Social Network Analysis Project
Traffic Flow Analysis Using Uber Movement Data

Mohammad Reza Ghasemi Madani, Riccardo Spolaor
Second Cycle Degree in Artificial Intelligence

Abstract

The proposed work presents a study of traffic flow in two cities (Bristol and Cincinnati) based on the Movement dataset provided by *Uber*. For each city, a temporal network defining the mean travel times during weekdays at different hours is provided along with a spatial graph defining the city structure. The analysis is divided into two steps. Firstly, node centrality is computed in order to investigate traffic behavior throughout different times of the day in comparison to the topographic structure of the cities. Next, community detection is applied to identify travel clusters and regions with similar characteristics. The findings clarify the patterns and characteristics of urban traffic in the studied areas which can be utilized in further studies such as transportation planning.

1 Context

This project outlines a study on *Traffic Flow Analysis*, a subfield of *Urban Planning* that focuses on understanding and modeling the movement of vehicles and people on roads, highways, and other transportation networks. By analyzing data on traffic volume, speed, density, and other factors, *Traffic Flow Analysis* seeks to identify patterns, predict congestion, and improve the efficiency and safety of transportation systems. *Urban Planning* research has long investigated these topics, but the increasing availability of GPS and mobile phone data has made them more relevant than ever. Since 2016, data from over two billion *Uber* trips taken in various cities worldwide have been available for analysis.

2 Problem and Motivation

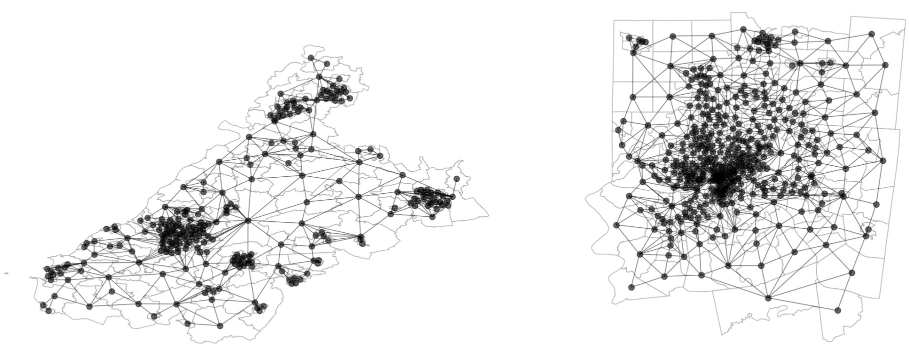
The study's objective is to replicate and extend the Stanford study, *"Traffic Flow Analysis Using Uber Movement Data,"*[1] which examines travel patterns and city structures. The work focuses on identifying and comparing mobility patterns among the cities of Bristol and Cincinnati. In particular, we try to highlight traffic bottleneck zones and compare them to the structural conformation of the cities. Moreover, we seek to spot the different travel communities present in each city. These two specific cities have been selected due to their different structural characteristics which would lead to different results in the analysis.

The identification of travel clusters and hot travel zones and their comparison with the city conformation are fundamental steps in order to predict possible traffic congestion and to improve the efficiency and safety of the transportation system and, ideally, the city infrastructure. This project aims at modeling the traffic flow in the two different cities throughout selected hours (namely: 00 : 00, 08 : 00, 13 : 00 and 20 : 00) that should roughly cover the evolution of travels during the day. Different centrality measures and community detection algorithms are applied and compared in order to verify that these methods could be valuable tools for modeling traffic flow and aid transportation planners in designing more efficient and effective traffic management strategies.

3 Datasets

Uber Movement Data is a large dataset containing information on billions of trips taken by Uber vehicles in cities around the world. The data is generated from GPS data obtained by Uber's ride-hailing service and includes information on trip origins, destinations, and aggregated travel times, including arithmetic and geometric means and standard deviations, between zones over a specified date range. It is also open to the public and can be downloaded in `.csv` format from the official website [2]. We considered for the task just information collected in the weekdays of the first quarter of 2020.

The data provided by Uber cannot be directly used to build a graph on which traffic congestion can be detected. A `GeoJSON` file describing the coordinates of the polygons which delimit the zones in a city is provided along with the data. Through the use of the python libraries `GeoPandas`[3] and `Shapely`[4], this file can be exploited to build the aforementioned spatial network that models the underlying city structure. In particular, the **spatial network** $G_s = (V_s, E_s, W_s)$ is built as an undirected graph where the nodes (V_s) correspond to different blocks of the cities and the edges (E_s) contain the distance (W_s) between neighboring regions. Two regions are connected with an edge if the polygons identified by their coordinates are adjacent. The distances are computed as the *Haversine distance*[1][5] between the centroids of the neighboring zones through the implementation provided by the Python library `Scikit-learn`[6]. The spatial graphs can be observed in Figure 1.



(a) Bristol. (b) Cincinnati.

Figure 1: Spatial Networks for the considered cities.

Further, a **temporal directed graph** $G_t = (V_t, E_t, W_t)$ with dynamic weights on the edges is defined for each city. The nodes correspond to different regions and are a subset of the ones

[1] It measures the angular distance between two points on the surface of a sphere and it is particularly indicated for geographic measurements involving coordinates.

present in the spatial network ($V_t \subseteq V_s$), while the edges (E_t) include the average travel time (W_t) over the different considered hours of the day across the regions. The dynamicity of the temporal network is handled by building a dictionary that for each considered hour, contains the view of the temporal network at that specific hour.

Certain community detection algorithms are expensive due to their exponential complexity and others are generally unsuited for directed networks. To handle the community detection task given the mentioned constraints, the temporal network of both cities is turned into an **undirected** counterpart $G_t' = (V_t, E_t', W_t')$ where if edges existed among two nodes in V_t they would be turned in an undirected edge in E_t' containing the mean value among them in W_t'. This way, the edge dimension of the network is reduced and the network is turned to undirected. Due to the fact that edges in both versions of the temporal networks can connect even distant regions, the densities of these are naturally larger than the ones of the spatial networks which contains edges just among neighbouring zones. The `Networkx`[7] library has been used to build the aforementioned graphs and exact or slightly modified versions of their algorithms have been utilized to pursue the analysis and obtain the results. Table 1 shows the networks specifications used in the projects.

(a) Spatial Graphs.

City	***Bristol***	***Cincinnati***
# of nodes	268	454
# of edges	719	1385
Density	0.02	0.01

(b) Bristol Temporal Graph.

Day time	***00:00***	***08:00***	**13:00**	**20:00**
# of nodes	193	191	200	198
# of edges	5435	6458	5711	6546
Density	0.15	0.18	0.14	0.17

(c) Cincinnati Temporal Graph.

Day time	***00:00***	***08:00***	**13:00**	**20:00**
# of nodes	440	442	441	442
# of edges	31685	36653	36880	35075
Density	0.16	0.19	0.19	0.18

(d) Bristol Undirected Temporal Graph.

