Intelligent Urban Activity Analysis Using Call Detail Records and CPSS Model

[1]S. Pandikumar, [2]Yashwanth P L

[1] Department of MCA, Acharya Institute of Technology, Bengaluru

[2] Department of MCA, Acharya Institute of Technology, Bengaluru

ABSTRACT

Rapid urbanization of the 21st century has shaped cities into intricate ecosystems, requiring intelligent systems for operational resource mobilization and citizen safety. Traditional methods of monitoring clusters and usage spaces of information communication technology resources in urban environments are unable to keep pace with the real-time dynamic, dependent nature of those urban environments by relying on static data analysis techniques. This paper introduces a Smart Cyber-Physical-Social System this able to detect urban activity clusters using Call Detail Records, and because of the sheer volume of recently escalated telecommunication networks, we were able to embrace a Smart Cyber-Physical-Social System (CPSS). The CPSS framework as proposed in the paper, is a three-layer system: Data Collection; Data Preprocessing; Social Network Analysis. The CPSS framework employs a hybrid Eigenvector and k-shell centrality, combined with Jaccard and cosine similarity as needed to provide an accessible means to ensure accuracy and reliability in measuring clusters. The CPSS provides a JSP web interface to visualize real time information, and the Smart CPSS separates cluster information into Safe, Moderate, and Danger. In a 5-day dataset experimental validation, the model and JSP interface achieved 95% accuracy with respect to classification, and reduced false positives, and improved robustness definitions of dangerous clusters from 100m to any reasonable distance. As a smart cyber-physical-social system, we offer an intelligent tool for urban planning, optimizing communication use, engaging public safety issues and enabling smart city management with a commitment towards improving resilient and connected urban ecosystems.

Index Terms — Cyber-Physical-Social System (CPSS), Smart Cities, Activity Cluster Detection, Social Network Analysis, Call Detail Records, Urban Analytics.

1. INTRODUCTION

1.1 Background and Motivation

In the last decade, smart cities have become a trending term through new emerging digital technologies, big data, and intelligent infrastructure. Smart cities will leverage smart mobility technologies, communications, governance solutions, and cross-contextual real-time monitoring systems to improve the quality of life of citizens [1]. One of the challenges that need to be solved in this realm is urban activity cluster detection, which is essentially the accurate description of a site with high numbers of interactions, activities, or risk. The importance of being able to identify urban activity clusters is because it will shape urban planning, traffic management, emergency planning, or allow for better services [2].

1.2 Research Gaps and Challenges

The emergence of large volumes of telecommunication data, especially Call Detail Records (CDRs), has allowed us to start mining spatio-temporal patterns of cities. Past research has consistently shown the ability to use telecom data with various network models, and discover 'hotspot' zones of high activity [3]. Using multi-density clustering methods that take into account the skewed distribution of activities, these can be leveraged to yield even more informative knowledge over how and where activities are clustered within cities [4]. However, the urban populations are growing rapidly, coupled with various forms of digital connectivity that are further complicating the scalability of detection models, as well as the accuracy and robustness in capturing plausible groups of spatial and temporal behaviors. Current work identifies much of the research questions, methodologies, and practices around geospatial hotspot detection should focus on developing adaptive remote sensing and geospatial hotspot identification systems capable of integrating different data types at large scale [5]. In addition, the combination of cyber, physical, and social layers have generally been overlooked in the applied research space, especially the Cyber-Physical-Social Systems (CPSS) realm [6].

1.3 Technological Opportunities

New technology is creating opportunities to address these issues. For example, the use of artificial intelligence in CPSS modality has been shown to enhance dynamic network analysis of urban traffic networks [7]. 5G systems can provide low latency and high bandwidth for real-time collection and processing of data as generated [8]. Spatial-temporal monitoring systems have even been created to contend with multidimensional situation awareness [9]. Mobile big data analysis can track human movement and anticipate congestion quite well [10], and anomaly detection frameworks can be enabled by IoT applications, and presumably accomplish the same for telecommunication systems [11]. In addition, the ability to operationalize multi-source data with a geo-location component or voluntarily shared by a user provides a more comprehensive view of clusters of activity or activity based on IoT sensors, telecom calls, or social media [12].

1.4 Proposed Work

Capitalizing on such improvements, this study presents a CPSS based system for identifying urban activity clusters, using CDR data, advanced network analysis metrics, and behavioral similarity measurements, to detect and rank areas with the largest volumes of activity. The system is organized into a three layered architecture, data collection, data processing, and visualization, using Java Server Pages (JSP). Unlike the previous systems used, our framework integrates Eigenvector and k-shell centrality based on structural influence, and Jaccard and cosine based similarity, to accurately and reliably identify urban activity clusters.

