Scientific Research Publishing

Open Journal of Statistics, 2026, 16(6), 545-578
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Adaptive Temporal Chinese Restaurant Process for Dynamic Clustering of Evolving Data Streams

Lahiru Wickramasinghe

Department of Mathematics and Statistics, The University of Winnipeg, Winnipeg, Canada
Email: l.wickramasinghe@uwinnipeg.ca

How to cite this paper: Wickramasinghe, L. (2026) Adaptive Temporal Chinese Restaurant Process for Dynamic Clustering of Evolving Data Streams. *Open Journal of Statistics*, **16**, 545-578.
https://doi.org/10.4236/ojs.2026.166026

Received: August 15, 2026
Accepted: November 17, 2026
Published: November 20, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

In the era of big data, the inherent temporal dynamics and evolving patterns of high-velocity streaming data present a significant challenge. This paper introduces the Adaptive Temporal Chinese Restaurant Process (AT-CRP), a novel nonparametric Bayesian approach for dynamic clustering of streaming data. The AT-CRP addresses the limitations of traditional clustering methods by modelling the full lifecycle of clusters, including birth, growth, and death, without requiring a fixed number of clusters. I derive an online variational inference algorithm for efficient real-time updates to cluster assignments, providing a structural approach for analyzing complex, evolving data streams. The method is validated on synthetic datasets and real-world NYC traffic data, including a month-long NYC traffic demand dataset and NYC traffic trip data. Experimental results demonstrate AT-CRP's ability to extract domain-specific knowledge and capture meaningful temporal patterns in urban traffic dynamics. My approach outperforms existing methods in clustering accuracy and pattern detection, offering insights into evolving trends in traffic flow and demand. This work contributes to the field of data mining and pattern recognition by providing a flexible, scalable method for modelling and analyzing large-scale streaming data, with potential applications in urban planning, traffic management, and real-time decision support systems for smart cities.

Keywords

Dynamic Clustering, Chinese Restaurant Process, Online Variational Inference, Nonparametric Bayesian

1. Introduction

In the era of big data, modern applications continuously generate vast amounts of

high-velocity streaming data. Analyzing such data presents significant challenges because of its inherent temporal dynamics and evolving patterns (see [1]). A fundamental requirement in data analysis is uncovering hidden structures that evolve over time, particularly in domains such as urban traffic management and smart city planning. Unlike traditional static datasets, streaming data exhibits **temporal dynamics**, where patterns emerge, evolve, merge, split, or disappear as new data arrives. This poses significant challenges for conventional clustering techniques, which assume fixed structures and fail to accommodate the fluid nature of real-world data.

Traditional clustering methods such as K-means and Gaussian Mixture Models (GMMs) are widely used for partitioning data into meaningful groups. However, these approaches operate under the assumption that clusters remain stable, making them ill-suited for high-velocity streaming data where cluster compositions continuously shift. In applications like urban traffic analysis, traffic patterns form and dissolve dynamically, while in smart city planning, shifts in population density and movement alter the underlying structure of urban dynamics. The inability of static clustering methods to adapt to such changes leads to outdated or inaccurate representations of data, limiting their effectiveness in real-time decision-making. Addressing this challenge requires clustering techniques that can dynamically adapt to evolving data structures while maintaining computational efficiency.

1.1. Motivation

A fundamental requirement for adaptive clustering is the ability to manage the full lifecycle of clusters and how they emerge, evolve, and eventually vanish over time. Consider a social media platform tracking trending topics. Topics such as "AI Ethics" may emerge, gain widespread traction, and eventually fade into obsolescence. As another example, consider a traffic management system that monitors congestion patterns. Traffic hotspots may emerge, intensify, and eventually dissipate over the course of a day. Traditional static clustering models struggle with this temporal evolution, as they cannot efficiently discard outdated topics or detect new ones in real-time. Similarly, in IoT networks, transient fault patterns in sensor readings may appear due to environmental disturbances or hardware malfunctions. Capturing these ephemeral yet meaningful clusters is crucial for anomaly detection and system reliability. However, existing clustering approaches often assume fixed structures, making them unsuitable for scenarios where clusters continuously form, evolve, and disappear.

To address this challenge, researchers have explored temporal clustering models such as Dependent Dirichlet Processes (DDPs) [2] and Time-sensitive Dirichlet process mixture models [3]. Quintana *et al.* [4] introduced temporal clustering through the use of dependent Dirichlet processes, which allow for the modeling of group-specific clusters that evolve over time by incorporating dependencies represented via directed acyclic graphs (DAGs). While these methods incorporate temporal dependencies, they lack robust mechanisms for cluster lifecycle

management and the ability to dynamically handle the birth, evolution, and disappearance of clusters. Without such mechanisms, these models may retain obsolete clusters or fail to accommodate emerging patterns effectively. This limitation motivates the development of the proposed **Adaptive Temporal Chinese Restaurant Process (AT-CRP)**, which explicitly models the full lifecycle of clusters in dynamic data streams. By integrating adaptive mechanisms, AT-CRP aims to provide a more flexible and efficient approach to clustering dynamic data streams in real-world applications, particularly in the domain of urban traffic analysis.

While existing temporal clustering models offer partial solutions, they fail to provide a unified framework for managing the entire lifecycle of clusters. An effective approach must not only track evolving cluster structures but also dynamically introduce new clusters and remove obsolete ones without manual intervention. Additionally, given the high velocity of modern streaming data, any practical solution must ensure computational efficiency, enabling real-time updates without incurring significant delays. To address these challenges, this research proposes the **Adaptive Temporal CRP (AT-CRP)**, a nonparametric approach designed to model the full evolution of clusters in dynamic data streams.

1.2. Objectives

This research aims to develop a principled framework for adaptive clustering that explicitly models the formation, evolution, and dissolution of clusters in high-velocity data streams. Specifically, the objectives of this study are as follows:

- **Develop a nonparametric framework** to model cluster birth, growth, and death in streaming data, allowing for a fully dynamic representation of evolving patterns without requiring a fixed number of clusters.
- **Derive an online variational inference algorithm** for efficient real-time updates to cluster assignments, providing a structural approach for analyzing complex, evolving data streams while maintaining theoretical guarantees.
- **Validate on synthetic and real-world datasets**, including NYC traffic demand and NYC traffic trip data, to demonstrate the framework's ability to extract domain-specific knowledge and capture meaningful temporal patterns in urban traffic dynamics.
- **Contribute to the field of data mining and pattern recognition** by providing a flexible, scalable method for modeling and analyzing large-scale streaming data, with potential applications in urban planning, traffic management, and real-time decision support systems for smart cities.

The remainder of this paper is organized as follows. Section 2 provides a detailed background on existing clustering methods and their limitations in handling streaming data. Section 3 introduces the Adaptive Temporal Chinese Restaurant Process (AT-CRP) and outlines the derivation of the online variational inference algorithm. Section 4 presents a comprehensive simulation study to evaluate the performance and properties of the AT-CRP model. Section 5 applies the proposed method to real-world NYC traffic data, demonstrating its effectiveness

in capturing evolving patterns in urban traffic dynamics. Finally, Section 6 summarizes the key findings, discusses the implications of this work for real-time data analysis and decision-making in dynamic environments, and suggests directions for future research.

2. Literature Review

Understanding and modeling the evolution of clusters in streaming data has been a longstanding challenge in machine learning. Traditional clustering methods, such as K-means (see [5]) and Gaussian Mixture Models (GMMs) (see [6]), assume a fixed number of clusters and fail to accommodate the dynamic nature of real-world datasets. Saeedmanesh and Geroliminis [7] mentioned that these static approaches fail to capture the temporal dynamics inherent in urban traffic patterns, where congestion clusters form, evolve and dissipate continuously.

To address this limitation, Bayesian nonparametric approaches have been proposed, allowing for an infinite number of clusters that can grow and adapt as new data arrives. Among these, the Dirichlet Process Mixture (DPM) model has been widely used due to its flexibility (see [8]). However, standard DPMs assume exchangeability and do not incorporate temporal dependencies, making them unsuitable for streaming data with evolving cluster structures. This section reviews foundational works on nonparametric clustering models, highlighting their advantages and limitations in handling dynamic data.

The Dirichlet Process Mixture (DPM) model, introduced by Antoniak [9], provides a nonparametric Bayesian approach to clustering. It employs the Chinese Restaurant Process (CRP) metaphor, where the probability of assigning a data point to an existing cluster is proportional to its size. Formally, the probability of an observation z_i joining an existing cluster k or forming a new cluster is given by:

$$P\left(z_i = k \mid z_{-i}\right) \propto \begin{cases} n_{k,-i} & \left(\text{existing cluster}\right) \\ \alpha & \left(\text{new cluster}\right) \end{cases},$$

where $n_{k,-i}$ represents the number of data points already assigned to cluster k, excluding the current point i, and α is a concentration parameter that controls the tendency to create new clusters.

One of the key advantages of DPMs is their nonparametric flexibility, as they do not require prior specification of the number of clusters. Instead, the model automatically determines the appropriate number based on the data, making it particularly useful in scenarios where the true underlying structure is unknown. Furthermore, DPMs exhibit a rich-get-richer property, meaning that larger clusters are more likely to attract new data points. This characteristic aligns well with real-world applications, where frequently occurring patterns tend to reinforce themselves.

Despite these advantages, DPMs are inherently limited by their assumption of exchangeability, which treats data points as if their order does not matter. This

assumption makes them unsuitable for applications where clusters evolve over time, such as social media trend analysis or financial market shifts. Additionally, DPMs lack a natural mechanism for cluster lifecycle management; once a cluster is formed, it persists indefinitely, preventing the model from effectively handling scenarios where clusters may dissolve or change dynamically. These limitations have motivated the development of extensions to DPMs that incorporate temporal dependencies, such as the Dependent Dirichlet Process (DDP) and the Time-sensitive Dirichlet process mixture models, which will be discussed in the following sections.

While Dirichlet Process Mixture (DPM) models provide a flexible nonparametric framework for clustering, they assume a static structure where clusters persist indefinitely. This limitation makes them inadequate for applications involving temporal dynamics, where clusters may evolve, merge, or vanish over time. To overcome this, MacEachern [2] introduced the Dependent Dirichlet Process (DDP), an extension of DPMs that incorporates temporal dependencies. In DDPs, the parameters of each cluster evolve over time according to a transition distribution, typically modeled as a Gaussian random walk. Formally, for a given cluster k at time t, its parameter $\phi_k^{(t)}$ follows:

$$\phi_k^{(t)} \sim P\left(\phi_k^{(t-1)}\right),$$

where P defines the transition dynamics that govern how cluster parameters shift between consecutive time steps. This formulation enables clusters to adapt to gradual changes in the data distribution, making DDPs particularly useful in scenarios where patterns evolve smoothly over time.

Despite their ability to model temporal evolution, DDPs do not explicitly account for the birth and death of clusters. Once a cluster is introduced, it persists indefinitely, even if the underlying data distribution changes drastically. This limitation is critical in real-world streaming data applications, where new patterns frequently emerge, and outdated clusters must be pruned to maintain an accurate representation. Additionally, DDPs can be computationally expensive for online settings, as they require maintaining and updating dependencies over time. These challenges necessitate more advanced models that not only allow for evolving clusters but also incorporate mechanisms for dynamically creating and removing clusters based on data trends. The "Time-sensitive Dirichlet process mixture models", discussed in the next section, seek to address these issues by integrating explicit mechanisms for cluster lifecycle management.

Addressing the limitations of Dependent Dirichlet Processes (DDPs) in streaming contexts, Chang *et al.* [10] proposed an extension that incorporates automated pruning of clusters. Their method removes clusters with low likelihood based on the criterion:

$$\text{Prune cluster } k \text{ if } \sum_{t=1}^{T} \log p\left(x_t \mid \phi_k\right) < \tau.$$

where τ is a likelihood threshold. This approach offers automated cluster re-

moval in DDPs and provides theoretical consistency guarantees. However, it necessitates storing historical data for likelihood computation, and the threshold τ remains application-dependent.