Day time	***00:00***	***08:00***	**13:00**	**20:00**
# of nodes	193	191	200	198
# of edges	3406	3849	3326	3926
Density	0.18	0.21	0.17	0.20

(e) Cincinnati Undirected Temporal Graph.

Day time	***00:00***	***08:00***	**13:00**	**20:00**
# of nodes	440	442	441	442
# of edges	20281	22595	22382	21411
Density	0.21	0.23	0.23	0.22

Table 1: Networks specifications.

3.1 Validity and Reliability

Validity and reliability are critical aspects of network data, which determine the quality and consistency of the information stored and collected in a network. ***Validity*** is an important aspect of the data quality that refers to the accuracy of the information and whether the data actually represents what it is intended to represent. Therefore, considering the validity of collected data is crucial to ensure the trustworthiness of the experiments performed on it.

The *spatial network* of the city of Bristol is built directly on geographic information provided by *Ordnance Survey*[8] and the one of Cincinnati considers data from *TomTom*[9]. The validity of their data is assumed to be high, due to their rigorous processes and standards to collect and maintain it. Each region of the city is assigned to a specific node and the distance between regions is accurately computed from the provided coordinates. Hence, no ***omission*** and ***commission errors*** are expected in the spatial graph.

The validity of the information present in the temporal networks can be verified as Uber described their data gathering and processing pipeline in a transparent way in the document [10].

Anonymized and aggregated travel data is collected by Uber through the information provided by the driver-partners smartphones. This approach is well-suited for providing accurate data in areas with limited commercial traffic and where fixed sensor infrastructure is not cost-effective. Travel time statistics are removed for zone pairs with insufficient trip numbers or unique riders for privacy reasons, resulting in some ***omission errors***. However, the missing nodes and edges relate to non-busy areas and low mean travel times, so they are unlikely to cause ***commission errors***. The final data still contains key nodes and edges, and the missing information is unlikely to have a significant impact on the quality of the measures.

To ensure objectivity and repeatability in the built models and analysis, the ***reliability*** of the proposed work is examined. The information present in the temporal networks is obtained through objective data collected and processed from digital devices. The process used to build spatial networks is based on computing distances among coordinates from authoritative sources. The analysis is performes through generally deterministic measures applied algorithmically for high repeatability of the experiment. However, specific algorithms may require tweaking, such as for community detection, which may use randomness-based heuristics or specific hyperparameters. To address this, the provided values of hyperparameters and random seeds will be illustrated to guarantee result repeatability. The proposed work can be considered reliable.

4 Measures

Centrality and community detection algorithms are applied on both the spatial and temporal networks of the different cities in order to extract insights about the traffic flow.

4.1 Centrality Measures

We examine centrality metrics to identify bottlenecks or peak traffic zones in the temporal network and core/peripheral zones in our spatial network, aiming to better understand the structural characteristics of urban traffic flow and human activity in the considered cities.

- **Node Degree:** In temporal graphs, nodes with high weighted in-degree and out-degree are likely to correspond to areas of high traffic congestion. Specifically, high in-degree nodes indicate areas that are the destination of many travelers, while high out-degree nodes indicate areas that are the origin of many travels.
- **Betweenness Centrality:**[11] Betweenness centrality is a measure of a node's importance in a network based on the number of shortest paths that pass through it. Hence, nodes with high betweenness centrality are considered to be structurally important for connecting regions in the spatial graph. In the temporal graph the betweenness values of the nodes can be seen as traffic movement. Changes in this metric during different hours of the day can evaluate traffic movement and predict traffic flow.
- **Closeness Centrality:**[12] Closeness centrality measures nodes' importance in a network based on how quickly it can reach all other nodes in the network. Regions with high closeness centrality can quickly access other areas in the network. So in the spatial and temporal graph the nodes in the center should have the highest closeness centrality scores.
- **PageRank:**[13] PageRank centrality considers a node's importance based on the number and quality of links pointing to it. In the context of traffic analysis, links to central regions

count more compared to the low transit regions. The PageRank of nodes in the temporal graph reflects the amount of traffic directed to them or adjacent regions and it can be employed to locate hotspots in cities during various times of the day. In the spatial graph, it highlights zones that connect densely joined regions.

- **HITS:**[14] It gives each node in a directed network two centrality scores: the authority and the hub centrality. High authority nodes are the "more resourceful", hence the ones receiving more links. Differently, high hub scores are given to nodes that connect to high authority nodes. In the temporal graphs, authority can be interpreted as a measure of total time invested to reach a given region, while hub scores are a measure of the total time spent to drive to main travel regions. The results of this metric on the spatial graph are not particularly important.

4.2 Community Detection

The community detection algorithms aim to discover groups within the networks as part of our objective to reveal insights about the traffic flow structure. The results on the temporal graphs may help locate specific areas within the cities which display similar traffic behavior during certain times of the day. Regarding the spatial network, the found communities should indicate structurally similar areas. As explained in the *Dataset* section we applied community detection on an undirected version of the temporal graphs. The **modularity**[15] measure has been considered as an objective evaluation metric for the quality of the communities. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

- **Girvan-Newmann Strength of Weak Ties algorithm:**[16] In this algorithm the edges with the highest weighted betweenness are gradually removed and the measure is re-computed after every removal until the network falls off to smaller components. In the temporal graph the obtained communities should consist in travel clusters which can be connected through others by highly traffic-busy edges. In the spatial graph it should define densely connected areas which are highly distant from each other. The parameter considering the number of communities to obtain has been set as 4.

- **Community Detection based on Weighted K-Cores:**[17] A weighted k-core is a connected set of nodes where each has a weighted degree of value at least k. Once a k-core community is computed, the algorithm is repeated for the remaining subgraph until all nodes are assigned to a community. This algorithm should define in the temporal graph similarly connected traffic-busy clusters, while it should create a partition of the spatial network based on groups of connected zones with similar distance between each other. The parameter k is re-assigned every iteration as the average weighted degree among all nodes non assigned to a community.

- **Clique Percolation Method:**[18] A community detection algorithm that computes communities from *k-cliques*. In addition, in order for a k-clique to be considered, the geometric mean among the weight of its edges (or *intensity*[19]) should be less or equal than a threshold l. The algorithm is repeated in the remaining subgraph until all nodes are assigned to a community. It should identify group of nodes more densely connected between each others, such as connected regions in the spatial graph or zones connected by travels in the temporal ones. The intensity component should penalize weak connections in the communities. k In this work is initialized as 2 for the spatial graph and as 4 for the temporal graph., while l is automatically re-assigned at each iteration as the geometric mean between all the edges among the not yet classified nodes.

