show that LocationTrails outperforms existing network representation learning methods (including Deepwalk) in terms of learned node quality, running time, and memory consumption. Moreover, given the localized nature of its random walks, LocationTrails can be trained locally on edge devices using federated learning. Therefore, it can mitigate user's privacy concerns.

Quantitative analysis

The Tables 1 shows the Normalized Mutual Information (NMI) [11] between the clusters computed by the methods, respectively. NMI quantify the amount of overlap between clusters. We observe that both metrics have high overlap values between Deepwalk, LINE, Metis and Graclus. However, the overlap between LocationTrails and the rest of the methods is relatively lower.

Cluster Analysis: BiNE

From Figure 3, we observe that BiNE fails to identify meaningful communities for the bipartite co-location network. There are two rationales for this poor performance: existence of common hubs and low authority scores. Recall that BiNE constructs two homogeneous networks (individual-individual network and location-location network) from the individual-location co-location network. An edge between two individuals exists if they visit common locations. However, the human-mobility co-location network consists of few hubs such as malls, schools, and recreation centers that are visited by multiple individuals. As a result, there exist a lot of noisy individual-individual edges. This can be observed by the relatively high density of individual-individual network (density = 2.4%), while the density of co-location network is 0.04%. Note that the density of the network is $2m/n(n-1)$ where m and n are the number of edges and the number of nodes, respectively. The second rationale for BiNE's performance is due to low authority scores. Recall that the number of random walks per node is dependent on the authority score of the node. We observe that for the human-mobility co-location network, these authority scores are too low (mean authority for users and locations are 0.001 and 0.0003, respectively). The low authority scores result in reducing the number of random walks per node, thereby affecting the quality of the node representations. We also performed an experiment where we increased minimum number of random-walks

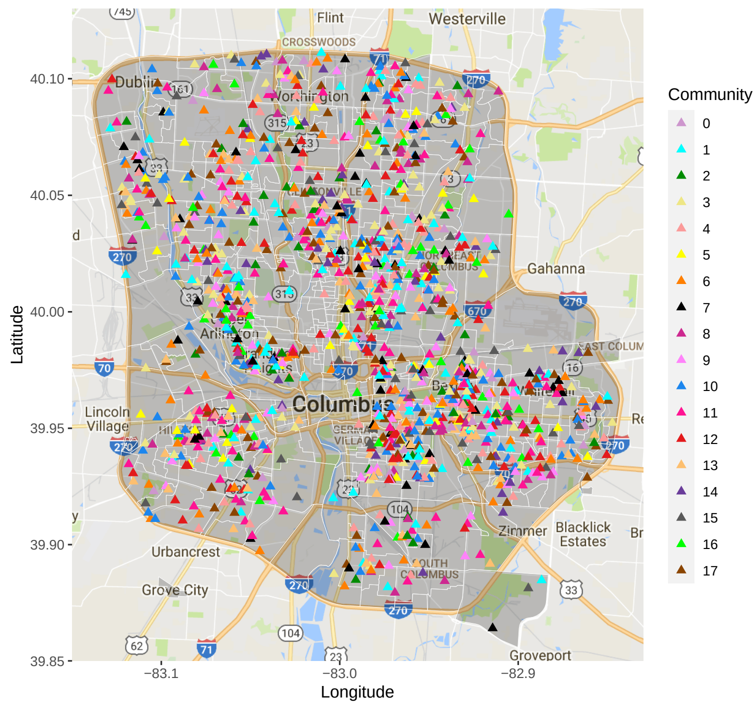

Figure 3: BiNE: Number of clusters = 18. (Home location anonymized)

per node parameter for BiNE, but the learned representations were still of poor quality due to noisy individual-individual edges.

Modularity based community detection

In Figure 4, we present the clusters identified by algorithm presented in Clauset et al. [12]. We use R library 'condor' to compute the communities. From Figure 4, we observe that the algorithm proposed by Clauset et al. [12] identifies residentially proximate clusters in both white-dominated and black-dominated neighborhoods. The NMI between clusters identified by algorithm presented in Clauset et al. [12] and other methods on white-dominated neighborhoods are Deepwalk (0.22), LINE (0.19), LocationTrails (0.19), LDA (0.16), Metis (0.19) and Graclus(0.17). The NMI between clusters identified by algorithm presented in Clauset et al. [12] and other methods on black-dominated neighborhoods are Deepwalk (0.15), LINE (0.16), LocationTrails (0.13), LDA (0.17), Metis (0.17) and Graclus(0.14).