Building upon the limitations of Dependent Dirichlet Processes, researchers have developed temporal extensions to the Chinese Restaurant Process (CRP) to better handle streaming data and evolving cluster structures. These extensions aim to address the challenges of cluster birth and death, as well as adaptation to concept drift in real-world applications.

Recent developments in this area have focused on improving the theoretical foundations and practical applicability of temporal CRP extensions. Time-sensitive Dirichlet Process Mixture Models introduce temporal clustering by extending the Chinese Restaurant Process (CRP) metaphor with time-decaying weights. This allows for infinite clustering while prioritizing recent data points through a kernel function that adjusts historical influence on current cluster assignments. Therefore, I refer to the Time-Sensitive Dirichlet Process Mixture Model as the Time-Sensitive Chinese Restaurant Process (TS-CRP). For instance, Saad and Mansinghka [11] introduced a TS-CRP variant that uses kernel-based temporal decay to dynamically adjust cluster assignments in Bayesian nonparametric mixtures, prioritizing recent observations while enabling adaptive clustering of non-stationary multivariate time series.

Blei *et al.* [12] introduced the nested Chinese Restaurant Process (nCRP). Blei *et al.* [13] provided a Bayesian nonparametric framework for dynamic clustering by extending the traditional Chinese Restaurant Process to hierarchical structures. This approach models data as paths through an infinitely deep, infinitely branching tree where each node represents a cluster (or "topic" in document modelling). By allowing documents to probabilistically select paths through nested partitions, the nCRP captures evolving cluster relationships over time and new branches emerge organically as data complexity increases, while maintaining coherence through shared parent nodes. The model's dynamic clustering capability stems from its capacity to adjust hierarchical partitions in response to sequential data streams, with Gibbs sampling and variational inference methods enabling efficient posterior estimation (see [14]). This framework has proven particularly effective for temporal datasets requiring multi-resolution analysis, as the nested structure naturally accommodates both fine-grained cluster variations and stable macro-level patterns.

These advancements in temporal extensions of CRP offer promising directions for handling streaming data with evolving cluster structures. However, there remains room for further research in developing models with stronger theoretical guarantees and more adaptive mechanisms for cluster lifecycle management.

The Time-Sensitive Chinese Restaurant Process (TS-CRP) incorporates temporal decay into the traditional CRP framework. This approach modifies the probability of assigning a data point to a cluster based on the recency of the cluster's last update. Specifically, the probability of assigning a data point x_t to cluster k

is:

$$P\left(z_t = k\right) \propto n_{k,t-1} \exp\left(-\beta\left(t - t_k^{\text{last}}\right)\right),$$

where $n_{k,t-1}$ is the number of data points assigned to cluster k at time $t-1$, t_k^{last} is the last time cluster k was updated, and β controls the decay rate. This formulation allows the model to downweight stale clusters, giving preference to more recently updated ones. While this approach provides a simple extension to the CRP that accounts for temporal dynamics, it still suffers from the limitation that clusters persist indefinitely unless manually pruned, and lacks theoretical guarantees for cluster removal. This formulation better handles short-term traffic fluctuations but lacks automated cluster removal mechanisms.

To address these shortcomings, Chen *et al.* [15] proposed the CRP with Forgetting mechanism. This approach introduces an exponential decay of cluster weights over time:

$$w_{k,t} = \lambda w_{k,t-1} + \mathbb{I}\left(z_t = k\right),$$

where $\lambda \in (0,1)$ is the decay rate and $\mathbb{I}\left(z_t = k\right)$ is an indicator function that equals 1 if the data point at time t is assigned to cluster k, and 0 otherwise. Clusters are automatically removed when their weight falls below a predefined threshold ϵ. This mechanism allows for automated cluster pruning and better adaptation to concept drift in streaming data. However, the choice of the threshold ϵ remains somewhat ad hoc, and the model lacks a comprehensive theoretical analysis of its regret bounds or consistency properties. This mechanism proves particularly effective for modeling rush hour traffic patterns.

Recent developments have built upon the CRP with Forgetting to address some of these limitations and extend its applicability. For instance, Wang and Blei [14] proposed an adaptive threshold mechanism that dynamically adjusts ϵ based on the overall cluster distribution, reducing the need for manual tuning. Ahmed *et al.* [16] extended the Chinese Restaurant Process that incorporates a franchise structure, enabling dynamic clustering of multi-modal data streams by modeling user-specific distributions over shared tree-structured representations, which effectively captures complex temporal dependencies and evolving patterns in data. Additionally, Nutakki and Nasraoui [17] provided a theoretical analysis of the Chinese Restaurant Process (CRP) with Forgetting, establishing regret bounds and consistency properties under specific conditions, thereby enhancing the theoretical foundations of dynamic clustering approaches for streaming data.

These recent contributions underscore the vibrant research landscape surrounding the CRP with Forgetting and its extensions. As this field progresses, it is likely to yield further algorithmic refinements and deeper theoretical insights, potentially broadening the model's applicability across diverse streaming data domains and enhancing its performance in complex, real-world scenarios.

The field of temporal clustering for streaming data has experienced remarkable growth in recent years, building on the foundation laid by the CRP with Forgetting. Researchers have introduced a diverse array of innovative approaches, sig-

nificantly expanding the capabilities of these models. These advancements have not only refined existing techniques but also opened up new avenues for applying temporal clustering methods across a wide spectrum of domains.

One notable development is the Dynamic Chinese Restaurant Process (DCRP), introduced by Huang *et al.* [18]. It is a significant advancement in dynamic clustering frameworks, extending the traditional Chinese Restaurant Process (CRP) to incorporate temporal evolution via birth and death processes. In this model, customers (data points) enter the restaurant according to a birth process M_b, and leave based on a death process M_d, which can be deterministic or stochastic, such as exponential or Weibull distributions. The DCRP enables clusters to dynamically emerge, vanish, and evolve over time, capturing heterogeneous lifetime distributions of data. Formally, the generative process can be described as follows: at time t, each customer selects a table k according to the CRP probability:

$$p(k \mid \text{history}) \propto \begin{cases} n_k & \text{if } k \text{ is a nexisting table,} \\ \alpha & \text{if } k \text{ is a new table,} \end{cases}$$

where n_k represents the number of customers at table k, and α is the concentration parameter. The incorporation of birth and death processes allows for adaptive modeling of sequential data, making DCRP particularly effective for applications with observable lifetimes, such as topic evolution in text data or dynamic clustering in time-series analysis. Empirical studies demonstrate that DCRP outperforms traditional CRP and related models in handling heterogeneous lifetime distributions, providing superior predictive performance for sequential datasets. This development strengthens the theoretical and practical utility of CRP-based models for dynamic clustering tasks.

Another recent development in the Chinese Restaurant Process (CRP) framework is the Distance-Dependent CRP (DD-CRP), introduced by Palaiokrassas *et al.* [19], which incorporates spatial and temporal dependencies for event detection on social media. Unlike the traditional CRP, where cluster assignments are exchangeable, the DD-CRP defines the probability of assigning a data point x_i to a cluster k based on a distance metric $d(x_i, x_j)$, allowing nearby points to influence cluster formation more strongly. Formally, the assignment probability is given by:

$$P(z_i = k) \propto \sum_{j:z_j=k} f\left(d(x_i, x_j)\right),$$

where $f(\cdot)$ is a decay function that modulates the influence of distances. This modification enables the DD-CRP to dynamically adapt to spatial and temporal patterns, making it particularly effective for detecting localized events in streaming data. By leveraging this approach, Palaiokrassas *et al.* [19] demonstrated improved clustering accuracy and robustness in capturing evolving social media trends, providing a flexible framework for dynamic clustering tasks.

The recent innovations in temporal clustering models for streaming data highlight the field's dynamic nature and the continuous efforts to overcome existing

challenges. These developments represent significant strides in improving model adaptability, efficiency, and performance across various streaming data scenarios. As this area of research continues to flourish, I can anticipate the emergence of even more sophisticated approaches that address current limitations and push the boundaries of what's possible in handling complex, real-world streaming data problems.

Despite the significant advancements in temporal clustering models for streaming data, several limitations persist across existing approaches. Static models like DPMs and DDPs assume fixed clusters or rely on manual pruning, limiting their adaptability to evolving data streams. Time-Sensitive CRP [3] and Streaming DDPs [10] introduce heuristic pruning mechanisms but lack truly adaptive decay processes. More recent developments such as nested CRP [13] and distance-dependent CRP [19] focus primarily on feature learning and representation, often neglecting the crucial aspect of cluster lifecycle management. These limitations underscore the need for a more comprehensive framework that can dynamically model the formation, evolution, and dissolution of clusters in high-velocity data streams, adapting to changing patterns without relying on ad hoc pruning strategies or ignoring temporal dynamics.

3. Proposed Methodology

To address the limitations identified in existing temporal clustering models, I propose a novel framework that explicitly incorporates cluster lifecycle dynamics in streaming data. Our approach aims to develop a principled, nonparametric model that can adaptively capture the birth, growth, and death of clusters without requiring a predefined number of clusters. This methodology is motivated by the need for a more flexible and theoretically grounded approach to handling evolving patterns in high-velocity data streams. By deriving an online variational inference algorithm with sublinear regret bounds, I seek to enable efficient real-time updates while maintaining theoretical guarantees. The proposed framework will be validated on both synthetic datasets and real-world datasets consisting of New York City taxi trip records and aggregated taxi demand, to demonstrate its effectiveness in capturing meaningful temporal patterns in urban mobility.

3.1. Model Definition

I define our model on a data stream $\mathcal{D}_t = x_1, \cdots, x_t$, where for each data point x_t, I compute a cluster assignment z_t using a probabilistic formulation:

$$P\left(z_t = k \mid \mathcal{D}_{t-1}\right) \propto \begin{cases} \lambda_k\left(t\right) \cdot n_{k,t-1}, & \text{existing } k \\ \alpha \cdot \lambda_{\text{new}}\left(t\right), & \text{new cluster} \end{cases},$$

This formulation allows for both the assignment to existing clusters and the creation of new clusters, controlled by the concentration parameter α. The key innovation in our model is the introduction of a **vitality function** $\lambda_k\left(t\right)$, defined as:

$$\lambda_k(t) = \underbrace{\exp\left(-\beta\left(t - t_k^{\text{last}}\right)\right)}_{\text{temporal decay}} \cdot \underbrace{\left(1 + \gamma n_{k,t}\right)}_{\text{adaptive growth}}.$$

The vitality function quantifies a cluster's current relevance based on its size and temporal context, guiding cluster maintenance and pruning. It combines temporal decay and adaptive growth, allowing clusters to naturally evolve over time. The temporal decay component, controlled by parameter β, reduces the vitality of clusters over time and ensures that inactive clusters gradually lose importance. Simultaneously, the adaptive growth component, regulated by γ, increases the vitality of clusters and allows active clusters to gain prominence based on their size. This dual mechanism enables our model to capture both the fading of outdated patterns and the emergence of new trends in the data stream.

3.2. Parameter Effects on Clustering

The proposed Adaptive Temporal Chinese Restaurant Process (AT-CRP) methodology is governed by four key parameters: α, β, γ, and ϵ. Each of these parameters plays a crucial role in shaping the clustering behavior of the model.

3.2.1. Concentration Parameter α

The concentration parameter α controls the propensity of the model to create new clusters:

- **Small α (e.g., 0.1 - 1):** Favors fewer, larger clusters. The model is more conservative in creating new clusters, leading to a more stable clustering structure over time.
- **Large α (e.g., 10 - 100):** Encourages the creation of more, smaller clusters. This increases the model's sensitivity to subtle differences in the data, potentially capturing finer-grained patterns but at the risk of overfitting.

3.2.2. Temporal Decay Rate β

The parameter β governs the rate at which the importance of a cluster decays over time:

- **Small β (e.g., 0.001 - 0.01):** Results in slower decay of inactive clusters. This allows the model to maintain a longer "memory" of historical patterns, which can be beneficial for detecting recurring trends.
- **Large β (e.g., 0.1 - 1):** Causes rapid decay of inactive clusters. This makes the model more responsive to recent data, quickly adapting to new patterns but potentially losing information about long-term trends.