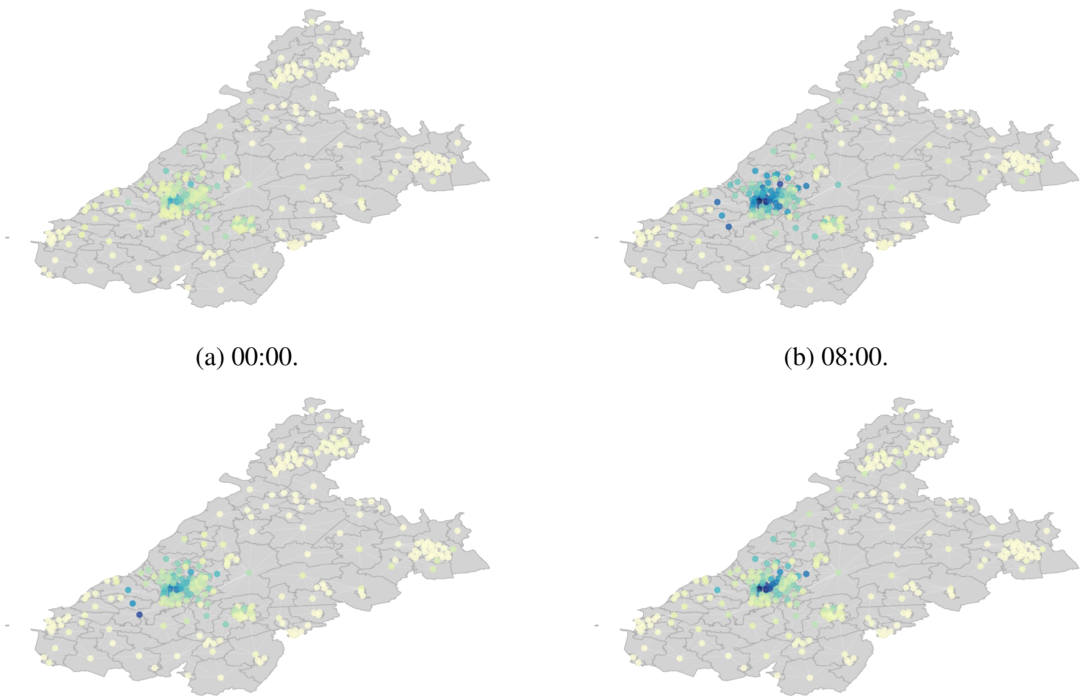
- **Louvain Community Detection Algorithm:**[15] A heuristic community detection algorithm that computes communities by optimizing the modularity of the network. Given this premise, the obtained partition will obviously be the ones reching the highest modularity scores. Hence, the results of this methodology are used as a baseline to compare the quality of the other community detection algorithms. The "resolution" parameter is initialized as 1 favouring average sized communities. The modularity gain threshold for each iteration is assigned as 10^{-8}. The seed for Random Number Generation is set at 42.

5 Results

The results of the analysis are presented in this section. Both qualitative and objective insights are presented and comparisons among the results on the spatial and temporal network at the different considered times are provided when meaningful.

5.1 Centrality Measures

Regarding the temporal graph, the centrality values are normalized in a scale between 0 and 1 considering the centrality value of the nodes across all the different hours. The same normalization range is applied to the spatial graph results. Blue-colored nodes have centrality close to 1, while the yellow ones are near 0.



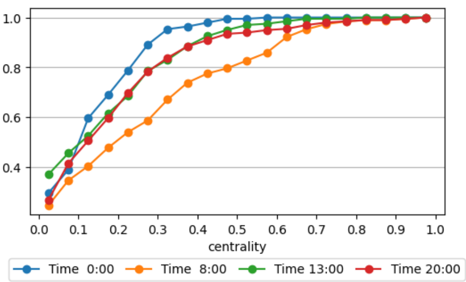
(a) 00:00. (b) 08:00.

(c) 13:00. (d) 20:00.

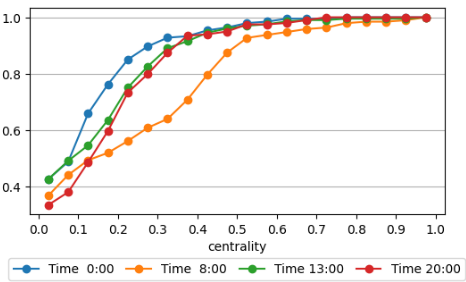
Figure 2: Bristol in-degree centrality on the temporal graph.

- **Node Degree:** For both cities it can be observed that the highest in- and out- degree centralities are found in the central areas meaning that travel activities mostly occur there.

Regarding in-degree, both cities show a similar pattern. Namely, at midnight the centrality value is lower than at other times of the day. In contrast, the highest activity is present at $8:00$, while a lower, but still stable activity can be seen at $13:00$ and $20:00$. These results inform that at midnight travels to destination zones are absent or sparse. On the other hand, a lot of travels to central zones of the city occur at $08:00$, which makes sense since it coincides with the time where most people move to work. The results for the city of Bristol are illustrated in Figure 2. Differently, out-degree centrality values change less during the day, with still higher peaks present at daytime hours such as $8:00$ and $13:00$. This result can be interpreted as the fact that during the day most travels originate from just a small amount of zones and they are constant during the day.



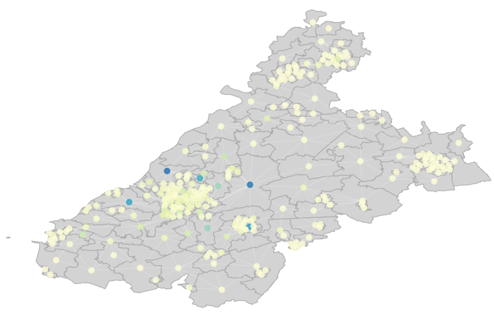
(a) In-degree.



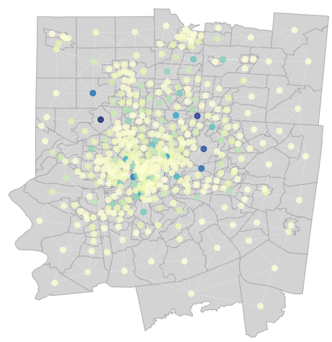
(b) Out-degree.

Figure 3: In-/out-degree centrality cumulative distribution in the temporal graphs.

Looking at the cumulative graphs in Figure 3, it can be observed that the curve at midnight increase sharply, meaning that number of low centrality nodes is high. On the other hand, the curve at $8:00$ grows much more slightly in comparison, since more nodes with higher centrality are present. Furthermore, in the out-degree chart it can bee seen that curves at times $00:00$, $13:00$ and $20:00$ start to behave in a similar manner after considering centralities of value 0.3. This furtherly approves the given interpretation.