Cluster Analysis: Hyper-parameter results

In this section, we present the adolescent clusters identified by the selected methods with different hyper-parameter settings. Overall, we observe the identified clusters to be stable with minor variations. The mobility pattern-related inferences (such as the presence of residentially proximate clusters in white-dominated neighborhoods) drawn for methods are consistent across hyper-parameters. Note that, due to the limitation of space, we are presenting only a few hyper-parameter visualizations.

Author details

References

1. Gurukar, S., Vijayan, P., Srinivasan, A., Bajaj, G., Cai, C., Keymanesh, M., Kumar, S., Maneriker, P., Mitra, A., Patel, V., et al.: Network representation learning: Consolidation and renewed bearing. arXiv preprint

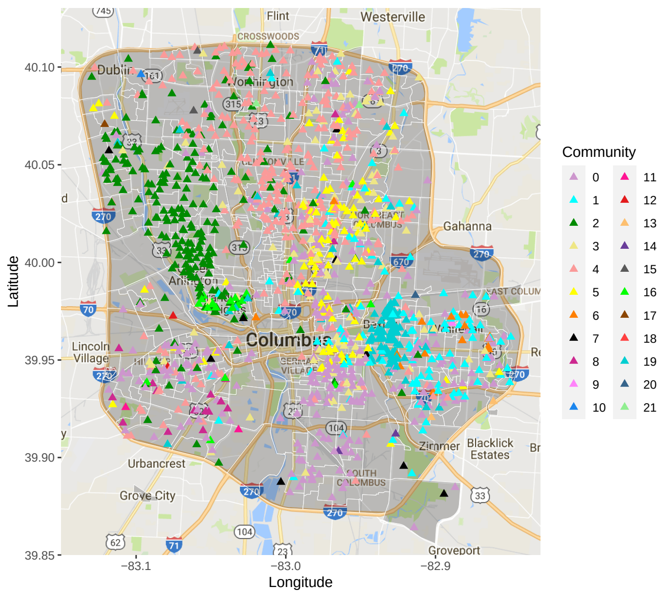

Figure 4: Clauset et al. [12]

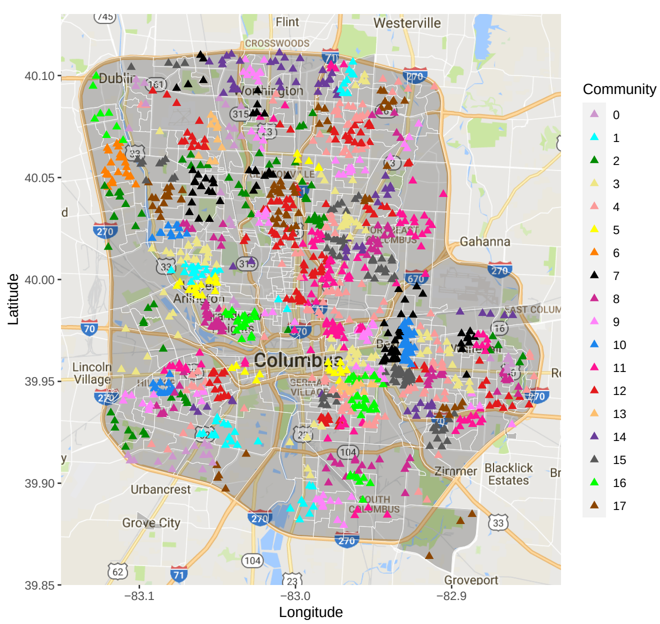

Figure 5: Deepwalk: walk length=5, number of walks=80, context window=5

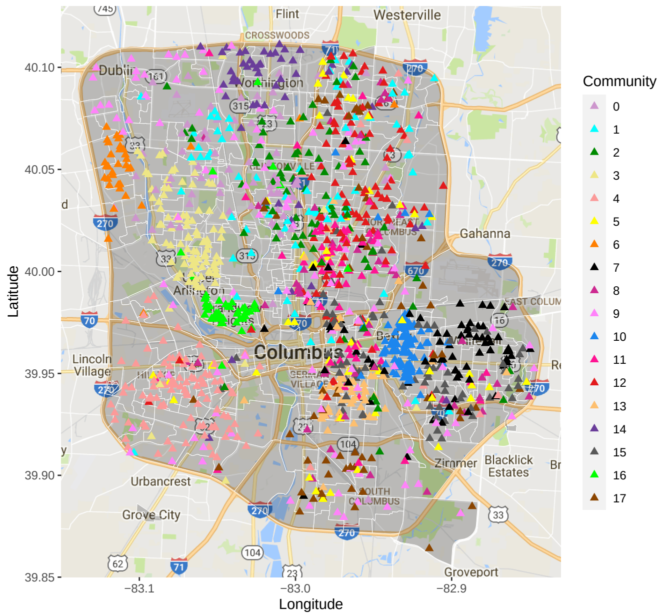

Figure 6: Deepwalk: walk length=10, number of walks=80, context window=10

arXiv:1905.00987 (May, 2019)

2. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710 (2014)
3. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. NeurIPS'13 (2013)
4. Mnih, A., Hinton, G.E.: A scalable hierarchical distributed language model. Advances in neural information processing systems **21**, 1081–1088 (2008)
5. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1067–1077 (2015)
6. Gao, M., Chen, L., He, X., Zhou, A.: Bine: Bipartite network embedding. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 715–724 (2018)
7. Kleinberg, J.M., Newman, M., Barabási, A.-L., Watts, D.J.: Authoritative Sources in a Hyperlinked Environment. Princeton University Press, ??? (2011)
8. Xi, W., Calder, C.A., Browning, C.R.: Beyond activity space: Detecting communities in ecological networks. Annals of the American Association of Geographers **110**(6), 1787–1806 (2020)
9. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
10. Gurukar, S., Parthasarathy, S., Ramnath, R., Calder, C., Moosavi, S.: Locationtrails: A federated approach to learning location embeddings. In: 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (To Appear) (2021). IEEE
11. Estévez, P.A., Tesmer, M., Perez, C.A., Zurada, J.M.: Normalized mutual information feature selection. IEEE Transactions on neural networks **20**(2), 189–201 (2009)
12. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. Physical review E **70**(6), 066111 (2004)

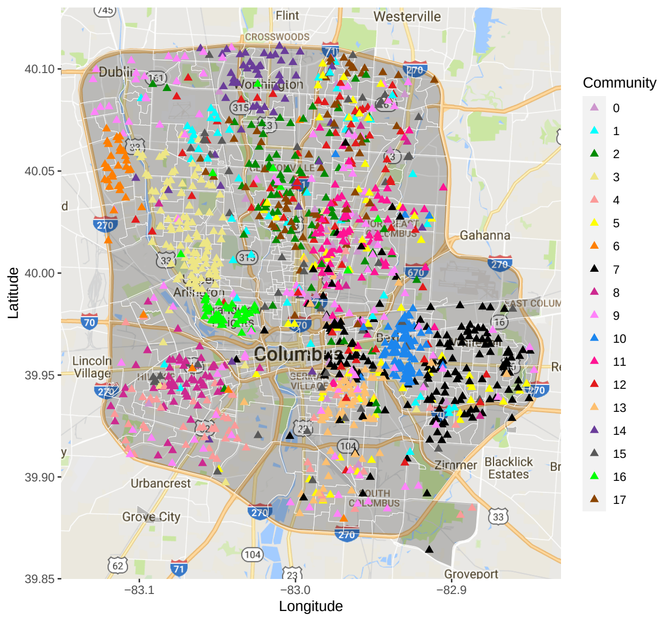

Figure 7: Deepwalk: walk length=20, number of walks=40, context window=3

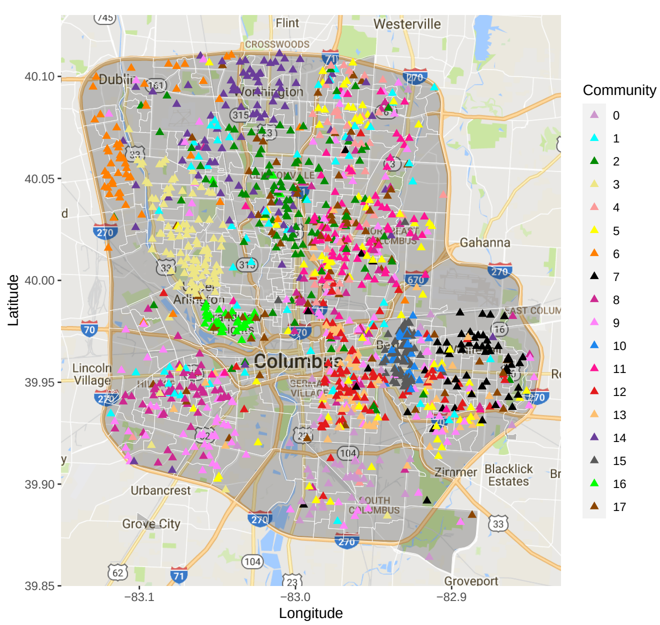

Figure 8: LINE: Number of negatives=3, number of samples=5 billion

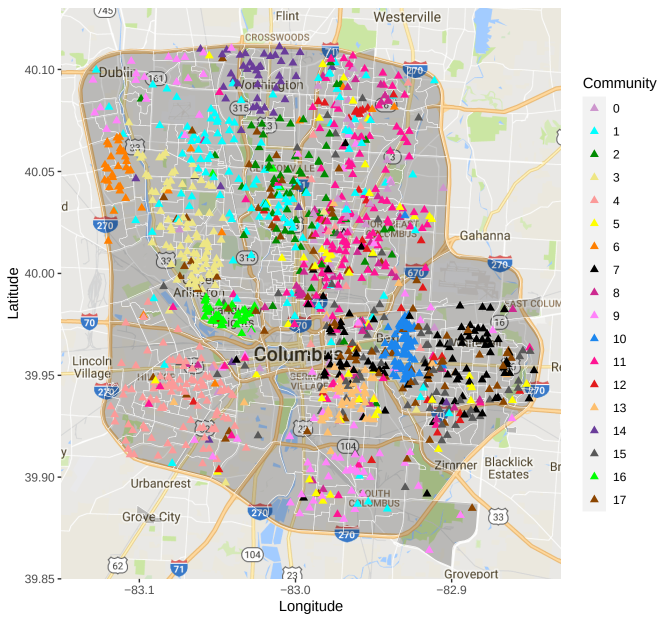

Figure 9: LINE: Number of negatives=10, number of samples=10 billion

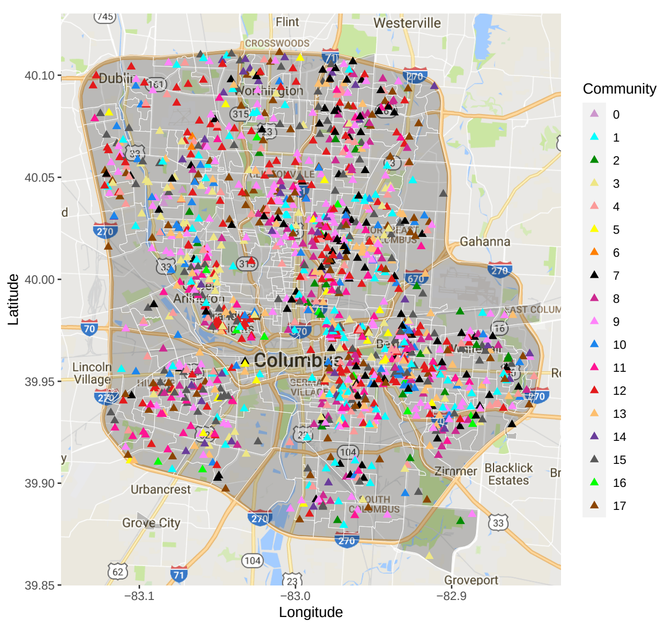

Figure 10: BINE: Number of negatives=1, window size=5, probability=0.15, beta=1, gamma=0.1

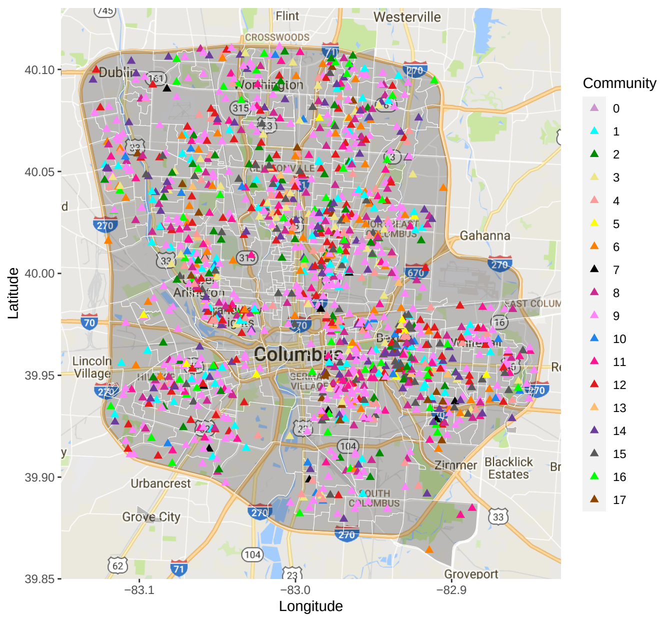

Figure 11: BINE: Number of negatives=4, window size=9, probability=0.15, beta=0.01, gamma=1