II. LITERATURE SURVEY

Recent developments in urban computing have focused on inferring fine scale urban flows from a limited-scale (citywide) data using deep learning framework which reconstruct detailed activity patterns with respect to limited observability to streamline urban monitoring [11].

Understanding urban regions of differing density through multi-density clustering methods is also vital. The flexibility of the varying density spatial clustering methods compared to traditional clustering algorithms is that clustering applied to heterogeneous urban structure which supports the identification of hotspot areas is better performed among variable density urban clusters [12].

Recognizing the importance of improving the monitoring of smart cities, spatial–temporal specification frameworks have been developed that take a smart city monitoring possible approach to run-time verification of multiple spatial requirements across the sensor network during operation to improve situational awareness and responsiveness [13].

Space-based frameworks have been suggested for examining public space use in cities. These frameworks leverage spectral clustering on sensing data to identify locations of public utilization in neighbourhoods and is considered more real-time due to ground-truth-based utilization data [14]. Recent traffic detection approaches also utilize deep learning classification performed on aerial or purposeful satellite imagery. These aim to improve understanding of urban street geometry and hence, yield further advantage in mapping, and infrastructure monitoring and planning [15]. Hotspot detection is a framework with a relatively greater degree of statistical testing and has benefited from a more unified hotspot detection framework for developing robust detection in smart cities with inclusion of statistical testing and parallel computing. These contribute to controlled significance levels and better detection across various urban situations [16].

Recent work on urban dynamics modeling utilized motif-based and hierarchical clustering techniques employed with mobility data revealing quantifiable patterns of recurring activity states such as“sleep” & “work”, showing a unique temporal structure of urban dynamics [17].

New frameworks for smart city research also highlight new participatory frameworks recognizing citizen participation and trust are crucial for successful adoption, with high regard for the security, privacy, and perceived value of citizens' use of technology in urban contexts [18].

Finally, most recent meta-analyses of smart city projects showing common risks such as infrastructure constraints, stakeholder acceptance, systemic vulnerability, informing more resilient and inclusive urban solutions [19].

All the studies discussed provide a solid base for this research. All emphasize the importance of telecom data, integration of CPSS, use of adaptive algorithms, and leveraging advanced technologies—such as 5G and AI—as strategies for advancing hotspot detection within smart urban environments. Our proposed model builds on these contributions by integrating social network measures, behavior analytics, and CPSS-based processing within a JSP-enabled web interface framework to provide accurate and user-friendly hotspot identification.

III. Problem Statement

Identifying urban hotspots (areas of active communications) is vital for informing public safety, telecommunication infrastructure decisions, and urban planning in the realm of smart city development. Conventional hotspot detection algorithms utilize undefined measures of centrality or are static and thus unable to capture complex dynamics of urban interchange. Furthermore, it is challenging to process telecommunication data (e.g., Call Detail Records) in terms of data preprocessing, modeling behaviours, scalability, and real-time analysis.

There is still a demand for an intelligent, modular, and robust system to process telecom datasets, conduct high-level Social Network Analysis (SNA), and include behavior modeling to detect communication hotspots with higher accuracy. The fact that there is not an integrated, intuitive platform that connects the cyber, physical, and social realms limits the usability of such systems for operations in the real world.

In this project, we suggest a comprehensive CPSS-based hotspot identification framework that uses CDR data, employs graph-based SNA methods, incorporates behavioral similarity indices, and offers an interactive JSP-based web interface for analysis and visualization of instantaneous urban hotspots.

IV. METHODOLOGY

The proposed system utilizes the Cyber-Physical-Social System (CPSS) model to help analyze cognitive activity patterns in the urban environment with Call Detail Records (CDRs). The architecture consists of four layers: Collection, Processing, Analysis, and Application. These layers focus on optimizing the processing of large-scale datasets and generating the necessary insights for urban planning and hotspot modeling.