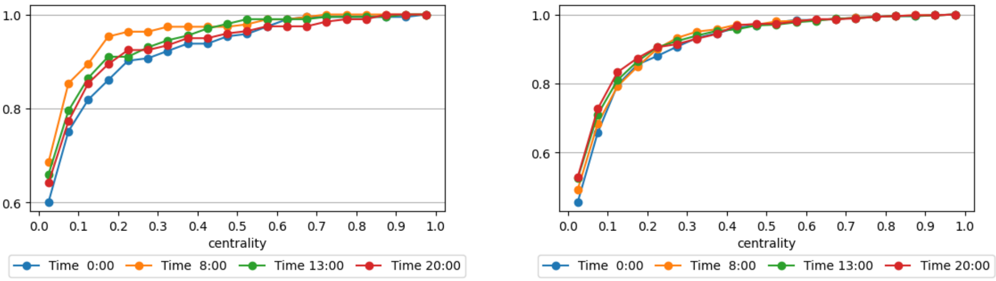
(a) Bristol Betweenness at 8:00.



(b) Cincinnati Betweenness at 8:00

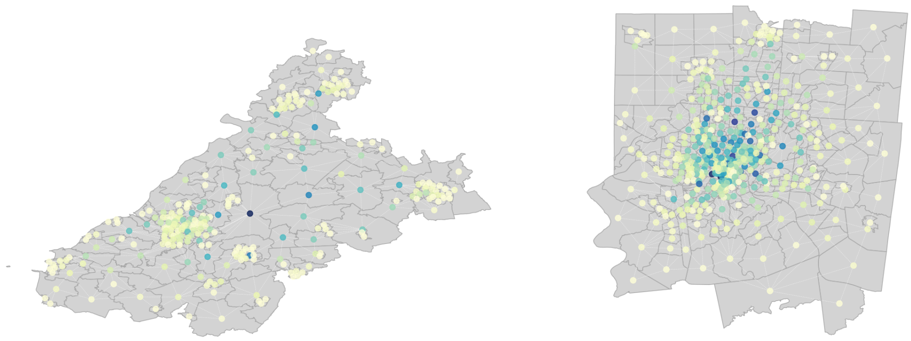
Figure 4: Betweenness Centrality on the temporal graphs.

- **Betweenness Centrality:** For both cities, the distribution of the betweenness centrality remains roughly the same during the day, meaning that the same regions are highly traversed to reach different destinations. Moreover, just a few nodes have high betweenness and they seem to connect spatially dense areas of the cities. Figure 4 shows the value of the betweennes centrality of the various regions at time $08:00$ in the two cities. The cumulative distributions in Figure 5 confirms this behavior.



(a) Bristol. (b) Cincinnati.

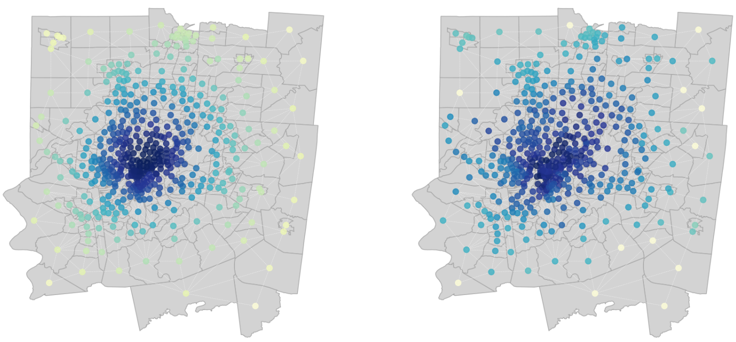
Figure 5: Bristol and Cincinnati Betweenness centrality cumulative distribution in the temporal graphs.



(a) Bristol. (b) Cincinnati.

Figure 6: Betweenness Centrality of the spatial graphs.

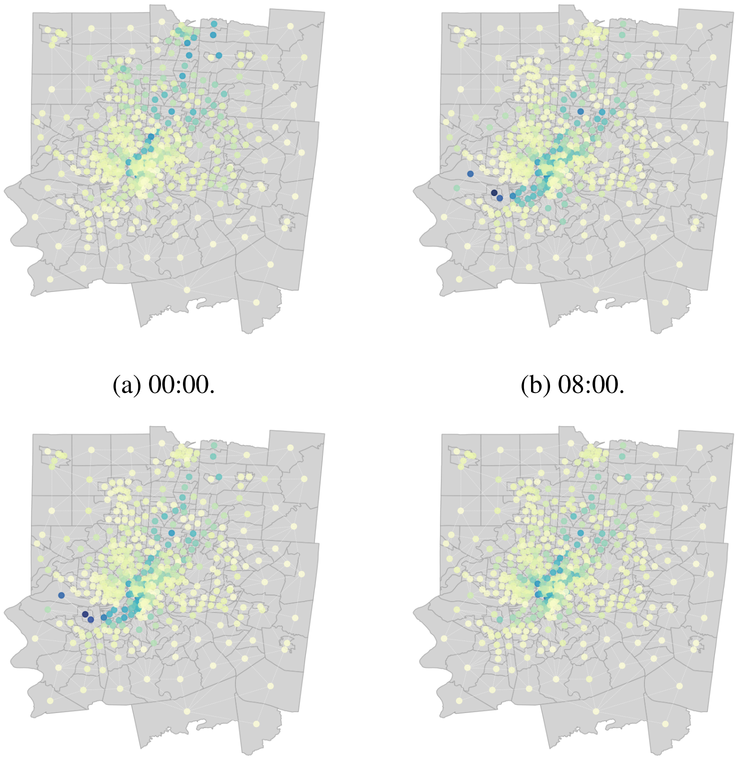
The betweenness values obtained in the spatial network highlight zones that are important to connect different regions. Regarding Bristol, it appears that these zones correspond to the high centrality nodes found in the temporal network. Contrarily, for Cincinnati it seems that traffic-busy zones do not correspond one-on-one to structurally important zones. Figure 6 illustrates the results.



(a) Spatial Graph. (b) Temporal Graph at 20:00.

Figure 7: Closeness Centrality of Cincinnati.

- **Closeness Centrality:** As expected, in the spatial graph of both cities the nodes in the city center obtain the highest closeness centrality, while the ones at the borders present lower values. Generally speaking, the temporal networks of both cities present a similar closeness distribution during the day. Furthermore, the nodes in the city centers present the highest values as in the spatial counterparts, meaning that they are central travel zones. An interesting difference in contrast to the spatial network is that in the temporal networks even a selected part of periphery nodes obtain reasonable closeness values. This denotes that even some non-central areas are determinant in shaping urban mobility. Figure 7 illustrates the described behavior for Cincinnati, considering the traffic situation at 20 : 00.