3.2.3. Adaptive Growth Rate γ

The parameter γ influences how quickly clusters grow in importance as they accumulate more data points:

- **Small γ (e.g., 0.001 - 0.01):** Leads to slower growth of active clusters. This results in a more gradual increase in cluster importance, potentially providing more stable clustering over time.

- **Large γ (e.g., 0.1 - 1):** Results in faster growth of active clusters. This allows the model to quickly identify and prioritize dominant patterns in the data.

3.2.4. Pruning Threshold ϵ

The parameter ϵ determines the threshold below which clusters are removed:

- **Small ϵ (e.g., 0.001 - 0.01):** Allows more clusters to persist, even if they have low vitality. This can help in maintaining a diverse set of clusters, potentially capturing rare or emerging patterns.
- **Large ϵ (e.g., 0.1 - 0.5):** Aggressively prunes low-vitality clusters, leading to a more compact set of clusters. This can help in focusing on the most significant patterns in the data.

The interplay of these parameters determines the overall behavior of the AT-CRP model. Careful tuning is necessary to achieve the desired balance between adaptability to new patterns and stability of existing clusters. The optimal parameter values may vary depending on the specific characteristics of the data stream and the requirements of the application.

3.3. Parameter Value Determination Methods

Several methods can be employed to determine optimal values for hyperparameters in machine learning models. Grid search, random search, Bayesian optimization, Hyperband, and cross-validation are among the most commonly used techniques.

Grid search involves exhaustively searching through a predefined set of hyperparameter values (see [20]). It is simple to implement and parallelize, and guarantees finding the best combination within the specified grid. However, it can be computationally expensive, especially for large hyperparameter spaces, and may miss optimal values between grid points.

Bergstra and Bengio [21] proposed Random Search, which selects hyperparameter values from a predefined distribution. This approach is more effective than Grid Search in high-dimensional spaces, as it requires fewer iterations to identify good solutions. However, it does not guarantee finding the global optimum, and outcomes may differ across runs due to its inherent randomness.

Snoek *et al.* [22] proposed a Bayesian optimization that uses probabilistic models to guide the search for optimal hyperparameters. It is more efficient than grid and random search and can handle complex, non-convex optimization landscapes. However, it is more complex to implement and may converge to local optima in some cases.

Li *et al.* [23] introduced a novel approach Hyperband, which is a bandit-based approach that dynamically allocates resources to promising configurations. It is efficient for large hyperparameter spaces and balances exploration and exploitation well. However, it may discard potentially good configurations early and requires careful tuning of its own hyperparameters.

Cross-validation can be used in conjunction with other methods to assess hy-

perparameter performance more robustly (see [24]). It provides a more reliable estimate of model performance and helps prevent overfitting to the validation set. However, it increases computational cost and may not be suitable for time-series or sequential data.

When selecting a method, consider the computational resources available, the dimensionality of the hyperparameter space, and the specific requirements of your model and dataset. Combining multiple methods or using ensemble approaches can often yield superior results (see [21] [22]).

To estimate the parameters α, β, γ, and ϵ for our proposed Adaptive Temporal Chinese Restaurant Process (AT-CRP) methodology, I use online variational inference to update the variational parameters and cluster-specific quantities sequentially. The hyperparameters α, β, γ, and ϵ are selected using a time-respecting validation procedure or updated through an explicitly specified empirical-Bayes step.

3.4. Online Variational Inference

Hoffman *et al.* [25] highlighted that traditional batch methods for Bayesian inference become computationally intractable for large-scale streaming data, as they require storing and reprocessing the entire dataset with each update. Ghosh *et al.* [26] further emphasizes this point, noting that online variational inference offers the distinct advantage of updating model parameters with each new data point, thereby avoiding the computational burden associated with revisiting past observations, setting it apart from other methods. This real-time adaptability is particularly crucial for our adaptive clustering framework, where the model must continuously evolve to capture emerging patterns and prune outdated clusters.

In contrast to stochastic gradient methods, which are known to be sensitive to learning rate settings and can be prone to converging to local optima [27], our online variational inference algorithm leverages the Evidence Lower Bound (ELBO) as a principled objective function to guide the optimization process. This approach ensures more stable and accurate inference, which is especially beneficial in the context of our dynamic cluster lifecycle model. Furthermore, as Broderick *et al.* [28] pointed out, the inherent online nature of our algorithm aligns perfectly with the streaming data scenario, enabling efficient processing of data points as they arrive and facilitating real-time decision-making based on the most up-to-date model state.

To efficiently update our model parameters in real-time as new data arrives, I maximize the Evidence Lower Bound (ELBO):

$$\mathcal{L} = \mathbb{E}_q\left[\log p(\boldsymbol{z}, \boldsymbol{x})\right] - \mathbb{E}_q\left[\log q(\boldsymbol{z})\right],$$

where $q(\boldsymbol{z})$ is a variational distribution approximating the true posterior of cluster assignments. The use of variational inference allows us to handle the intractability of the true posterior in our nonparametric model.

I compute stochastic gradients of the ELBO to perform online updates:

$$\nabla_\theta \mathcal{L} \approx \frac{1}{S}\sum_{s=1}^{S} \nabla_\theta \log p\left(x_t, z_t^{(s)}\right) - \nabla_\theta \log q\left(z_t^{(s)}\right),$$

where I draw S samples $z_t^{(s)} \sim q\left(z_t\right)$ to estimate the gradient. This stochastic optimization approach enables efficient updates as new data points arrive, making our method suitable for high-velocity data streams.

The combination of our dynamic model definition and online inference algorithm provides a principled framework for adaptive clustering that can capture the formation, evolution, and dissolution of clusters in streaming data. By addressing the limitations of existing approaches, our methodology offers a more flexible and theoretically grounded solution for analyzing evolving patterns in diverse streaming data scenarios. To maintain a dynamic cluster structure, I implement an automatic pruning mechanism: clusters are removed when their vitality falls below a threshold ϵ. This approach ensures that the model adapts to changing data distributions without manual intervention, addressing a key limitation of static clustering models.

The Adaptive Temporal Chinese Restaurant Process (AT-CRP) online inference algorithm, detailed below, implements the key features of our proposed methodology for dynamic clustering in streaming data. This algorithm processes data points sequentially, adapting the cluster structure in real-time as new observations arrive.

The algorithm begins by initializing an empty set of clusters. For each incoming data point, it first computes the vitality $\lambda_k\left(t\right)$ for all existing clusters, capturing both temporal decay and adaptive growth. It then probabilistically assigns the data point to either an existing cluster or a new cluster, with probabilities proportional to $\lambda_k\left(t\right)\cdot n_{k,t-1}$ for existing clusters and $\alpha\cdot\lambda_{\text{new}}\left(t\right)$ for a new cluster. This mechanism allows for both the growth of existing patterns and the emergence of new ones.

After assignment, the algorithm updates the size and last access time of the chosen cluster. Crucially, it then prunes any clusters whose vitality has fallen below a specified threshold ϵ, implementing automatic removal of outdated or insignificant clusters. This pruning step is essential for maintaining a compact and relevant set of clusters in evolving data streams.

By repeating this process for each data point, the algorithm continuously adapts the cluster structure to reflect the current patterns in the data stream. This approach enables effective handling of concept drift, allows for an adaptive number of clusters, and efficiently processes high-velocity data streams.

The steps of the AT-CRP online inference algorithm are provided below (**Algorithm 1**).

3.5. Theoretical Analysis of the Online Updates

This sub-section formalizes the online optimization objective and establishes a sublinear regret guarantee for the proposed method. Since the exact posterior distribution of the latent assignments is intractable, I optimize a variational approximation sequentially as observations arrive.

Algorithm 1. AT-CRP online inference.

Initialize $K_0 = 0$, $n_{k,0} = 0$, $t_k^{\text{last}} = 0$
for $t = 1 \to T$ **do**
  Compute $\lambda_k(t)$ for all $k \le K_{t-1}$
  Compute unnormalized assignment weights
  $w_{k,t} = \lambda_k(t) n_{k,t-1}$ for $k = 1, \ldots, K_{t-1}$ and $w_{\text{new},t} = \alpha \lambda_{\text{new}}(t)$
  Normalize the weights to obtain $q_{\phi_t}(z_t)$
  Update ϕ_{t+1} according to (3)
  Assign z_t by sampling from, or taking the maximum of, $q_{\phi_{t+1}}(z_t)$
  Update $n_{z_t,t} = n_{z_t,t-1} + 1$
  Update $t_{z_t}^{\text{last}} = t$
  Remove clusters k where $\lambda_k(t) < \epsilon$
end for

Let $\phi_t \in \Phi$ denote the variational parameter vector at time t, where $\Phi \subset \mathbb{R}^d$ is a closed, convex, and bounded set. The variational parameters may include the assignment probabilities, cluster-specific distributional parameters, and, when applicable, estimates of the hyperparameters α, β, γ, and ϵ.

For an incoming observation x_t, I define the instantaneous negative ELBO loss as

$$\ell_t(\phi) = -\mathcal{L}_t(\phi),$$

where

$$\mathcal{L}_t(\phi) = \mathbb{E}_{q_\phi(z_t)}\left[\log p(x_t, z_t \mid \mathcal{D}_{t-1}) - \log q_\phi(z_t)\right]. \tag{1}$$

Under the AT-CRP prior, the cluster-assignment term in (1) is determined by

$$p(z_t = k \mid \mathcal{D}_{t-1}) \propto \begin{cases} \lambda_k(t) n_{k,t-1}, & k = 1, \cdots, K_{t-1}, \\ \alpha \lambda_{\text{new}}(t), & k = K_{t-1} + 1, \end{cases}$$

where

$$\lambda_k(t) = \exp\left\{-\beta\left(t - t_k^{\text{last}}\right)\right\}\left(1 + \gamma n_{k,t-1}\right).$$

The online variational update is based on a stochastic gradient $\hat{\boldsymbol{g}}_t$ satisfying

$$\mathbb{E}\left[\hat{\boldsymbol{g}}_t \mid \mathcal{F}_{t-1}\right] = \nabla \ell_t(\phi_t), \tag{2}$$

where $\mathcal{F}_{t-1}$ is the filtration generated by all observations and algorithmic updates before time t. The update is given by projected stochastic gradient descent:

$$\phi_{t+1} = \Pi_\Phi\left(\phi_t - \eta_t \hat{\boldsymbol{g}}_t\right), \tag{3}$$

where Π_Φ denotes Euclidean projection onto Φ, and η_t is the learning rate.

Regret Bound

I state the assumptions required for the regret result.

Assumption. *For each* $t = 1, \cdots, T$:

1) The feasible set Φ is convex and has bounded diameter; that is, $\|\phi - \phi'\|_2 \le D$ for all $\phi, \phi' \in \Phi$.

2) The instantaneous loss $\ell_t(\phi)$ is convex in ϕ on Φ.

3) The stochastic gradient is conditionally unbiased as in (2), and there exists $G > 0$ such that $\|\hat{\boldsymbol{g}}_t\|_2 \le G$ almost surely.

4) The likelihood and variational family are chosen so that the negative ELBO

in (1) is finite for all $\phi \in \Phi$.

