A CPSS-Driven Framework for Dynamic Urban Hotspot Detection Leveraging Socio-Spatial and Social Intelligence

[1]**Dr.G.Apparao Naidu**
Professor of CSE, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: naiduag@gmail.com

[2]**Errakolla Shailaja**
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email:shailajayadav354@gmail.com

[3]**Datla Alekhya**
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email:alekhyadatla77@gmail.com

[4]**Arra Smayana**
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: arrasmayanareddy@gmail.com

***Abstract*— With integrated services like automated housing, intelligent transportation, sophisticated communication networks, and effective infrastructure, smart cities are being created more and more to improve urban living. Finding locations with unusually high communication activity known as telecom hotspots is a major challenge in this field. For telecom companies looking to maximize network performance and resource allocation, identifying these areas is essential. Despite the potential of Cyber-Physical-Social Systems (CPSS) for this task, problems with large-scale data management, computational efficiency, and result reliability still exist. This work presents a sophisticated CPSS framework for hotspot detection using telecom datasets. Three functional layers make up the architecture: a data processing layer that handles data cleaning, storage, and analytical modeling; a hotspot analysis layer that conducts graph-based social network analysis; and a data acquisition layer that gathers raw Call Detail Records (CDRs). The model combines behavior-based metrics like cosine similarity and Jaccard similarity with network-based metrics like Eigenvector centrality and k-shell decomposition to improve detection accuracy. A five-day telecom dataset is used to assess the system and look at how communication patterns change over time. The robustness and accuracy of the system are confirmed by validating the identified hotspots using statistical techniques like autocorrelation and cross-correlation. Experimental findings demonstrate that the proposed approach correctly pinpoints regions of high activity, offering a scalable means of examining communication in smart city environments.**

Keywords: social network analysis, telecom data, hotspot detection, smart cities, and cyber-physical-social systems (CPSS).

I. INTRODUCTION:

The creation of smart cities has been greatly impacted by the quick development of information and communication technologies (ICT), which permit better urban living through intelligent systems, infrastructure, and services. This approach enables accurate hotspot detection and offers a privacy-aware, scalable solution for smart city management. A smart city integrates cyber, physical, and social domains to deliver the sustainable, efficient, and citizen-centric environments. The identification and analysis of urban hotspots areas with high levels of human activity, communication, or economic engagement is one of the many operational challenges in smart cities. These areas are crucial for infrastructure development, telecom service planning, and resource optimization.Traditional hotspot detection relies on telecom data like Call Detail Records (CDRs), which are often restricted due to privacy and availability issues. To overcome this, we suggest a Smart Cyber-Physical-Social System (CPSS) model that infers hotspots without the need for direct communication records by using socioeconomic data. Through the use of Principal Component Analysis (PCA) and Social Network Analysis (SNA) techniques like Eigenvector Centrality, K-shell, Jaccard, and cosine similarity, the model builds a similarity-based graph.This approach enables accurate hotspot detection and offers a privacy-aware, scalable solution for smart city management.

II. LITERATURE REVIEW:

Urban hotspot identification is a critical research area in smart city planning and telecommunication optimization. Early studies, such as Onnela et al. [1], analysed large-scale Call Detail Records (CDRs) by modelling communication networks as weighted graphs, employing metrics like clustering coefficients and node degrees to uncover human communication patterns. Similarly, Nanavati et al. [2] characterized telecom graphs using degree and neighbourhood distributions but faced challenges related to scalability and robustness when applied to dynamic, large-scale networks. To improve data quality, Nattapon et al. [3] improved the accuracy of network analysis by introducing filtering techniques to eliminate anomalies from telecom datasets. Ahmad and associates. [4] combined machine learning and social network analysis (SNA), demonstrating that graph features could accurately forecast customer attrition. Mededovic and others.[5] focused explicitly on hotspot detection, applying centrality measures such as degree and closeness centrality to identify areas of high communication activity. However, these traditional centrality metrics often failed to capture complex relational structures in dense urban environments. Further advancements include Modarresi et al. [6], who proposed resilient graph topologies to improve fault tolerance in smart home networks, and Seufert et al. [7], who applied spatial distribution methods to model public Wi-Fi hotspots. Peiyan et al. [8] developed an entropy-based "Hoten" metric that incorporates human mobility patterns, while Brdar et al. [9] suggested combining multiple centrality measures to enhance hotspot characterization. Despite these efforts, there remains a lack of comprehensive approaches that integrate social, physical, and cyber layers for urban hotspot detection. Behavioural similarity measures like Jaccard and Cosine similarity, which capture relational dynamics beyond mere connectivity, have been underutilized. Amin et al. In order to increase the accuracy and resilience of hotspot detection, [10] presented a Cyber-Physical-Social System (CPSS) framework that combines graph-based SNA with socio-behavioural metrics. However, current CPSS models frequently lack robust validation techniques like autocorrelation and cross-correlation analyses as well as layered architectures to manage large-scale data processing effectively. This disparity drives our suggested multi-layered Smart CPSS architecture, which makes use of similarity metrics like Jaccard and Cosine similarity, sophisticated social network metrics like Eigenvector Centrality and K-shell decomposition, and Principal Component Analysis (PCA) for dimensionality reduction. Our approach not only enhances computational efficiency but also captures more nuanced social and behavioural relationships, leading to robust and scalable hotspot identification applicable to smart city planning and telecom optimization. Emphasizes on automation, cost-effectiveness, and real-time decision-making.