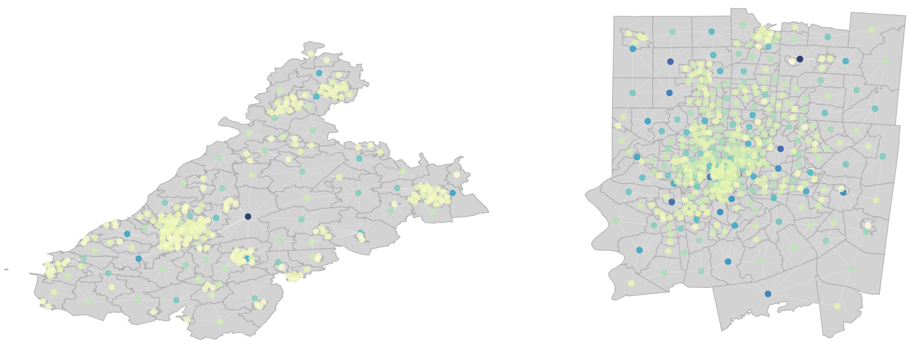


(a) 00:00. (b) 08:00.

(c) 13:00. (d) 20:00.

Figure 8: PageRank Centrality of Cincinnati temporal graphs.

- **PageRank:** Regarding Bristol, some central nodes reach high centrality during the day. An interesting observation is that at 20 : 00, which is a rush hour, even a small amount of periphery nodes seem to obtain high PageRank value meaning that they turn to be important zones to reach high-demand travel areas.



(a) Bristol. (b) Cincinnati.

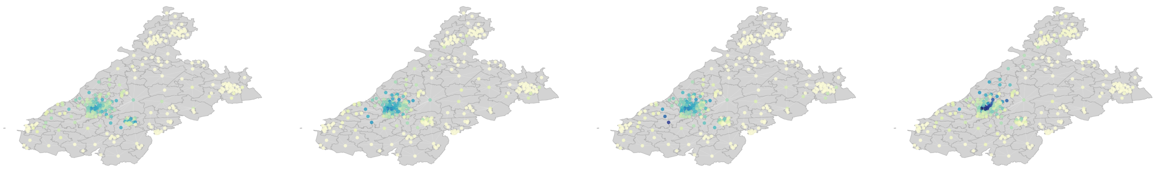
Figure 9: PageRank Centrality on the spatial graphs.

Similar behavior can be observed in Cincinnati (Figure 8). In particular, the central-northern area contains the nodes with the highest centrality measures. During working hours, namely 08 : 00 and 13 : 00 even the central-west part of the city becomes relevant, judging by the high PageRank of the regions in this area.

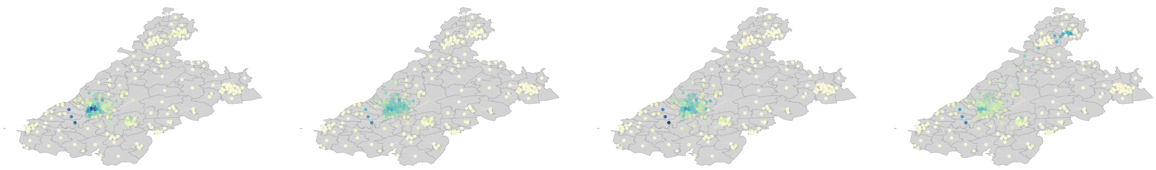
In the spatial network, a high centrality is assigned to nodes that connect dense regions and are thus far away from their neighbors. Some similarities can be observed with the results obtained by betweenness centrality, especially regarding the city of Bristol as seen in figure Figure 9.

- **HITS:** Regarding Bristol, it can be observed that the central area always contains high authority zones during the day. Interestingly, at 20 : 00, just a small part of these central nodes receives high authority, although their value overcomes the highest ones obtained at different hours. This can be interpreted as the fact that there are less highly reached areas at 20 : 00, although the travel times to get to them are particularly high. Hubs in the same city are again observed in the central areas during the day. A peculiar characteristic is that at 00 : 00 a small portion of these central nodes obtain higher values, meaning that these zones are popular departures to densely reached parts of the city. Moreover, at 20 : 00, even a cluster of periphery nodes in the northern area of the city receives high hub centrality. The results are observed in Figure 10.

  Similar results are obtained for Cincinnati, where central areas receive high authority and hub scores, with small variations during specific times of the day.



(a) Authorities at 00:00. (b) Authorities at 08:00. (c) Authorities at 13:00. (d) Authorities at 20:00.



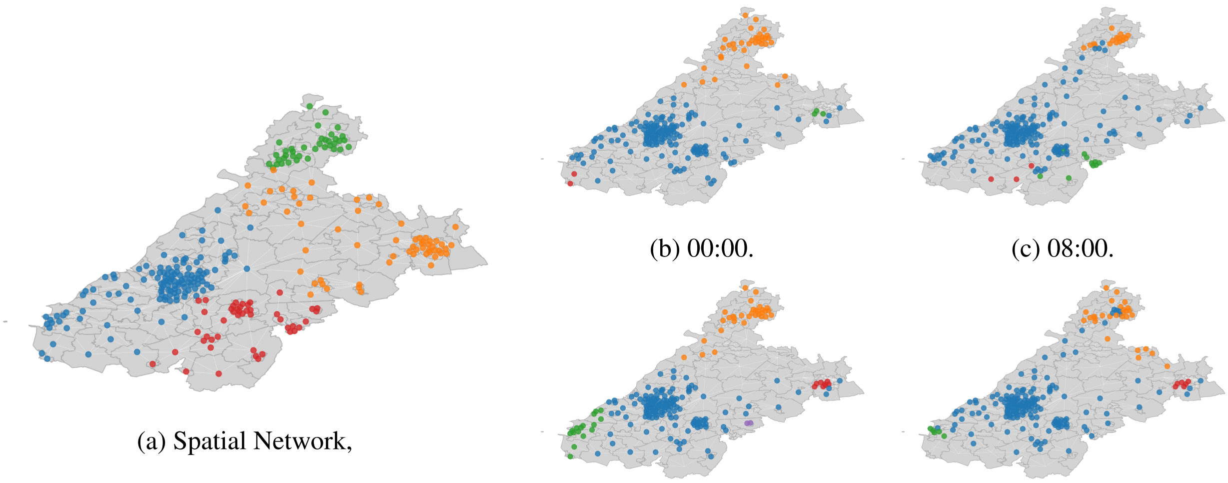
(e) Hubs at 00:00. (f) Hubs at 08:00. (g) Hubs at 13:00. (h) Hubs at 20:00.

Figure 10: Authorities and Hubs Centrality on the temporal graphs of Bristol.

5.2 Community Detection

After a qualitative observation of the communities obtained from the temporal and spatial graphs, the strength of each graph partition is evaluated through the modularity score for a more objective insight.