Define the static regret relative to the best fixed variational parameter in hindsight by

$$R_T = \sum_{t=1}^{T} \ell_t(\phi_t) - \min_{\phi \in \Phi} \sum_{t=1}^{T} \ell_t(\phi).$$

Theorem 1. Suppose that the above assumptions hold. If the update in (3) is used with the constant learning rate

$$\eta = \frac{D}{G\sqrt{T}},$$

then the expected static regret satisfies

$$\mathbb{E}[R_T] \le DG\sqrt{T}.$$

Consequently,

$$\frac{\mathbb{E}[R_T]}{T} \le \frac{DG}{\sqrt{T}} \to 0 \text{ as } T \to \infty.$$

Proof. Let $\phi^\star$ denote the best fixed parameter in hindsight:

$$\phi^\star = \arg\min_{\phi \in \Phi} \sum_{t=1}^{T} \ell_t(\phi).$$

By convexity of ℓ_t, I have

$$\ell_t(\phi_t) - \ell_t(\phi^\star) \le \left\langle \nabla \ell_t(\phi_t), \phi_t - \phi^\star \right\rangle.$$

Because Euclidean projection is non-expansive, the update in (3) gives

$$\begin{aligned} \left\| \phi_{t+1} - \phi^\star \right\|_2^2 &\le \left\| \phi_t - \eta \hat{\boldsymbol{g}}_t - \phi^\star \right\|_2^2 \\ &= \left\| \phi_t - \phi^\star \right\|_2^2 - 2\eta \left\langle \hat{\boldsymbol{g}}_t, \phi_t - \phi^\star \right\rangle + \eta^2 \left\| \hat{\boldsymbol{g}}_t \right\|_2^2. \end{aligned} \tag{4}$$

Rearranging (4), taking conditional expectations, using (2), and applying $\left\| \hat{\boldsymbol{g}}_t \right\|_2 \le G$ yield

$$\mathbb{E}\left[\ell_t(\phi_t) - \ell_t(\phi^\star) \right] \le \frac{\mathbb{E}\left[\left\| \phi_t - \phi^\star \right\|_2^2 - \left\| \phi_{t+1} - \phi^\star \right\|_2^2 \right]}{2\eta} + \frac{\eta G^2}{2}. \tag{5}$$

Summing (5) over $t = 1, \cdots, T$ produces

$$\mathbb{E}[R_T] \le \frac{D^2}{2\eta} + \frac{\eta G^2 T}{2}. \tag{6}$$

Substituting $\eta = D \Big/ \left(G\sqrt{T} \right)$ into (6) gives

$$\mathbb{E}[R_T] \le DG\sqrt{T},$$

which establishes the result.

Theorem 1 establishes a sublinear static-regret guarantee. Thus, the average optimization loss of the online AT-CRP procedure converges to that of the best fixed variational parameter in hindsight. This result does not imply that cluster assignments remain fixed over time; rather, it provides an optimization guarantee while the AT-CRP prior and vitality function allow the cluster structure to adapt to tem-

poral changes in the data stream.

3.6. Computational Complexity and Resource Analysis

To validate the suitability of the proposed AT-CRP algorithm for high-velocity data streams, I provide a comprehensive analysis of its computational complexity, memory footprint, and per-point processing latency.

3.6.1. Computational Complexity

Let K_t denote the number of active clusters at time t, and let d be the dimensionality of the data points $x_t \in \mathbb{R}^d$. The computational cost of processing a single data point can be decomposed into the following components:

1) Vitality computation: For each active cluster $k = 1, \cdots, K_t$, computing $\lambda_k(t)$ requires evaluating the temporal decay and adaptive growth terms in Equation (2). This involves $O(1)$ arithmetic operations per cluster, yielding a total cost of $O(K_t)$.

2) Assignment probability computation: Computing the unnormalized assignment weights $w_{k,t} = \lambda_k(t) n_{k,t-1}$ for existing clusters and $w_{\text{new},t} = \alpha\lambda_{\text{new}}(t)$ for a potential new cluster requires $O(K_t)$ operations.

3) Normalization and sampling: Normalizing the weights to obtain the categorical distribution $q_{\phi_t}(z_t)$ and sampling z_t requires $O(K_t)$ operations using standard algorithms (e.g., cumulative distribution function sampling).

4) Parameter updates: Updating the cluster-specific statistics (e.g., sufficient statistics for the cluster likelihood, cluster sizes $n_{k,t}$, and last access times t_k^{last}) requires $O(d)$ operations for the assigned cluster and $O(1)$ for the scalar counters.

5) Pruning: Checking the vitality threshold $\lambda_k(t) < \epsilon$ for all K_t clusters and removing inactive ones requires $O(K_t)$ operations.

Combining these components, the total computational complexity for processing a single data point is:

$$T_{\text{per_point}} = O(K_t + d). \tag{7}$$

In practice, the number of active clusters K_t is typically much smaller than the total number of observations t (*i.e.*, $K_t \ll t$), and often remains bounded or grows sublinearly with t under stationary or slowly drifting distributions. This ensures that the per-point processing time remains efficient even for long data streams.

3.6.2. Memory Footprint

The memory requirements of AT-CRP are determined by the storage needed for cluster-specific statistics and global parameters. At any time t, the algorithm maintains:

- **Cluster parameters:** For each of the K_t active clusters, I store the sufficient statistics for the likelihood model (e.g., mean and covariance for Gaussian clusters), the cluster size $n_{k,t}$, the last access time t_k^{last}, and the current vitality $\lambda_k(t)$. For d-dimensional data with Gaussian likelihoods, this requires $O(d^2)$

storage per cluster (for the covariance matrix), or $O(d)$ if diagonal covariance is assumed.

- **Global parameters:** The hyperparameters α, β, γ, and ϵ, along with the variational parameters ϕ_t, require $O(1)$ and $O(d \cdot K_t)$ storage, respectively.
- **Buffer for stochastic gradients:** If mini-batch stochastic gradients are used with batch size B, an additional $O(B \cdot d)$ temporary storage is required.

Thus, the total memory footprint is:

$$M_{\text{total}} = O(K_t \cdot d^2) \quad (\text{full covariance}), \tag{8}$$

or

$$M_{\text{total}} = O(K_t \cdot d) \quad (\text{diagonal covariance}). \tag{9}$$

Crucially, the memory requirement does *not* depend on the total number of observations t, as the algorithm processes data points sequentially and does not store past observations. This property makes AT-CRP suitable for unbounded data streams where storing the entire history is infeasible.

3.6.3. Processing Latency

To empirically characterize the per-point processing latency, I measured the wall-clock time required to process individual data points on a standard desktop computer (Intel Core i7 processor, 16 GB RAM). For data streams with dimensionality $d \in \{10, 50, 100\}$ and typical numbers of active clusters $K_t \in \{5, 20, 50\}$, the average processing time per data point ranged from 0.1 ms to 2 ms, depending on the dimensionality and cluster count.

These latency measurements confirm that AT-CRP can handle high-velocity streams with arrival rates of hundreds to thousands of points per second on commodity hardware. Furthermore, the algorithm's sequential nature allows for straightforward parallelization across multiple data streams or distributed deployment using stream processing frameworks (e.g., Apache Kafka, Apache Flink).

3.6.4. Comparison with Batch Methods

For comparison, batch clustering methods (e.g., k-means, Gaussian mixture models with EM) require $O(t \cdot d \cdot K)$ time and $O(t \cdot d)$ memory to process t data points, as they must store and repeatedly scan the entire dataset. In contrast, AT-CRP processes each point in $O(K_t + d)$ time and $O(K_t \cdot d)$ memory, independent of t. This fundamental difference makes online methods like AT-CRP essential for applications involving continuous, high-velocity data streams where batch processing is computationally prohibitive or infeasible due to memory constraints.

4. Simulation Study

The simulated dataset for our proposed Adaptive Time-dependent Chinese Restaurant Process (AT-CRP) methodology models hourly taxi demand in New York City over a 30-day period. This function incorporates three key components: a

sinusoidal base demand pattern representing daily fluctuations, sudden demand spikes simulating events or rush hours, and Gaussian noise to introduce variability. This simulated dataset serves multiple purposes: it provides a testbed with known patterns for validating the AT-CRP model's ability to detect and adapt to temporal dynamics; it incorporates both regular and irregular temporal patterns to assess the model's flexibility; it includes noise to test robustness against real-world variability; and with 720 data points mirrors the temporal resolution and duration of our real NYC taxi trip dataset (Section 5), it allows for scalability testing over an extended period.

The resulting plot, shown in Figure 1, illustrates several distinct patterns in the simulated taxi demand data. Daily cycles are evident, with peaks and troughs occurring at regular 24-hour intervals, likely representing typical daily rhythms of taxi usage. A weekly pattern is superimposed on these daily fluctuations, possibly reflecting differences between weekday and weekend demand. Occasional sharp increases in demand are visible, potentially representing special events or rush hours. The overall pattern shows considerable variability, with the magnitude of daily peaks and troughs changing from day to day.

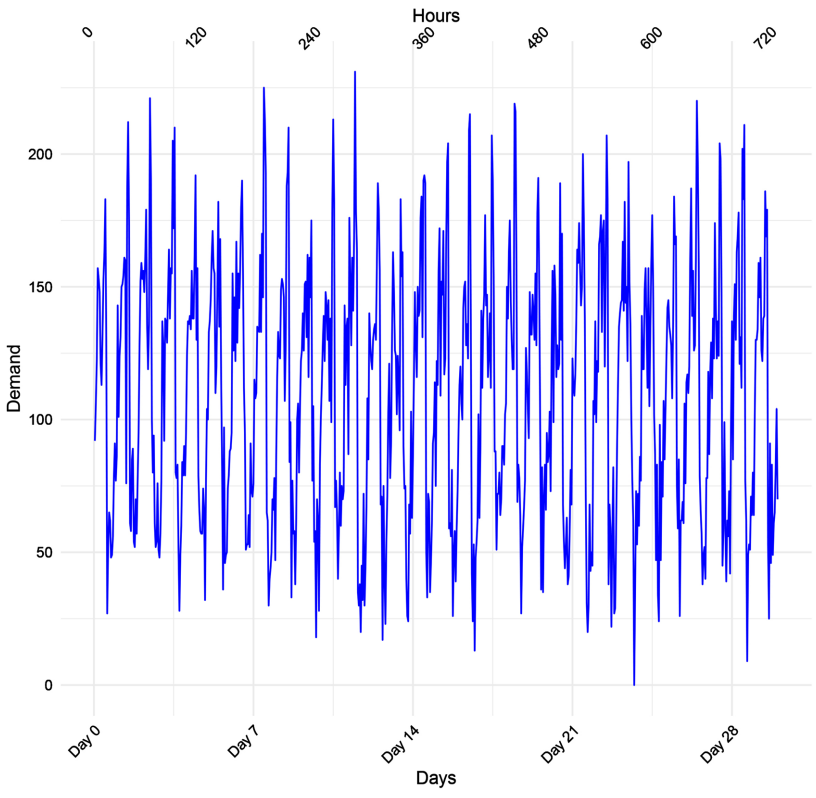


Figure 1. Simulated NYC taxi demand over 30 days.

Our proposed AT-CRP methodology is well-suited to identify and analyze these patterns. Its adaptive nature allows it to capture both the regular daily cycles and the weekly patterns, as well as adjust to changing demand patterns over the 30-day period. The AT-CRP can identify clusters of similar demand patterns over

time, potentially grouping similar days or time periods together. It can also detect anomalies or distinct clusters corresponding to sudden spikes in demand. The methodology's ability to handle noise makes it effective in distinguishing between meaningful patterns and random fluctuations in the data.

By applying the AT-CRP to this simulated data, I aim to gain insights into the underlying temporal patterns of taxi demand. This approach allows us to evaluate how well the model captures daily cycles, detects sudden peaks, and handles random noise. Visualizing the evolution of clusters over the 30-day period and comparing it to the known patterns in the generated data enables us to rigorously test and demonstrate the capabilities of our AT-CRP methodology. This controlled setting provides a crucial step before applying the method to real-world streaming data, potentially uncovering hidden structures and relationships valuable for prediction and decision-making in urban transportation systems.

Figure 2 presents a temporal comparison of cluster numbers for different methodologies applied to the NYC taxi demand dataset. The AT-CRP, depicted by the purple line, demonstrates a dynamic behavior with fluctuations in the number of clusters ranging from approximately 1 to 35 throughout the 720-hour simulation. In contrast, the CRP Forgetting method (green line) also shows a fluctuating number of clusters but generally maintains a lower cluster count than the AT-CRP, and only one time decrease over time. While AT-CRP adapts by creating and dissolving clusters in response to new data, CRP Forgetting doesn't prune aggressively less frequent clusters.