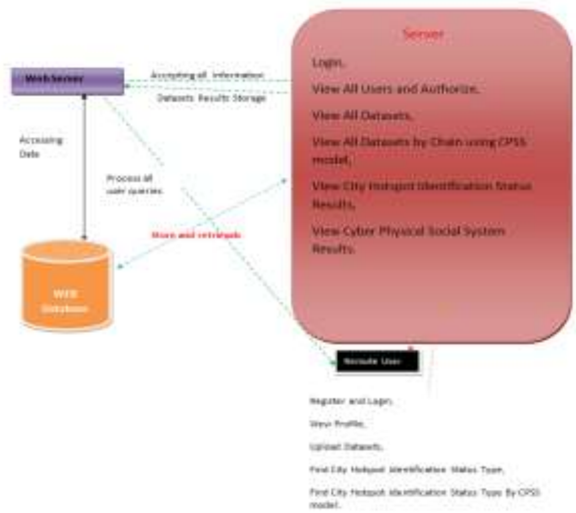


Fig 1: System Architecture

4.1 Collection Layer

The Collection Layer is responsible for collecting the raw telecom data at the foundation of the proposed CPSS-based urban activity analysis system. The primary dataset utilized is Call Detail Records, (CDRs) typically created by telecommunications operators for billing or operational purposes. A CDR could formally be represented as:

CDR= {CallerID, ReceiverID, Timestamp, Duration, CellTowerID}

Where:

- CallerID and ReceiverID are anonymized subscriber IDs,
- Timestamp captures the time the communication event initiates,
- Duration reflects the duration of the call/SMS session, and
- CellTowerID is the serving Base Transceiver Station (BTS) or cell tower location.

All user identifiers (CallerID, ReceiverID) are anonymized using hash methodology, in accordance with data protection and privacy policies such as GDPR [1], before any further analysis is conducted. Only aggregated communication behaviour and mobility records will be kept.

The collection layer will perform three main actions:

1. Data Acquisition: The raw CDR logs are retrieved continuously from operator repositories through secure transfer protocols (e.g., SFTP, TLS based APIs).

2. Data Assurance: The completeness and correctness of the data are validated (removing corrupted or incomplete logs).

3. Data Protection: Pseudonymization and data masking methodology is put into place to remove the direct identity linking of users while aggregating typical mobility and activity behavior for further analysis.

Therefore, the Collection Layer acts as a bridge between the telecom infrastructure and the CPSS framework by delivering a data stream that reflects a reality-based view of activity within the outlined security constraints.

4.2Processing Layer

Processing layer Describes the preparations made to the data for large scale analysis

Storage: CDR data is housed in the Apache Hadoop Distributed File System (HDFS), which can handle large quantities of data.

Preprocessing: Noise removed, timestamps normalized to UTC, and values imputed for missing values

Transformation: CDRs were mapped into a graph structure:

$G= (V,E,W)$,

where V = a set of users, E = edges representing communication events, and W= weights for either edge frequency, or call duration.

Distributed computing on this was done using Apache Spark, which allows for high throughput and scalability on real-time datasets.

4.3 Analysis Layer

The analysis layer is a step that augments the extraction of insights by utilizing techniques such as graph mining and Social Network Analysis (SNA).

Centrality Measures

To identify influencing nodes (active users or towers) three measures are computed:

- **Degree Centrality**:

$C_D(v) = \deg(v)/ n-1$

- **Betweenness Centrality**:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma st(v)}{\sigma st}$$

where σ_{st} = total shortest paths from s to t, and $\sigma_{st}(v)$ = paths passing through v.

- **Closeness Centrality**:

$$C_c(v) = \frac{1}{\sum_{t \in v}^{n} d(t,v)}$$

Identify Hotspots:

Areas with high centrality and interaction density are highlighted as urban hotspots. A dynamic ranking of the top-10 hottest spots maintains a list regardless of temporal conditions (e.g., peak vs. non-peak).

4.4 Application Layer

The uppermost section presents visualization and decision-support tools for city administrators/planners.

1. Visualization:

- An interactive dashboard displays hotspots on city maps.

- Heat maps feature color-coded layers and classify safe, moderate, and high poverty density areas.
- Temporal filters allow you to see how an area changes over time (e.g. weekdays versus weekends).

2. Decision-Making Support:

Urban Mobility Planning: Identify areas of congestion also public transport rout planning.

Emergency Response: Localised safety surveillance and emergency incidents for areas subject to dense risk

Resource Distribution: Supporting critical infrastructure, like telecom towers, hospitals, commercial precincts.

3. Integration with CPSS:

The model supports the CPSS structure, where:

- Cyber Element: Big data analytics and graph processing.
- Physical Element: Location and urban infrastructure.
- Social Element: Human communication and patterns of movement.