- **Girvan-Newmann Strength of Weak Ties algorithm:** Regarding Bristol, we can generally identify a large travel community in the center during the whole day and another small one in the northern part. It is interesting to note how this last community has more presence during non rush hours. Another insight is that between 13 : 00 and 20 : 00 two additional important travel communities can be observed in the far east and west sides of the city. A general interpretation of this behavior is that at 08 : 00, when people move to work, the majority of travels occur in the center of the city, while at later times of the day people start to reach even other peripheral zones. An intuitive community division is finally observed in the spatial graph, where central densely connected zones generate their own partition alongside neighbouring peripheral nodes. A partial overlap between the travel and topographic communities can be observed, in particular for what concerns the central and northern areas of the city. Results can be observed in Figure 11.



(a) Spatial Network,

(b) 00:00.

(c) 08:00.

(d) 13:00.

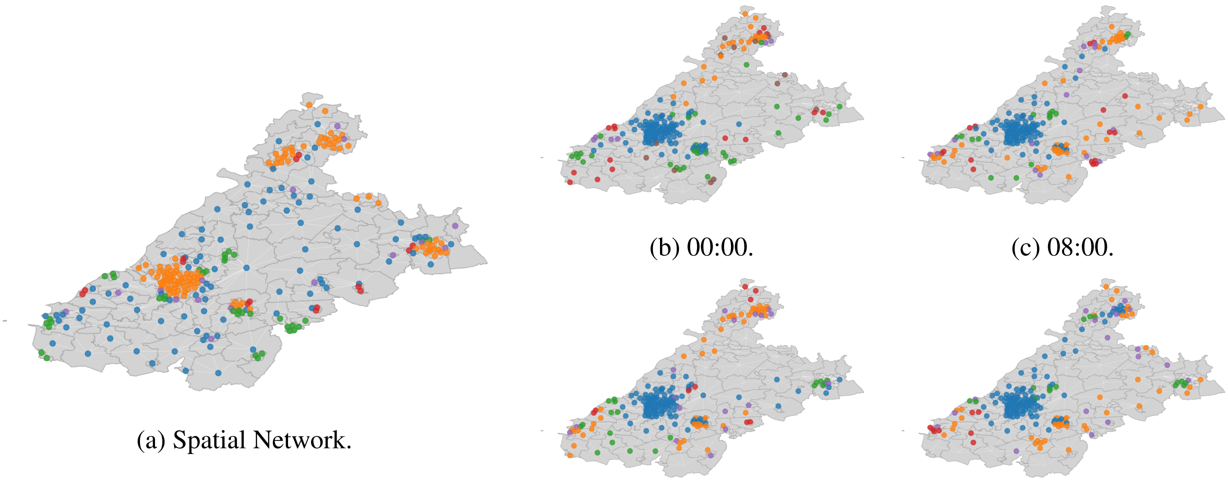
(e) 20:00.

Figure 11: Girvan-Newmann Communities of Bristol.

For what concerns Cincinnati, the results on the temporal network are not so meaningful or reliable, as it easily identify just a big travel community across all city. On the other hand, the results on the spatial graph once again meet the expectations, as densely connected zones are collected in their own clusters.

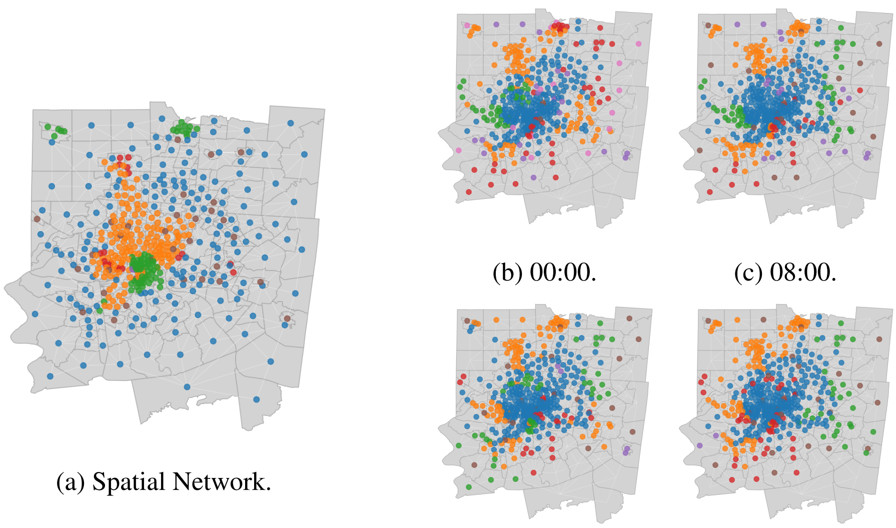
- **Community Detection based on Weighted K-Cores:** Regarding Bristol, it can be once again observed that a main community is found in the centre of the city. Similar travel behaviors can be generally sighted in the north and south section of the city during the day as highlighted by another big community that spreads in these areas. This last community is absent in the southern area at 00 : 00, while it has a scarce presence at 20 : 00 in the north of the city. This may indicate that similar travel behaviors are observed at opposite sides of the city during these two considered hours. In the spatial network, as expected, clusters are obtained from zones that have a similar density. The sole visible partial overlap of clusters between the spatial and temporal networks can be once again seen in the central zone of the city. Results are illustrated in Figure 12.

  In Cincinnati, it can once again be discerned a main travel cluster in the center, which remains roughly constant during the day. The other groups are more peripheral but manage to also reach parts of the center. A smaller community can also be spotted in the most topographically dense part of the city center. Another main cluster can be identified and it has a far less significant presence in the east part of the city at 08 : 00 and 20 : 00, which can be associated with rush hours. The most interesting insight is that another important community can be spotted and it slowly translates from the western part in the early hours to the eastern part in later times, meaning that similar travel behaviors can be observed in opposite parts of the city at the beginning and the end of the day. Once again the spatial network shows a clustering clearly defined by zones with similar distance between each other. The main overlaps with the temporal network communities are observed for the central areas. It is interesting to point out how the small community observed in the most dense part of the centre finds correspondence with a cluster observed even in the temporal network. Results are shown in Figure 13.



(a) Spatial Network. (b) 00:00. (c) 08:00. (d) 13:00. (e) 20:00.