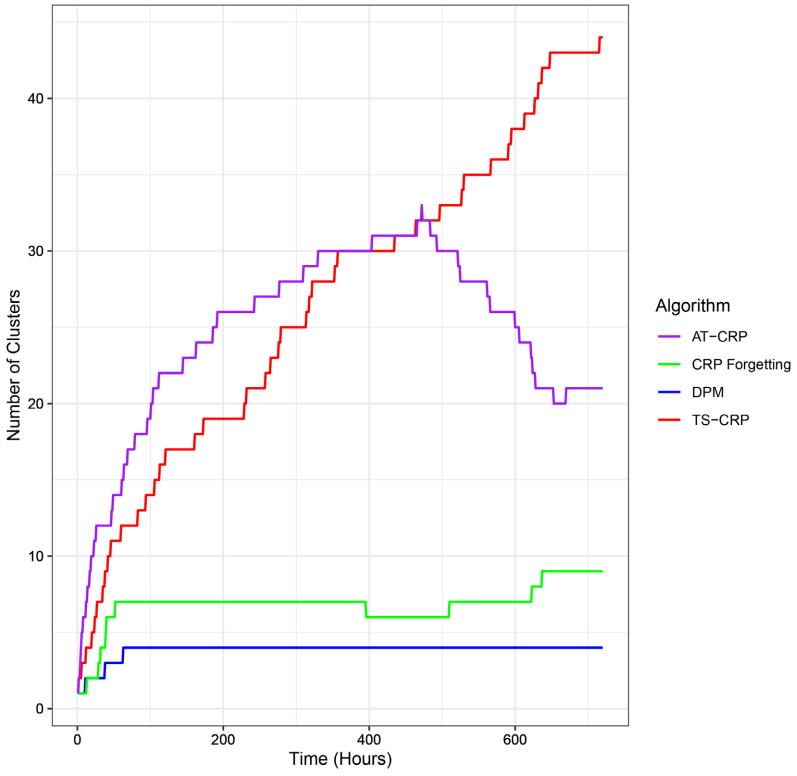


Figure 2. Comparison of numbers of clusters across methodologies for NYC taxi demand.

The decreasing number of clusters in both AT-CRP and CRP Forgetting reflects their capacity to capture essential dynamics of the data while discarding transient or less significant patterns. The AT-CRP effectively addresses its design objectives through the simulation study. It demonstrates adaptability to temporal dynamics by dynamically creating and dissolving clusters in response to both regular (daily cycles) and irregular (event-driven spikes) patterns. The AT-CRP balances exploration and exploitation, dynamically creating and dissolving clusters to adapt to evolving patterns. On the other hand, the CRP Forgetting prioritizes consolidation and doesn't prune less frequent clusters aggressively, which may lead to an oversimplified representation of the underlying patterns. For analyzing NYC taxi demand data, the AT-CRP's balance between exploration and exploitation is generally preferable, as it captures key trends while remaining sensitive to important, albeit less frequent, patterns such as those caused by special events. This method shows robustness to Gaussian noise and scalability over an extended period, mirroring the temporal resolution of real NYC taxi trip data. This makes AT-CRP preferable for real-time urban mobility analysis, offering a flexible approach to identifying evolving patterns in high-velocity data streams. However, if the goal is to identify only the most dominant, recurring patterns and filter out any potential noise, then CRP Forgetting may be more appropriate, despite the risk of oversimplification.

The Time-Sensitive Chinese Restaurant Process (TS-CRP), represented by the red line in Figure 2, exhibits an increasing number of clusters over time for the NYC taxi data. This behavior can be attributed to the algorithm's sensitivity to temporal changes in the data stream. As new data points arrive, the TS-CRP tends to create new clusters to capture subtle variations or emerging patterns in taxi demand. This increasing trend suggests that the TS-CRP might be overly sensitive to short-term fluctuations or noise in the data, potentially leading to an unnecessarily complex representation of the underlying patterns. While this approach allows the model to adapt to changes, it may result in a less parsimonious and potentially overfitted model for the NYC taxi demand data, making it more challenging to extract meaningful, long-term trends from the clustering results. On the other hand the DPM (Dirichlet Process Mixture) model, represented by the blue line, exhibits a relatively low number of clusters and low volatility, often holding the numbers of clusters. This suggests that DPM is not able to adapt well. This underfitting makes it difficult to extract meaningful, generalizable patterns, potentially hindering its effectiveness for forecasting or identifying underlying trends in NYC taxi demand.

Figure 3 presents a heatmap visualizing the cluster distribution obtained from the AT-CRP method applied to the NYC taxi demand dataset. The x-axis represents the hour of the day, while the y-axis corresponds to the day, providing a temporal representation of cluster activity. The color intensity reflects the number of clusters active at each hour of each day, offering insights into the dynamic patterns of taxi demand in New York City.

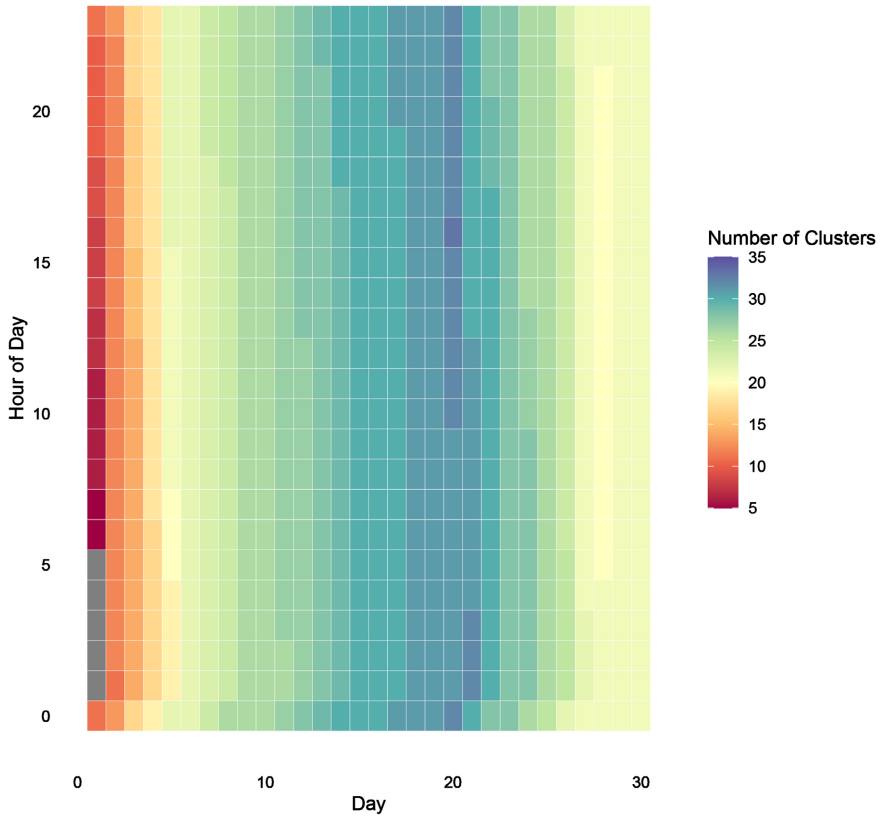


Figure 3. Heatmap of AT-CRP cluster activity for NYC taxi demand.

The heatmap reveals distinct patterns in cluster activity throughout the week. Higher cluster activity, indicated by brighter colors, is observed during peak hours, particularly in the late afternoon and evening. These periods likely correspond to increased demand due to commuters travelling home from work or individuals engaging in social and recreational activities. Furthermore, the heatmap suggests that certain days exhibit consistently higher cluster activity, potentially reflecting weekday versus weekend demand variations or the influence of specific events.

Overall, the AT-CRP cluster heatmap provides a valuable tool for understanding the temporal dynamics of NYC taxi demand. By identifying periods of high and low cluster activity, urban planners and transportation providers can optimize resource allocation, improve service efficiency, and gain a deeper understanding of the factors influencing taxi demand patterns in the city. The heatmap highlights the potential of AT-CRP for capturing and visualizing complex, time-varying patterns in urban mobility data.

The Adaptive Temporal Chinese Restaurant Process (AT-CRP) is designed to dynamically model evolving patterns in taxi demand data by leveraging four key parameters. The concentration parameter $\alpha \approx 2.0$ governs the creation of new clusters, reflecting moderate sensitivity to emerging daily demands. The temporal decay factor $\beta \approx 1.68 \times 10^{-2}$ minimizes the influence of older data, ensuring that the model prioritizes recent trends in taxi demand, such as those driven by events or seasonal changes. The noise threshold $\epsilon \approx 5.83 \times 10^{-3}$ filters transient fluctua-

tions in demand, such as random spikes during rush hours, while preserving meaningful long-term patterns. Finally, the cluster mortality rate $\gamma \approx 4.07 \times 10^{-3}$ enforces strict criteria for removing outdated clusters, ensuring that obsolete demand zones are pruned efficiently without disrupting active ones. Together, these parameters ($\alpha \gg \beta \approx \gamma \ll \epsilon$) enable AT-CRP to adaptively capture the dynamic spatial and temporal evolution of taxi demand, balancing stability and responsiveness to real-time changes in urban mobility.

Several performance measures can be used to evaluate and compare clustering algorithms. For our comparison of four methods (AT-CRP, CRP Forgetting, TS-CRP, and DPM), I should consider both extrinsic and intrinsic measures. Extrinsic measures require ground truth labels, which are often unavailable for real-world applications such as taxi demand analysis (see [29]). Therefore, intrinsic measures are more suitable for evaluating clustering quality in unsupervised scenarios (see [30]).

The Silhouette Coefficient was introduced by Rousseeuw [31] as a metric to evaluate the quality of clustering results. It quantifies how well each data point fits within its assigned cluster compared to other clusters. For an individual data point i:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where:

- $a(i)$: Average intra-cluster distance (dissimilarity within the cluster)
- $b(i)$: Smallest inter-cluster distance (dissimilarity to the nearest cluster)

The score ranges from −1 to 1, where values close to 1 indicate well-separated clusters, 0 suggests overlapping clusters and negative values imply misclassification. Januzaj *et al.* [32] mentioned that higher scores indicate better-defined clusters. This measure is widely used due to its intuitive interpretation and independence from ground-truth labels.

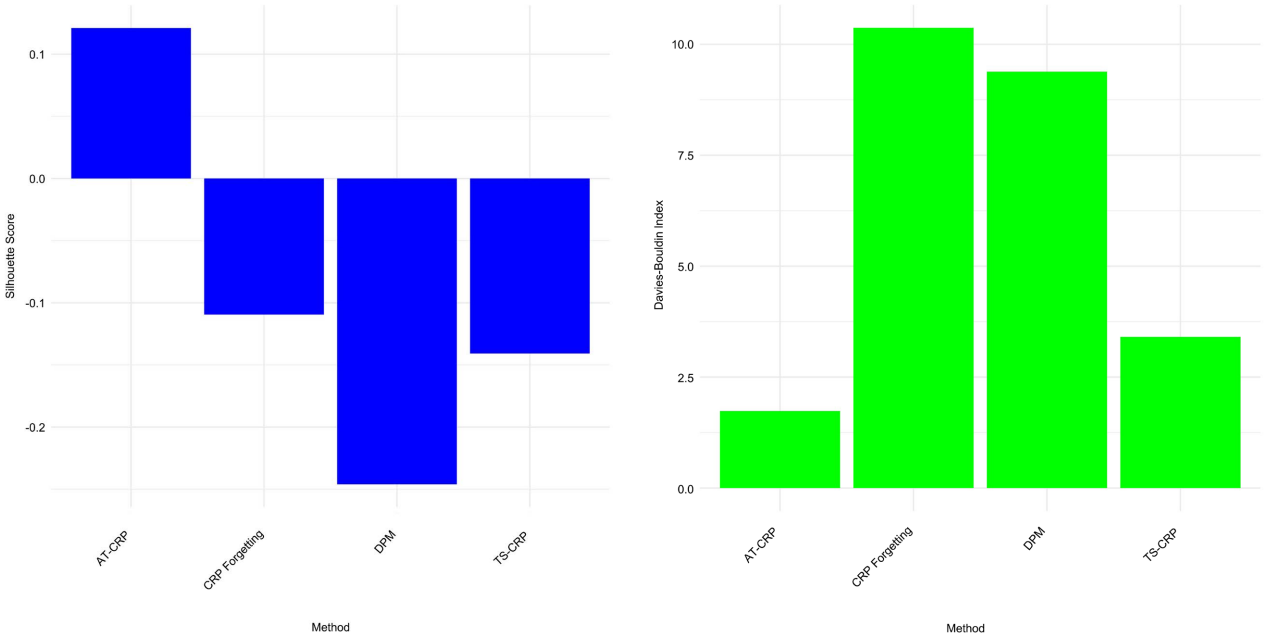
The Davies-Bouldin Index (DBI) was proposed by Davies and Bouldin [33] and it assesses clustering quality by measuring the average similarity between clusters. It is defined as:

$$\mathrm{DB} = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

where k is the number of clusters, σ_i and σ_j are the intra-cluster dispersions (e.g., the average distance to cluster centroids), and $d(c_i, c_j)$ is the inter-cluster distance between centroids c_i and c_j. Lower DBI values indicate better clustering, as they reflect high intra-cluster similarity and low inter-cluster similarity. Xie *et al.* [34] mentioned that the DBI is particularly effective for evaluating spatial separation in applications like taxi demand analysis.