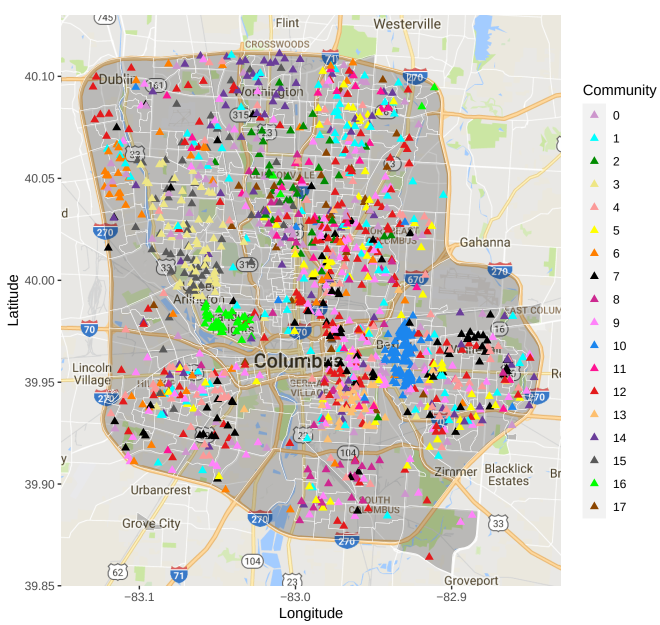

Figure 12: LocationTrails: Context window=5

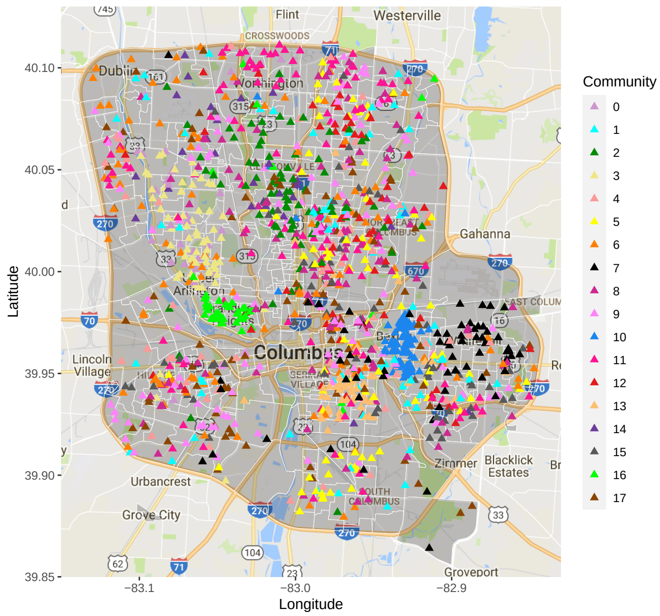

Figure 13: LocationTrails: Context window=10

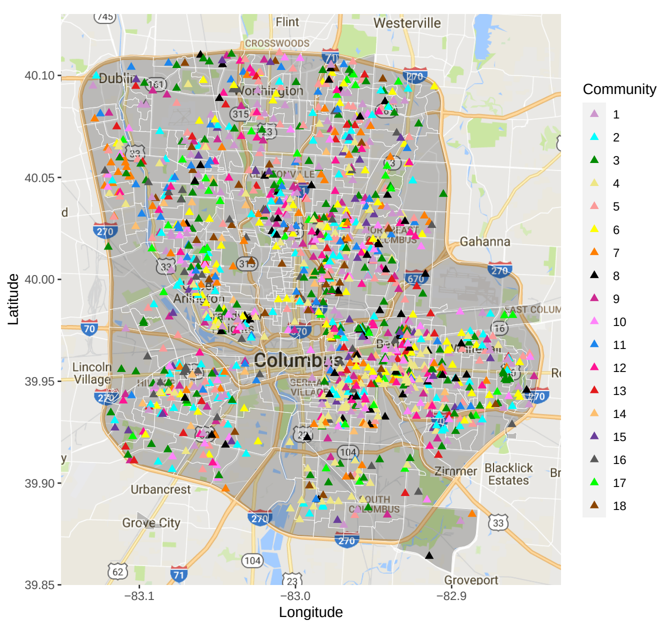

Figure 14: LDA: Gibbs number of iterations=10,000

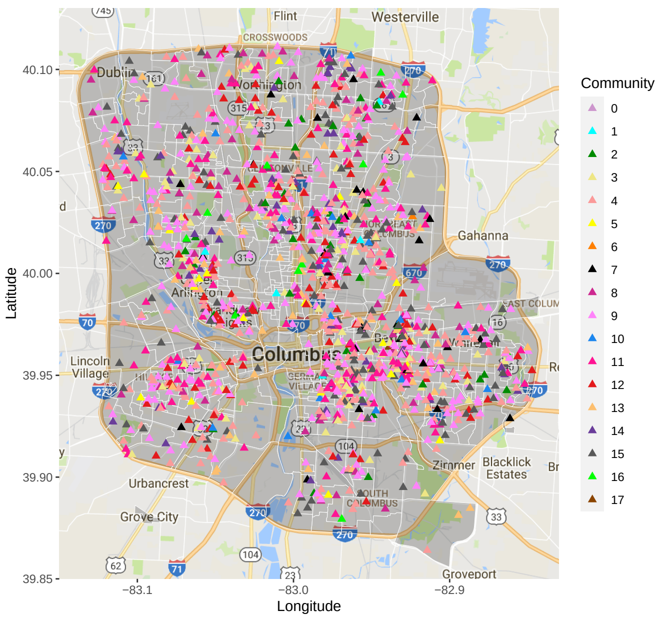