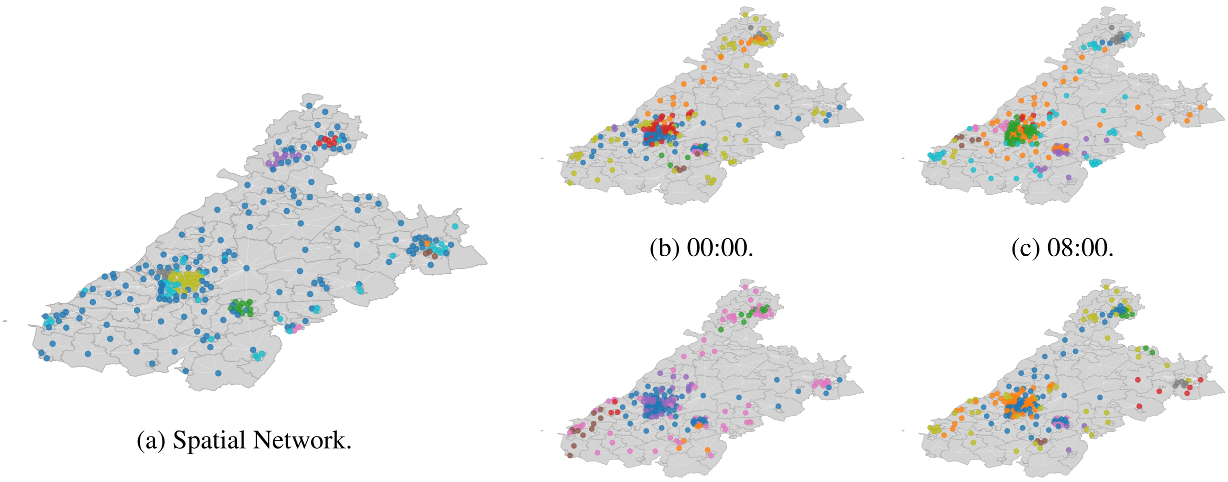
Figure 12: K-Core Communities of Bristol.



(a) Spatial Network. (b) 00:00. (c) 08:00. (d) 13:00. (e) 20:00.

Figure 13: K-Core Communities of Cincinnati.

- **Clique Percolation Method:** It is important to consider how, due to the nature of the algorithm, the obtained communities are not color-correspondent in the various views of the temporal graph. In Bristol it is peculiar how the centre of the city is now divided between two different travel communities during the day and one of the two spreads in the neighbourhood. Moreover, the other topographically dense regions of the city are generally all collected in another cluster pointing out that similar travel patterns occur among these regions. No noticeable changes happen during the day, possibly meaning that fully connected travel zones remain roughly the same at different hours. In the spatial graph we can observe two clusters in the city center that may partially overlap with the ones found in the temporal network, The other spatially dense areas are collected in their own clusters, while the sparse region around these dense areas form its own group. Figure 14 presents the results.



(a) Spatial Network.

(b) 00:00.

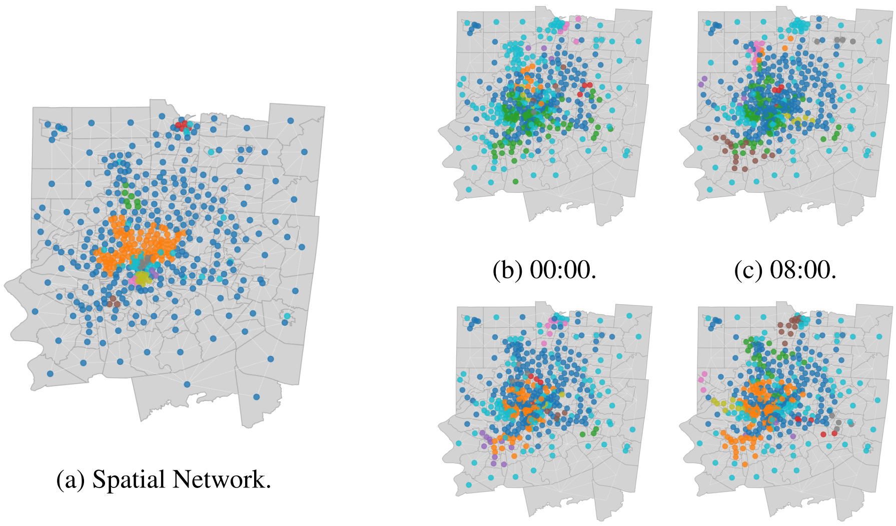
(c) 08:00.

(d) 13:00.

(e) 20:00.

Figure 14: Clique Percolation Communities of Bristol.

With respect to Cincinnati, a behavior in contrast with the previous algorithms results emerges in the temporal network, as the city center is contended among different communities which spread from the periphery to the center. This suggest a more chaotic behavior among travellers, which cannot be collected into specific patterns. The spatial network presents a more interesting division where, generally speaking, two main clusters can be observed in the city centre and in the periphery. Figure 15 highlights the results.



(a) Spatial Network.

(b) 00:00.

(c) 08:00.

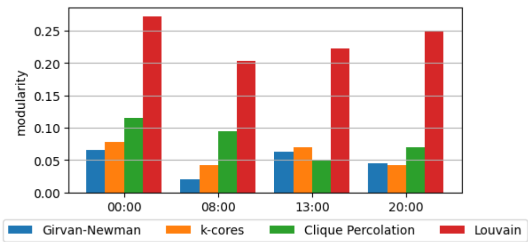
(d) 13:00.

(e) 20:00.

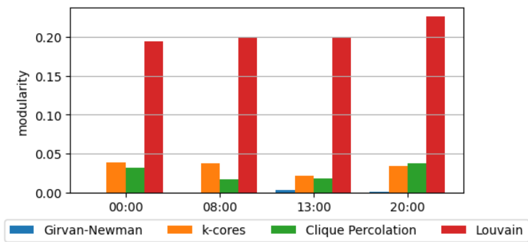
Figure 15: Clique Percolation Communities of Cincinnati.

- **Louvain Community Detection Algorithm:** In the temporal network of Bristol the city center is generally disputed between two main communities except for 08 : 00 where just a group is mainly present in there. This can be connected to the fact that 08 : 00 is a time of the day where most people move to work, hence a similar travel behaviour occurs in the whole city centre. The other smaller structurally dense areas are instead collected in other clusters. While the second behavior is generally shared among the other community detection algorithms (except for Girvan-Newmann), the former is peculiar of Clique Percolation. Considering the partitions obtained by Louvain as a baseline model, the results obtained by Clique Percolation are qualitatively the best ones, while the ones of Girvan-Newmann the worst. In the spatial network we can once again observe a main community in the centre and other bigger communities in the periphery. This interpretation visually roughly overlaps with the one obtained by Girvan-Newmann, whilst it is in contrast with the ones suggested by the other two algorithms.

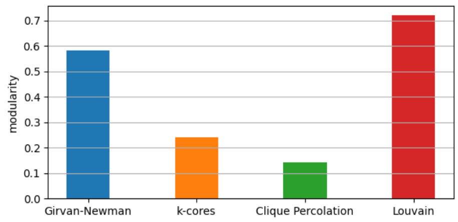
  For Cincinnati, the method finds communities that share parts of the center in a similar way to Clique Percolation and K-Core, and in complete opposition to Girvan-Newmann. The modularity value for Girvan-Newmann is always close to 0. Once again the spatial partitioning of Louvain is more closely similar to Girvan-Newmann. It is interesting to notice how even K-Cores obtains interesting modularity results for the spatial network partitioning.