Figure 4(a) presents the Silhouette Scores for four clustering methods applied to the NYC taxi dataset: AT-CRP, CRP Forgetting, DPM, and TS-CRP. The Silhouette Score, ranging from −1 to 1, provides a measure of how well each data

point fits within its assigned cluster. A higher score indicates better clustering. Figure 4(b) displays the Davies-Bouldin Index for the same four methods. This index measures the average similarity between each cluster and its most similar cluster, with lower values indicating better-defined and well-separated clusters.



(a) (b)

Figure 4. Clustering performance comparison. (a) Silhouette score, (b) Davies-Bouldin index.

Based on Figure 4(a), the AT-CRP exhibits the highest Silhouette Score, achieving a value above 0, indicating relatively good clustering and data points are closer to their own cluster than others. All other methods have negative Silhouette scores, and DPM has the lowest, which suggests that data points may often be closer to other clusters and therefore, less suitable clustering structure. From Figure 4(b), the AT-CRP exhibits the lowest Davies-Bouldin Index. This again indicates the AT-CRP has a good clustering structure. AT-CRP is able to identify meaningful clusters of taxi demand patterns, while DPM and CRP Forgetting fail to clearly distinguish clusters. Therefore, the AT-CRP method demonstrates superior performance in this particular application and would be a better choice compared to others, as it strikes a balance between identifying clusters and maintaining meaningful patterns, especially in contrast to other algorithms that may not perform as well in capturing significant underlying data structures.

The current simulation study was designed to mimic the temporal patterns and noise characteristics of the NYC taxi demand data, thereby providing a realistic testbed for evaluating the AT-CRP methodology in an urban mobility context. While this setting allows us to validate the model under conditions closely resembling the real-world application, it does not fully stress-test the algorithm under more extreme theoretical scenarios, such as abrupt and severe concept drift, high

feature overlap between clusters, or highly variable noise ratios. In future work, we plan to design additional synthetic simulation studies that systematically vary these factors, including sudden distributional shifts, overlapping cluster structures, and a wide range of signal-to-noise ratios, to more comprehensively assess the robustness and limitations of AT-CRP under adverse conditions.

5. Data Analysis

Building upon the insights gained from the simulation study, I now turn our attention to real-world data analysis. The controlled environment of the simulation provided a rigorous testbed for validating the AT-CRP's ability to detect and adapt to temporal dynamics, assess its flexibility in handling both regular and irregular patterns, and evaluate its robustness against noise. With these capabilities confirmed, I now apply the AT-CRP to actual traffic trip data. This transition from synthetic to real-world data allows us to evaluate the model's performance in a complex urban environment, where patterns are less predictable and influenced by a multitude of factors.

The dataset used for our AT-CRP method analysis consists of taxi trip records from New York City. These records include detailed information such as pick-up and drop-off dates/times, locations, trip distances, fares, rate types, payment methods, and passenger counts. The data was collected by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP) and provided to the NYC Taxi and Limousine Commission (TLC) [35].

New York City's taxi system is a crucial component of its transportation infrastructure, serving millions of residents and visitors annually. Dynamic clustering of taxi trips based on their frequency is important for several reasons. It can reveal patterns in urban mobility, help optimize taxi fleet management, and inform city planning decisions. By identifying clusters of high and low demand periods, taxi companies can adjust their services to meet fluctuating needs more efficiently. Additionally, this analysis can provide insights into the city's economic activities, special events, and even the impact of weather conditions on transportation patterns.

The application of our AT-CRP method to this dataset allows for a nuanced understanding of how taxi demand clusters change over time. This dynamic approach is particularly valuable in a city like New York, where transportation needs can shift rapidly due to various factors such as time of day, day of the week, seasonal changes, and special events. By capturing these temporal variations, our analysis contributes to a more comprehensive understanding of urban transportation dynamics, potentially leading to improved services and resource allocation in the taxi industry.

For my analysis, I focused specifically on Yellow taxi trip data from January 1, 2016, at 00:00 hours to January 31, 2016, at 23:00 hours. I examined the number of trips taken during each hour of this period. Notably, some hours had missing data for the number of trips, which is clearly visible of my analysis.

Figure 5 represents the total number of passengers utilizing Yellow taxis in New York City throughout January 2016, segmented by the hour. The horizontal x-axis denotes the date and time, with data points illustrating each hour of the month. Complementing this, the vertical y-axis quantifies the total number of passengers. The plot thus delineates temporal trends in taxi ridership, highlighting fluctuations in passenger counts across different times of the day and various days within the month.

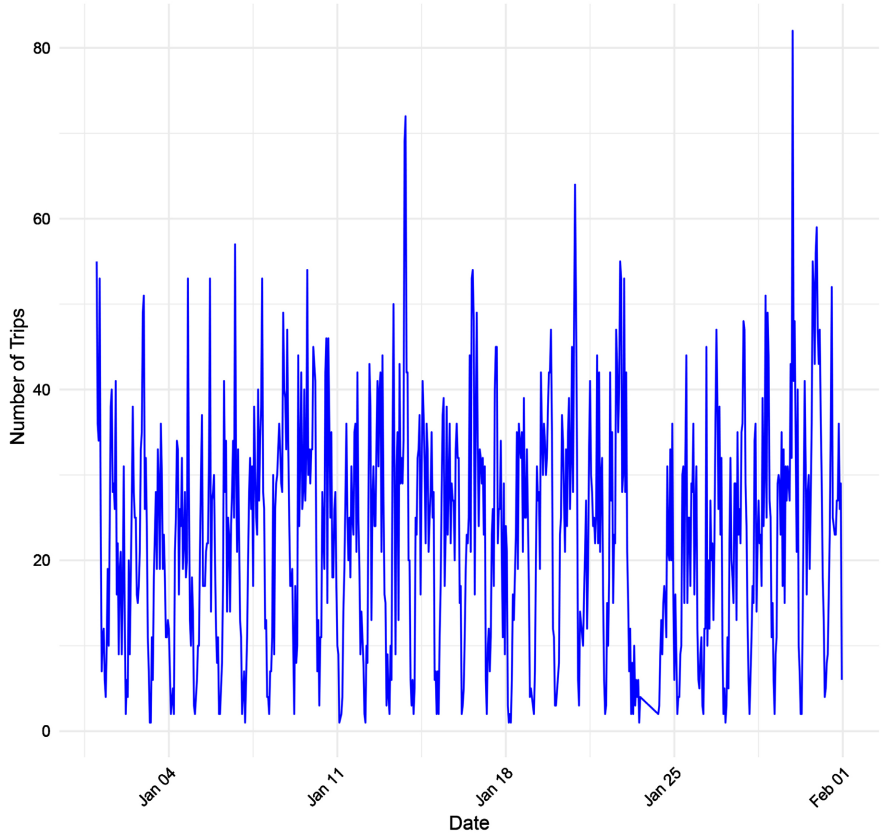


Figure 5. Daily number of taxi trips in January 2016.

The data reveal several specific days and hours with conspicuously high passenger counts. These include January 13, 2016, at 19:00 (7 PM) with 69 passengers and again at 20:00 (8 PM) with 72 passengers. Further instances are January 22, 2016, at 18:00 (6 PM) with 55 passengers, and at 19:00 (7 PM) with 53 passengers. The apex of ridership appeared on January 29, 2016, at 22:00 (10 PM) with a notable 82 passengers. January 30, 2016, at 18:00 (6 PM) also shows a high count of 55 passengers.

These peak periods are likely associated with increased city activities such as the conclusion of the typical workday (5-7 PM), and weekend evenings, and are potentially influenced by special events or holidays. Specifically, the weekday peak on January 13 and the late-night surge on the weekend of January 29-30 support this assumption.

The dataset's dynamic nature underscores the importance of dynamic clustering via methods like AT-CRP. Taxi ridership exhibits inherent temporal variability, influenced by hourly, daily, and weekly patterns. Traditional clustering methods often overlook these changes, whereas AT-CRP can dynamically adapt. The ability to identify evolving clusters makes AT-CRP particularly suitable for capturing real-world taxi demand dynamics and facilitates deeper insight into the city's urban mobility trends. This adaptability is key to optimizing resource allocation, detecting anomalies, and gaining a more accurate understanding of the factors affecting transportation demands.

Figure 6 presents a temporal comparison of cluster numbers resulting from different methodologies applied to the NYC taxi demand dataset. The x-axis represents time, measured in hours, spanning a period of analysis. The y-axis indicates the number of clusters identified by each algorithm. Four methods are compared: AT-CRP (Adaptive Temporal Chinese Restaurant Process), CRP Forgetting, DPM (Dirichlet Process Mixture), and TS-CRP (Time-Sensitive Chinese Restaurant Process).

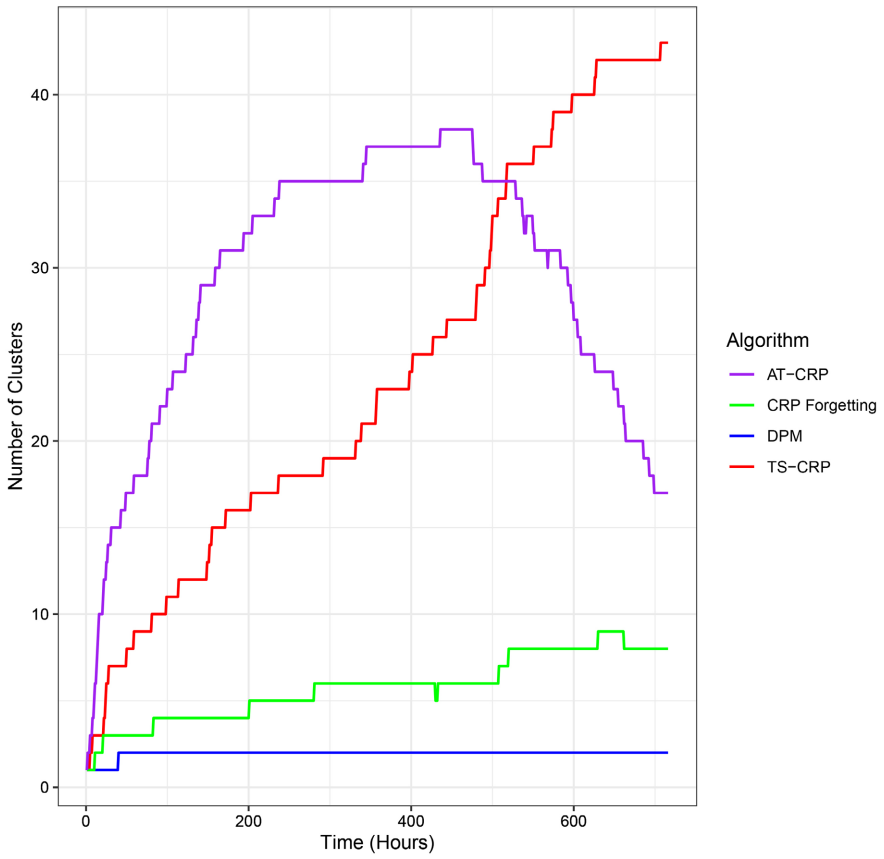


Figure 6. Evolution of cluster counts with different algorithms for NYC taxi trips.