Figure 15: BINE: Number of negatives=10, window size=5, probability=0.5, beta=0.001, gamma=0.01

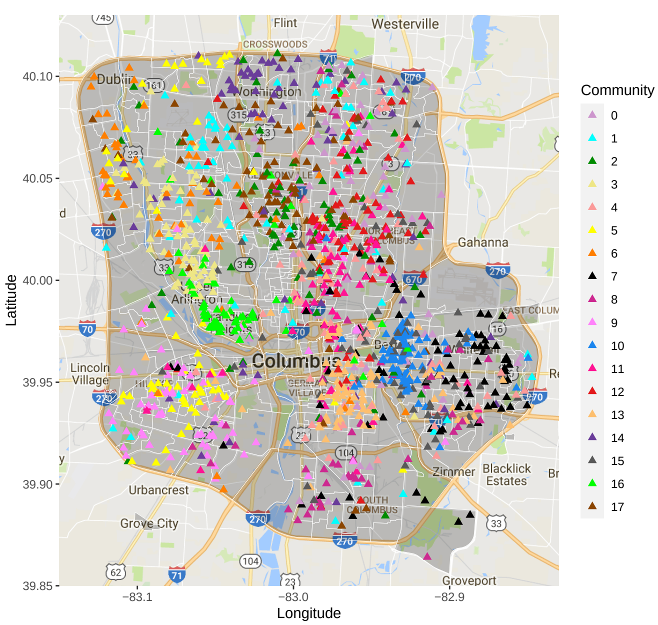

Figure 16: Metis: Cut objective="volume"

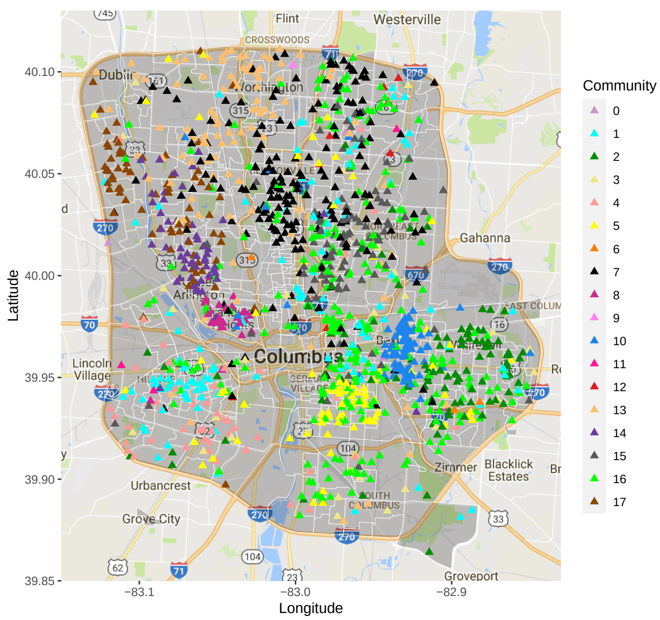

Figure 17: Graclus: Cut objective="ratio association"