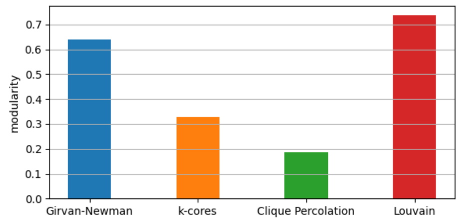
(a) Bristol temporal graphs modularity.



(b) Cincinnati temporal graphs modularity.



(c) Bristol spatial graph modularity.



(d) Cincinnati spatial graph modularity.

Figure 16: Modularity comparison for different community detection algorithms.

All of these qualitative considerations are confirmed by the modularity results as it can be seen in Figure 16. In particular, Girvan-Newmann seems to perform more closely to Louvain, hence better, when we consider sparse graphs as the spatial one, while it underperforms massively in densely connected graphs as the temporal ones. Differently, Clique Percolation and k-cores perform better on the temporal network than Girvan-Newmann. From a complexity point of view Girvan-Newmann performs very badly, while the fastest results are reached by k-cores and Louvain. Due to the fastness of k-core we can consider it appropriate for finding travel and space communities. Louvain remains the best solution with respect to our objective score, but unlike k-cores it cannot create non-connected communities.

6 Conclusion

In conclusion, this study has demonstrated the effectiveness of centrality measures and community detection in modeling traffic flow in Bristol and Cincinnati. By examining changes in centrality measures over time in our temporal networks and comparing them to the static spatial network, we were able to identify mobility patterns during different hours of the day. The measures can be exploited to obtain specific insights. Node degree informs about which zones are popular origins or destinations. Betweenness on the other hand highlights possible bottleneck zones, while closeness identifies central travel zones. Differently, PageRank identifies areas which are high-demand destinations or that are fundamental sources to reach other hot-spots. Finally, HITS centrality illustrates hub zones that are directly linked to other "authoritative" high-traffic areas. The topographic structure can be shaped by either Betweenness, PageRank or Closeness. While the first two generally highlight areas connecting dense parts of the city, the latter defines central nodes. The combination of these results helps understanding how traffic zones are structurally located in the city.

The results of the community detection algorithms all inform that travel behaviors generally maintain a core pattern in the center of the city during the day. More peripheral zones travel patterns can instead change over time as in some cases traffic would flow or expand to different parts of the cities during the day. Generally speaking, topographic communities observed in the spatial network distinguish between central dense regions and peripheral zones. Usually, the core territories roughly correspond to specific travel communities, while periphery areas experience more varied movement behaviors. Our study identifies Louvain as the most objectively reliable community detection algorithm, although Girvan-Newmann performs quietly well when it is tasked to identify spatial communities, and Clique Percolation and k-cores obtain nice results on the temporal networks.

7 Critique

Our results suggest that these methods could be valuable tools for transportation planners in designing more efficient and effective traffic management strategies. Moving forward, the next logical step can be to propose and implement solutions to reduce traffic congestion in the identified bottlenecks or to facilitate the transportation of the identified travel communities. Simulated data can be used to measure the impact of these solutions on traffic flow and has the potential to make significant contributions to the field of transportation planning. We believe our work is pretty exhaustive as it confronts the problem of shaping traffic insights using a large array of different metrics. One flaw of the project is that community detection, mainly for computational reasons, is applied to undirected versions of the temporal networks. Indeed, information is lost in the process of transforming these graphs. In order to address this problem, a faster implementation of the proposed community detection algorithms could be used by the means of lower-level programming languages or GPU acceleration. Indeed, all the community detection algorithms should be revised to address the newly imposed directionality of the networks.

References

[1] Mackenzie Pearson and Sofia Samaniego. "Traffic Flow Analysis Using Uber Movement Data". In: 2017. URL: http://snap.stanford.edu/class/cs224w-2017/projects/cs224w-11-final.pdf.

[2] *Uber Movement: Let's find smarter ways forward, together*. 2017. URL: https://movement.uber.com/?lang=en-US.

[3] Kelsey Jordahl et al. *geopandas/geopandas: 0.12.1*. Version v0.12.1. July 2020. URL: https://geopandas.org/en/stable/index.html.

[4] *Shapely*. Version v2.0.1. URL: https://shapely.readthedocs.io/en/stable/manual.html.

[5] Song Gao et al. "Understanding Urban Traffic-Flow Characteristics: A Rethinking of Betweenness Centrality". In: *Environment and Planning B: Planning and Design* 40.1 (2013), pp. 135–153. eprint: https://doi.org/10.1068/b38141. URL: https://doi.org/10.1068/b38141.

[6] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[7] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.

[8] *Ordnance Survey | See A Better Place*. 2023. URL: https://www.ordnancesurvey.co.uk/.

[9] *TomTom - Mapping and Location Technology*. 2023. URL: https://www.tomtom.com/.

[10] *Uber Movement: Travel Times Calculation Methodology*. 2017. URL: https://movement.uber.com/_static/c9bce307d99643c3.pdf.

[11] Ulrik Brandes. "A Faster Algorithm for Betweenness Centrality". In: *The Journal of Mathematical Sociology* 25 (Mar. 2004). URL: https://www.researchgate.net/publication/2887625_A_Faster_Algorithm_for_Betweenness_Centrality.

[12] Linton C. Freeman. "Centrality in social networks conceptual clarification". In: *Social Networks* 1.3 (1978), pp. 215–239. ISSN: 0378-8733. URL: https://www.sciencedirect.com/science/article/pii/0378873378900217.

[13] Lawrence Page et al. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab, Nov. 1999. URL: http://ilpubs.stanford.edu:8090/422/.

[14] Jon M. Kleinberg. "Authoritative Sources in a Hyperlinked Environment". In: *J. ACM* 46.5 (Sept. 1999), pp. 604–632. ISSN: 0004-5411. DOI: 10.1145/324133.324140. URL: https://doi.org/10.1145/324133.324140.

[15] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (Oct. 2008), P10008. DOI: 10.1088/1742-5468/2008/10/P10008. URL: https://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

[16] M. Girvan and M. E. J. Newman. "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12 (June 2002), pp. 7821–7826. URL: https://doi.org/10.1073%2Fpnas.122653799.

[17] Bin Liu and Feiteng Zhang. "Incremental Algorithms of the Core Maintenance Problem on Edge-Weighted Graphs". In: *IEEE Access* PP (Apr. 2020), pp. 1–1. URL: https://www.researchgate.net/publication/340391378_Incremental_Algorithms_of_the_Core_Maintenance_Problem_on_Edge-Weighted_Graphs.

[18] Gergely Palla et al. "Uncovering the overlapping community structure of complex networks in nature and society". In: *Nature* 435 (June 2005), pp. 814–818. URL: https://www.nature.com/articles/nature03607.

[19] Jukka-Pekka Onnela et al. "Intensity and coherence of motifs in weighted complex networks". In: *Physical Review E* 71.6 (June 2005). URL: https://doi.org/10.1103%2Fphysreve.71.065103.