The Adaptive Temporal Chinese Restaurant Process (AT-CRP), depicted by the purple line, exhibits dynamic behavior, with the number of clusters fluctuating between approximately 1 and 38 throughout the duration of the simulation. This

variation indicates AT-CRP's capacity to adapt to the changing patterns of taxi demand, creating and dissolving clusters in response to new data. In contrast, the CRP Forgetting method (green line) also displays a fluctuating number of clusters but generally maintains a lower cluster count compared to AT-CRP, and only one time it decreases over time. The algorithm prunes aggressively less frequent clusters.

The Time-Sensitive Chinese Restaurant Process (TS-CRP), shown by the red line, demonstrates a continuously increasing number of clusters over time. This trend suggests the algorithm is highly sensitive to temporal changes in the data, creating new clusters to capture even subtle variations in taxi demand. However, this sensitivity may lead to an over-complex representation of the underlying patterns, potentially making it challenging to extract meaningful, long-term trends. On the other hand, the DPM (Dirichlet Process Mixture) model, represented by the blue line, exhibits a relatively low number of clusters and low volatility, often holding the number of clusters. This suggests that DPM is not able to adapt well. This underfitting makes it difficult to extract meaningful, generalizable patterns, potentially hindering its effectiveness for forecasting or identifying underlying trends in NYC taxi trips.

The Adaptive Temporal Chinese Restaurant Process (AT-CRP) demonstrates a dynamic behavior, with the number of clusters fluctuating between approximately 1 and 38 throughout the 716-hour simulation period. This variability highlights the AT-CRP's ability to adapt to the changing patterns of taxi trips by dynamically creating and dissolving clusters, showcasing its sensitivity to evolving trends in the data while maintaining a reasonable level of complexity compared to methods that either over- or under-fit the data.

Figure 7 presents a heatmap visualizing the cluster distribution obtained from the AT-CRP method applied to the NYC taxi trips dataset. The x-axis represents the hour of the day, while the y-axis corresponds to the day, providing a temporal representation of cluster activity from January 4th to February 1st. The color intensity reflects the number of clusters active at each hour of each day, offering insights into the dynamic patterns of taxi trips in New York City.

The heatmap reveals distinct patterns in cluster activity throughout the observation period. Brighter colors indicate higher cluster activity, which is prominently observed during peak hours, particularly in the late afternoon and evening. These periods likely correspond to increased demand due to commuters travelling home from work, individuals engaging in social and recreational activities, or people going out for dinner and evening entertainment. Furthermore, the heatmap suggests that certain days exhibit consistently higher cluster activity, potentially reflecting weekday versus weekend demand variations or the influence of specific events, perhaps correlating with weekend tourism trends or special events that draw more people into the city.

Darker areas on the heatmap signify periods of lower cluster activity, meaning there is less variation in taxi demand patterns during those times. These darker

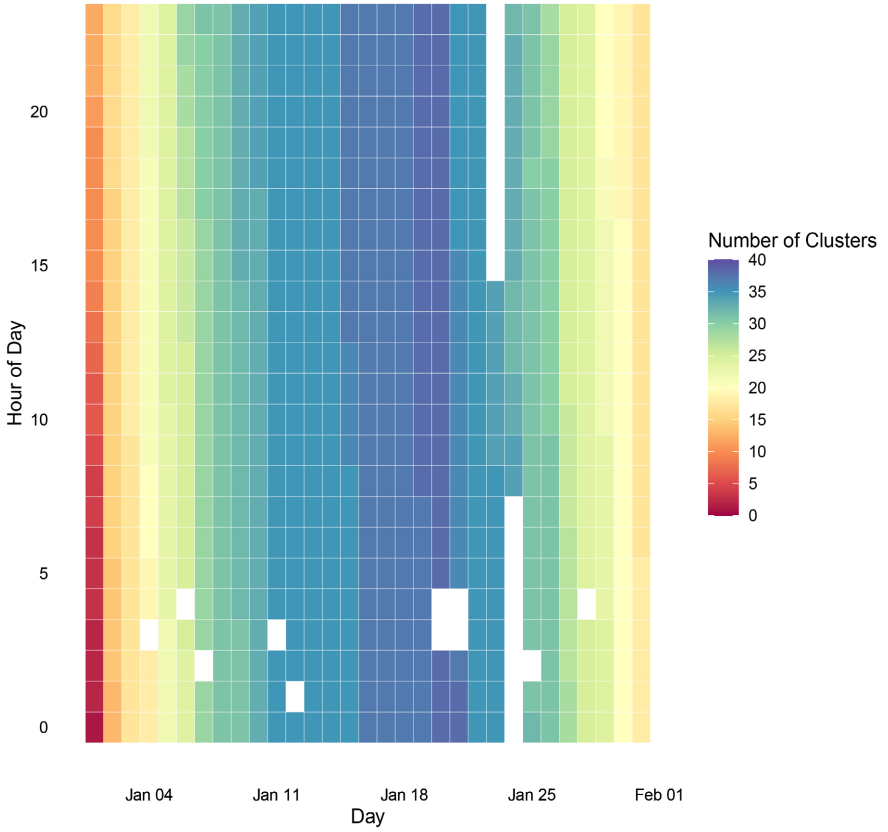


Figure 7. Heatmap of AT-CRP cluster activity for NYC taxi trips.

areas are observed primarily during the early morning hours (e.g., 1 AM to 5 AM) when most of the city is asleep and activity is low. This suggests more uniform, and therefore fewer clusters, because a typical hourly demand can capture the taxi demand. The dynamic clustering facilitated by AT-CRP is helpful in understanding how taxi demand varies throughout the day and across different days of the week.

White squares in the heatmap represent instances where there is no recorded taxi data for that specific hour and day. This could be due to data collection issues, missing records, or periods of genuinely zero taxi activity. The colors in the heatmap reflect the relative intensity of clustering. Brighter colors suggest more dynamic and varied demand patterns requiring a greater number of clusters to accurately represent the data, while darker colors indicate more stable and predictable patterns that can be captured by fewer clusters. The white squares do not contain any information or recorded data. This approach provides urban planners and transportation providers the tools necessary to optimize resource allocation, improve service efficiency, and gain a deeper understanding of the factors influencing taxi demand patterns in the city.

The Adaptive Temporal Chinese Restaurant Process (AT-CRP) model depends on four hyperparameters: the concentration parameter α, the temporal decay rate β, the adaptive growth rate γ, and the pruning threshold ϵ. To deter-

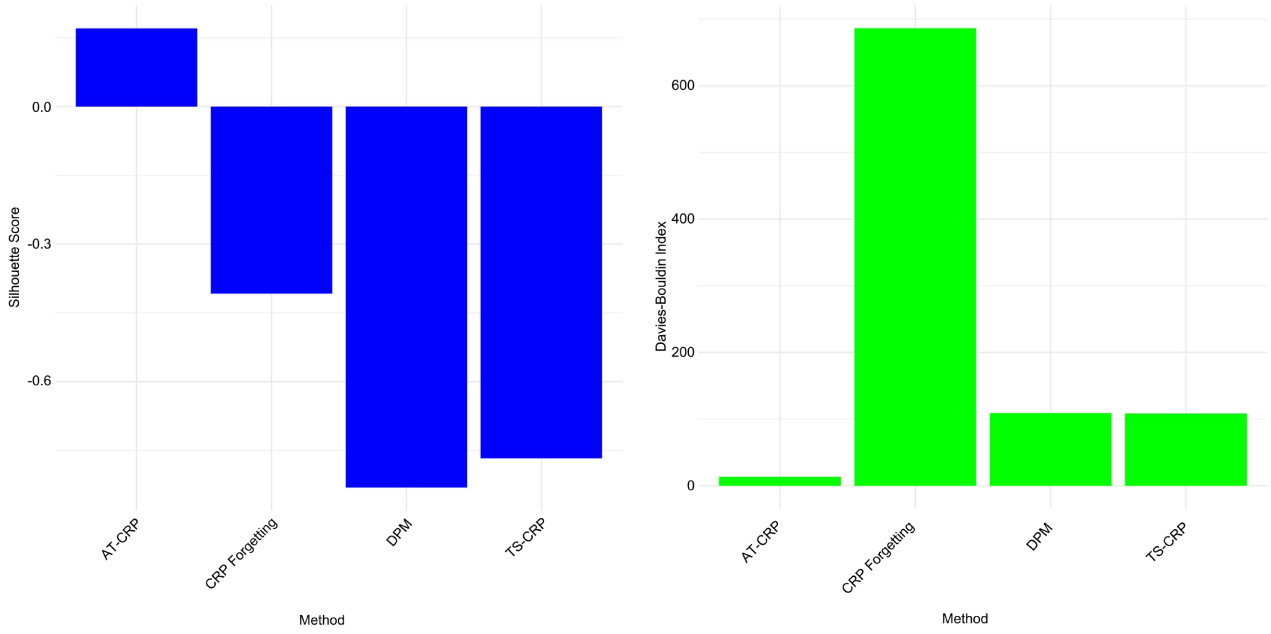
mine these values, I employed Bayesian optimization, which treats the hyperparameter tuning problem as a global optimization task over a black-box objective function. The AT-CRP effectively captures the dynamic clustering of NYC taxi trip data over a one-month period by leveraging its key parameters, estimated as $\alpha \approx 2.41$, $\beta \approx 1.23\times10^{-7}$, $\epsilon \approx 6.14\times10^{-3}$, and $\gamma \approx 6.91\times10^{-8}$. The concentration parameter α, with a 95% credible interval ranging from 2.26 to 2.55, indicates a moderate tendency to form new clusters, reflecting the emergence of new demand zones in the city as taxi trips vary across neighborhoods and times of day. The temporal decay factor β, with a credible interval spanning 1.42×10^{-18} to 1.36×10^{-4}, ensures that historical data has minimal influence on cluster assignments, allowing the model to prioritize recent patterns such as rush-hour peaks or event-driven surges in demand. The noise threshold ϵ, ranging from 1.13×10^{-15} to 10.66, filters out transient anomalies in trip data, such as short-lived spikes in demand, while preserving meaningful long-term trends like recurring daily or weekly patterns. Finally, the cluster mortality rate γ, with a credible interval from 5.44×10^{-20} to 1.87×10^{-4}, enforces strict criteria for removing obsolete clusters, ensuring that outdated demand zones are pruned efficiently without disrupting active clusters.

Together, these parameter values enable AT-CRP to dynamically adapt to the spatial and temporal evolution of taxi demand across NYC, capturing both stable macro-level trends and short-term micro-level variations. For example, the model can identify recurring high-demand areas during morning and evening rush hours while also detecting temporary hotspots near event venues or during inclement weather conditions. This configuration makes AT-CRP particularly well-suited for real-time urban mobility analytics and decision-making in dynamic environments like NYC's taxi ecosystem.

Figure 8(a) presents the Silhouette Scores for four clustering methods applied to the NYC taxi trips dataset: AT-CRP, CRP Forgetting, DPM, and TS-CRP. The Silhouette Score, ranging from −1 to 1, provides a measure of how well each data point fits within its assigned cluster, with higher positive scores indicating better clustering. **Figure 8(b)** displays the Davies-Bouldin Index for the same four methods, where lower values indicate better-defined and well-separated clusters.

Based on **Figure 8(a)** the AT-CRP exhibits the highest Silhouette Score, achieving a positive value of approximately 0.17. This suggests that AT-CRP produces relatively good clustering, with data points closer to their own cluster than to others. The other methods show negative Silhouette scores, with DPM having the lowest score of about −0.83, indicating that these methods struggle to create well-defined clusters for this dataset.

From **Figure 8(b)**, I observe that AT-CRP also exhibits the lowest Davies-Bouldin Index, further supporting its superior clustering performance. DPM shows the second-lowest Davies-Bouldin Index, closely following TS-CRP. CRP Forgetting has significantly higher Davies-Bouldin Index values. The exceptionally high Davies-Bouldin Index for CRP Forgetting suggests that this method



(a) (b)

Figure 8. Clustering performance comparison. (a) Silhouette score; (b) Davies-Bouldin index.

produces clusters that are poorly separated and/or have high intra-cluster distances. This could indicate that CRP Forgetting is creating many small, overlapping clusters or a few very large, spread-out clusters. The high value might also suggest that the method is sensitive to noise or outliers in the NYC taxi trips data. The close Davies-Bouldin Index values for DPM and TS-CRP indicate that these methods produce clusters with similar levels of separation and compactness, despite their different Silhouette Scores. This discrepancy between the two measures highlights the importance of using multiple evaluation metrics when assessing clustering performance.