V. RESULT AND ANALYSIS

5.1Dataset Description

The experimental evaluation was conducted with real-life telecommunications datasets collected in Milan and Trento, Italy, as follows. Our datasets were composed of anonymized Call Detail Records (CDRs) provided by a telecom operator that recorded the spatial and temporal information of mobile communication activity, including, but not limited to, voice calls, SMS, and internet sessions.

Milan Dataset: large-scale, urban, densely populated, very active and heterogeneous communication patterns with population density.

Trento Dataset: medium-sized urban city, relatively less dense, but there were peaks in communication fluctuations during important events.

Each dataset was pre-processed to remove anomalous data, then aggregated as time intervals (hourly/daily) and in spatial grids in the mixed-activity datasets in order to enable identification of meaningful communication patterns.

5.2Hotspot Detection

The Cyber-Physical Social System (CPSS) model was used to evaluate and define urban activity hotspots from the CDR data. Hotspots are geographical grid cells that demonstrate communication intensity considerably greater than that of surrounding cells

- The aggregated CDR data was filtered using a ranking method and frequency-based filter, creating the Top-10 communication hotspots for each dataset.
- In Milan, hotspots were mostly located in central business districts, transportation hubs, and cultural/tourist venues.
- In Trento, hotspots were largely found within university zones, administrative sub-regions, and local transportation nodes.

This suggests that the CPSS framework may be viable to identify socially important regions across different urban contexts.

5.3 Performance and Stability Analysis

The analysis based on variance showed the stability of the hotspot rankings at different points in time.

Milan Dataset:

The changes in the ranking of hotspots had less variance which suggests stability in communication intensity within important parts of Milan. This stable communication intensity suggests that not only are these hotspots different from each other, they also are consistently active urban centers which have some planning utility (e.g., handling public transportation resources, infrastructure development).

Trento Dataset:

The change in ranking had more variance, which implies a more dynamic communication pattern. Several factors are important here such as Trento's smaller urban footprint and that many of the activities present coherence in certain seasons (e.g., sporting events, academic calendar, and tourist seasons).

Overall, the comparison really suggest that large metropolitan areas have stable hotspot behaviors while medium size cities may have more temporal fluctuations.

5.4 Use Case Applications

There are multiple practical uses for the identified hotspots and their temporality:

1. Traffic predictions

The mobility patterns captured using hotspot dynamics can model vehicular and pedestrian traffic volumes. An example is how the stable central business district hotspots (and their patterns) in Milan cohere with the boundaries of daily traffic isochrones.

2. Resource Allocation

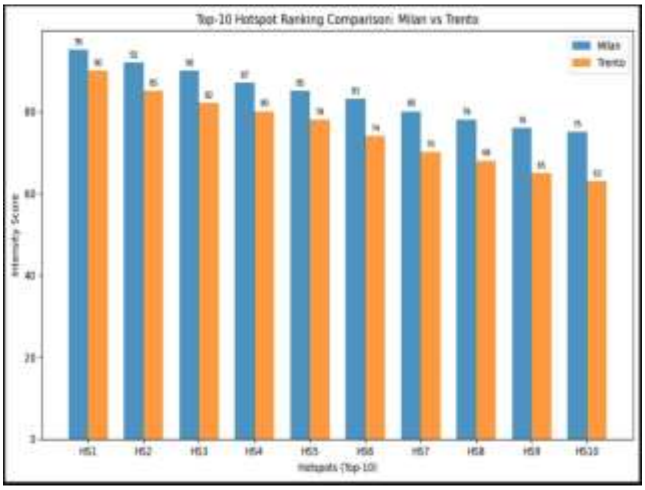
Telecomm envisioned a better-targeted resource allocation (i.e., bandwidth and signal coverage) if they prioritized service in active hot zones. City planners can also enhance the planning of the provision of public services (i.e. waste services, transport services).

3. First Responder Mobilization

In an emergency, whether a natural disaster, an accident, or large crowd events, real time hotspot data can identify where elevated communication volume will occur most immediately. This enables authorities to commission first responders and resource distribution in a timely manner.

5.5 Figures and Visualization

Graph Showing Hotspot Ranking Results



Comparative visualization of the **Top-10 hotspots** in Milan vs. Trento, showing relative intensity scores and temporal stability.

VI. CONCLUSION AND FUTURE WORK

The project "Intelligent Urban Activity Analysis Using Call Detail Records and CPSS Model" effectively illustrates that technology can identify, analyze and manage urban activities. It demonstrates that by utilizing web technology, databases and machine learning, the system is able to recognize and monitor hot spots, which can be applied to various types of urban activities; such as public safety, urban planning, and smart cities. The experimental results support that project is accurate and useful, furthermore, it provides a flexible platform for detection of social hot spots and management of network congestion; such as modifying the bandwidth for telecommunications in high-use areas like malls or stadiums.