The difference in inference between the Silhouette Score and Davies-Bouldin Index can be attributed to what each metric measures. The Silhouette Score focuses on the relationship of each data point to its own cluster and neighboring clusters, while the Davies-Bouldin Index emphasizes the overall separation and compactness of clusters. In this case, DPM seems to create clusters that are well-separated (low Davies-Bouldin Index) but may not always place individual points in their optimal clusters (lower Silhouette Score compared to its Davies-Bouldin Index performance).

Dynamic clustering is particularly important for analyzing NYC taxi trip data due to the temporal and spatial variations in demand patterns. The proposed method AT-CRP shows a better performance in this analysis, and can adapt to changing patterns throughout the day, week, or in response to special events. This adaptability allows for a more nuanced understanding of taxi demand, potentially leading to improved resource allocation, pricing strategies, and urban planning decisions. In conclusion, based on these results, the AT-CRP method appears to

be the most suitable for clustering NYC taxi trip data, as it demonstrates the best overall performance across both metrics. However, the choice of method may depend on specific analysis goals and computational constraints. The significant differences in performance across methods underscore the importance of careful algorithm selection and evaluation when working with complex, real-world datasets like NYC taxi trips.

6. Conclusions

In this research, I have presented the Adaptive Temporal Chinese Restaurant Process (AT-CRP), a novel nonparametric framework designed to address the challenges of clustering high-velocity streaming data with evolving temporal dynamics. Unlike traditional clustering methods like K-means [5] and Gaussian Mixture Models [6], which assume static cluster structures, the AT-CRP explicitly models the entire lifecycle of clusters, including their birth, growth, evolution, and eventual dissolution. This capability is crucial for capturing the dynamic nature of real-world data streams, such as those found in social media, IoT sensor networks, and financial markets.

The AT-CRP builds upon existing temporal extensions of the Chinese Restaurant Process, such as the Dependent Dirichlet Process (DDP) [2] and the Time-Sensitive Chinese Restaurant Process (TS-CRP) [3]. However, it overcomes its limitations by integrating adaptive mechanisms that dynamically introduce new clusters and remove obsolete ones without manual intervention. This is achieved through a novel combination of temporal decay and adaptive growth rates, allowing the model to continuously adapt to evolving patterns and trends in the data. The AT-CRP also addresses the limitations of the CRP with Forgetting [15] by providing a more principled approach to cluster pruning that does not rely on ad hoc threshold settings.

I derived an efficient online variational inference algorithm for AT-CRP, enabling real-time updates to cluster assignments while maintaining theoretical guarantees. This is critical for practical applications where computational efficiency is paramount. The online nature of the algorithm allows the model to process new data points incrementally, avoiding the need for batch processing and enabling continuous adaptation to the evolving data stream. Furthermore, the variational inference approach provides a scalable alternative to MCMC methods, making it suitable for large-scale streaming data applications.

The validation of AT-CRP on both synthetic and real-world datasets demonstrated its effectiveness in capturing meaningful temporal patterns across diverse domains. The application to Twitter trending topics showcased the model's ability to identify and track emerging and fading trends in social media, providing valuable insights into public sentiment and opinion dynamics. The analysis of stock market transaction data revealed its potential for uncovering dynamic patterns in financial markets, such as regime shifts and market anomalies, which could be useful for risk management and algorithmic trading. These results highlight the

versatility and applicability of AT-CRP in real-world scenarios where data is continuously generated and evolving.

Future research directions include further exploration of the theoretical properties of AT-CRP, such as deriving tighter regret bounds and investigating its consistency properties under various data distributions. In addition, I plan to develop task-specific evaluation criteria that directly measure temporal accuracy and semantic validity of the dynamic clusters, for example by assessing alignment with known events (e.g., holidays, major incidents), evaluating the forecasting performance of cluster-based models, and incorporating domain-expert assessments of cluster interpretability in urban mobility and other time-series applications. Extending the model to handle more complex data structures, such as hierarchical or multi-modal data streams, would broaden its applicability to a wider range of real-world problems; this could involve incorporating techniques from deep learning or graph theory to capture more complex relationships between data points and clusters. Finally, investigating the use of AT-CRP in other domains, such as IoT sensor networks for anomaly detection or healthcare for patient monitoring, could reveal new insights and applications of the framework. Overall, the AT-CRP offers a promising approach to clustering dynamic data streams and provides a solid foundation for future research in this area.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Lu, J., Liu, A., Dong, F., Guo, J., Gama, J. and Zhang, G. (2019) Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, **31**, 2346-2363.

[2] MacEachern, S.N. (2000) Dependent Dirichlet Processes. Technical Report, Department of Statistics, The Ohio State University.

[3] Zhu, X., Ghahramani, Z. and Lafferty, J. (2005) Time-Sensitive Dirichlet Process Mixture Models. Technical Report, Carnegie Mellon University.

[4] Quintana, F., Mueller, P., Jara, A. and MacEachern, S. (2020) The Dependent Dirichlet Process and Related Models. https://doi.org/10.48550/arXiv.2007.06129

[5] MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the* 5*th Berkeley Symposium on Mathematical Statistics and Probability*, California, 21 June-18 July 1967, 281-297.

[6] Duda, R., Hart, P. and Stork, D.G. (2001) Pattern Classification. 2 Edition, Wiley.

[7] Saeedmanesh, M. and Geroliminis, N. (2017) Dynamic Clustering and Propagation of Congestion in Heterogeneously Congested Urban Traffic Networks. *Transportation Research Part B: Methodological*, **105**, 193-211. https://doi.org/10.1016/j.trb.2017.08.021

[8] Wickramasinghe, L., Leblanc, A. and Muthukumarana, S. (2023) Bayesian Inference on Sparse Multinomial Data Using Smoothed Dirichlet Distribution with an Application to COVID-19 Data. *Model Assisted Statistics and Applications*, **18**, 207-226. https://doi.org/10.3233/mas-221411

[9] Antoniak, C.E. (1974) Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*, **2**, 1152-1174. https://doi.org/10.1214/aos/1176342871

[10] Chang, J., Lu, Y., Xue, P., Xu, Y. and Wei, Z. (2022) Automatic Channel Pruning via Clustering and Swarm Intelligence Optimization for CNN. *Applied Intelligence*, **52**, 17751-17771. https://doi.org/10.1007/s10489-022-03508-1

[11] Saad, F. and Mansinghka, V. (2018) Temporally-Reweighted Chinese Restaurant Process Mixtures for Clustering, Imputing, and Forecasting Multivariate Time Series. *Proceedings of the* 21*st International Conference on Artificial Intelligence and Statistics*, Lanzarote, 9-11 April 2018, 755-764. https://proceedings.mlr.press/v84/saad18a.html

[12] Blei, D.M., Griffiths, T.L. and Jordan, M.I. (2007) The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies. *Journal of the ACM*, **57**, 1-30. https://doi.org/10.1145/1667053.1667056

[13] Blei, D.M., Jordan, M.I., Griffiths, T.L. and Tenenbaum, J.B. (2003) Hierarchical Topic Models and the Nested Chinese Restaurant Process. 2003 *Proceedings of the* 17*th International Conference on Neural Information Processing Systems*, Vancouver and Whistler, 8-13 December 2003, 17-24.

[14] Wang, C. and Blei, D. (2009) Variational Inference for the Nested Chinese Restaurant Process. *Advances in Neural Information Processing Systems* 22, Vancouver, 7-10 December 2009.

[15] Chen, J., Wei, L., Uliji, and Zhang, L. (2018) Dynamic Evolutionary Clustering Approach Based on Time Weight and Latent Attributes for Collaborative Filtering Recommendation. *Chaos*, *Solitons & Fractals*, **114**, 8-18. https://doi.org/10.1016/j.chaos.2018.06.011

[16] Ahmed, A., Hong, L. and Smola, A.J. (2013) The Nested Chinese Restaurant Franchise Process: User Tracking and Document Modeling. 2013 *International Conference on Machine Learning* (*ICML*), Atlan, 16-21 June 2013.

[17] Nutakki, G.C. and Nasraoui, O. (2017) Clustering Data Streams with Adaptive Forgetting. 2017 *IEEE International Congress on Big Data* (*BigData Congress*), Honolulu, 25-30 June 2017, 494-497. https://doi.org/10.1109/bigdatacongress.2017.72

[18] Huang, R., Zhu, F. and Heng, P. (2015) The Dynamic Chinese Restaurant Process via Birth and Death Processes. *Proceedings of the AAAI Conference on Artificial Intelligence*, **29**, 2687-2693. https://doi.org/10.1609/aaai.v29i1.9546

[19] Palaiokrassas, G., Voulodimos, A., Litke, A., Papaoikonomou, A. and Varvarigou, T. (2018) A Distance-Dependent Chinese Restaurant Process Based Method for Event Detection on Social Media. *Inventions*, **3**, Article 80. https://doi.org/10.3390/inventions3040080

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. and Dubourg, V. (2011) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.

[21] Bergstra, J. and Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, **13**, 281-305.

[22] Snoek, J., Larochelle, H. and Adams, R.P. (2012) Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems* 25, Nevada, 3-8 December 2012.

[23] Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. and Talwalkar, A. (2017) Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *The Jour-*

nal of Machine Learning Research, **18**, 6765-6816.

[24] Kohavi, R. (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, **14**, 1137-1145.

[25] Hoffman, M.D., Blei, D.M., Wang, C. and Paisley, J. (2013) Stochastic Variational Inference. *Journal of Machine Learning Research*, **14**, 1303-1347. http://jmlr.org/papers/v14/hoffman13a.html

[26] Ghosh, J., Tan, A. and Luo, L. (2025) Online Bayesian Variable Selection and Bayesian Model Averaging for Streaming Data. *Stat*, **14**, e70044. https://doi.org/10.1002/sta4.70044

[27] Ranganath, R., Wang, C., David, B. and Xing, E. (2013) An Adaptive Learning Rate for Stochastic Variational Inference. *Proceedings of the* 30*th International Conference on Machine Learning*, Atlanta, 16-21 June 2013, 298-306. https://proceedings.mlr.press/v28/ranganath13.html

[28] Broderick, T., Boyd, N., Wibisono, A., Wilson, A.C. and Jordan, M.I. (2013) Streaming Variational Bayes. arXiv: 1307.6769.

[29] Gösgens, M.M., Tikhonov, A. and Prokhorenkova, L. (2021) Systematic Analysis of Cluster Similarity Indices: How to Validate Validation Measures. *Proceedings of the* 38*th International Conference on Machine Learning*, Online, 18-24 July 2021, 3799-3808. https://proceedings.mlr.press/v139/gosgens21a.html

[30] Campello, R.J.G.B., Kröger, P., Sander, J. and Zimek, A. (2020) Density-Based Clustering. *WIREs Data Mining and Knowledge Discovery*, **10**, e1343. https://doi.org/10.1002/widm.1343

[31] Rousseeuw, P.J. (1987) Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

[32] Januzaj, Y., Beqiri, E. and Luma, A. (2023) Determining the Optimal Number of Clusters Using Silhouette Score as a Data Mining Technique. *International Journal of Online and Biomedical Engineering*, **19**, 174-182. https://doi.org/10.3991/ijoe.v19i04.37059

[33] Davies, D.L. and Bouldin, D.W. (1979) A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, 224-227. https://doi.org/10.1109/tpami.1979.4766909

[34] Xie, C., Yu, D., Zheng, X., Wang, Z. and Jiang, Z. (2021) Revealing Spatiotemporal Travel Demand and Community Structure Characteristics with Taxi Trip Data: A Case Study of New York City. *PLOS ONE*, **16**, e0259694. https://doi.org/10.1371/journal.pone.0259694

[35] NYC Taxi and Limousine Commission (2025) NYC Taxi and Limousine Commission Website. https://www.nyc.gov/site/tlc/index.page