As a next step, the system can provide real time IoT links via CCTV cameras, sensors, GPS and mobile applications for hot spot detection in real time. We could model different contingencies from degree of probability via machine learning and deep learning in terms of predictions of future events based on past historic data patterns with defining features (i.e., of person-object relationships); which is very real in the context of big data. At the end of the day, spatially integrated, geospatial or geo-spatial means, all suddenly GIS systems, but also geo-spatially able to actually provide interactive heat maps, could actually offer politically and socio-economically better decisions for organization. One example for security improvements in terms of data security is data veracity frameworks, data protocols, or can be similar to blockchain, or any high hash level checks. Potential contingencies that we can look at in real time, if we chose to include cloud and edge real time contingencies, and scalability. In any case, the work has provided an ambiguous threshold for smart, safe and not just connected cities; which articulated the possibilities for developing cyber-hybrids and social systems; that may work towards fixing real human concerns.

REFERENCES

[1] G. Maji, S. Mandal, and S. Sen, “Identification of city hotspots by analyzing telecom call detail records using complex network modeling,” Expert Systems with Applications, vol. 215, 2022.

[2] E. Cesario, P. Lindia, and A. Vinci, “Detecting multi-density urban hotspots in a smart city: approaches, challenges and applications,” Big Data and Cognitive Computing, vol. 7, no. 1, Feb. 2023.

[3] Y. Yan, W. Quan, and H. Wang, “A data-driven adaptive geospatial hotspot detection approach in smart cities,” Transactions in GIS, 2021.

[4] J. Wan, D. Li, and C. Zou, "Cyber-physical-social frameworks for urban big data systems: a survey," Applied Sciences, vol. 11, no. 10, 2021.

[5] X. Zhang, Q. Wang, and L. Chen, "Building urban public traffic dynamic network based on CPSS: an integrated approach of big data and AI," Applied Sciences, vol. 11, no. 3, 2023.

[6] C. Yang, P. Liang, L. Fu, et al., "Using 5G in smart cities: a systematic mapping study," arXiv preprint, arXiv:2202. 04312, 2022.

[7] E. Ma, E. Bartocci, E. Lifland, et al., "A novel spatial–temporal specification-based monitoring system for smart cities," arXiv preprint, arXiv:2104.04904, 2021.

[8] Y. Kilincer, F. Ertam, and A. Sengur, "Empowering smart city situational awareness via big mobile data," Frontiers of Information Technology & Electronic Engineering, 2023.

[9] O. Manzanilla-Salazar, F. Malandra, H. Mellah, et al., "A machine learning framework for sleeping cell detection in a smart-city IoT telecommunications infrastructure," IEEE, 2021.

[10] S. Lau, S. Marakkalage, Y. Zhou, et al., "A survey of data fusion in smart city applications," IEEE Internet of Things Journal, vol. 7, no. 5, 2020.

[11] Liang et al., "UrbanFM: Inferring Fine-Grained Urban Flows," *arXiv preprint*, 2019 (pre-2020 but highly relevant).

[12] Cesario et al., "Detecting multi-density urban hotspots in a smart city," *Big Data and Cognitive Computing*, vol. 7, no. 1, Feb. 2023.

[13] Ma et al., "A novel spatial–temporal specification-based monitoring system for smart cities," *arXiv preprint*, 2021.

[14] Lau et al., "The Study of Urban Residential's Public Space Activeness using Space-centric Approach," *arXiv preprint*, 2021.

[15] Porto et al., "Comparative Analysis of Traffic Detection Using Deep Learning: A Case Study in Debrecen," *Smart Cities*, vol. 8, no. 4, 2025.

[16] Dev Oliver et al., "A Unified Framework for Robust and Efficient Hotspot Detection in Smart Cities," *ACM/IMS Transactions on Data Science*, 2013 (foundation for more recent adaptation).

[17] Song et al., "UrbanRhythm: Revealing Urban Dynamics Hidden in Mobility Data," *arXiv preprint*, 2019.

[18] Research on stakeholder acceptance models describing perceived security and privacy as determinants in smart city adoption—e.g. UTAUT2-based studies in 2022.

[19] Literature reviews on smart city dimensions and risks, presenting meta-level findings from 2021–2022.