III. METHODOLOGY:

System Architecture

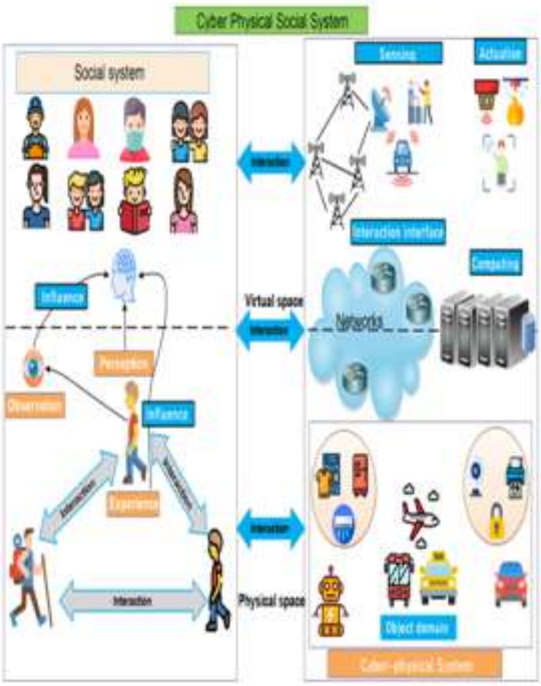

Fig: System Architecture

Data Collection layer: The Census Data Collection module is part of the Data Collection Layer, where the process starts. This module is in charge of collecting unprocessed census data, most likely from a variety of sources like public records, government databases, and surveys. The collected data serves as the foundation for the entire pipeline, ensuring that the subsequent layers have the input they require to accomplish their tasks. This layer only concentrates on gathering data, making sure it is complete and pertinent for subsequent processing. Several tools and methods may be used at this stage to collect data from various sources, including administrative records, national databases, and surveys. Numerous characteristics, including age distribution, income levels, educational attainment, housing conditions, etc., may be covered by the data gathered, depending on the analysis's objectives. After being gathered, the data is sent to the following layer for additional analytical preparation, transformation, and cleaning. The foundational stage of the analytical pipeline is the Data Collection Layer. This layer is in charge of acquiring raw data, which in this case is obtained through the collection of census data.

Data Management and processing Layer: The Data Management and Processing Layer is crucial for transforming unprocessed data into an analysis-ready format. The first step is the Exploratory Data Analysis (EDA) Module, which helps with understanding the distribution and organization of the data as well as identifying trends and anomalies. This is essential for assessing the data's usability and quality. The Principal Component Analysis (PCA) Module, a dimensionality reduction method, is used after EDA. PCA allows for more straightforward but effective analysis by minimizing the number of variables and preserving as much variance in the dataset as possible. The Graph Construction Module creates graph-based data representations following PCA. Finding connections between data points, like resemblances between communities or regions, can be aided by this. The Metric Computation Module then computes different statistical or network metrics, like density or centrality, which offer numerical insights into the data's structure. In general, this layer uses structured transformations to distill the data into meaningful formats and representations, readying it for interpretation.

Data Interpretation layer: The pipeline's last step, the Data Interpretation Layer, is devoted to extracting meaning from the data that has been processed. The first step is Hotspot Identification, which identifies important trends or areas of interest, like densely populated areas, areas with economic inequality, or other anomalies. Decision-makers can more easily focus quickly on areas that require additional research or policy intervention as a result. Hotspot identification frequently entails statistical or spatial analysis using the calculated metrics from the preceding layer. The Visualization and Reporting module converts complex data findings into intelligible and useful insights after hotspot identification. Creating maps, graphs, and summary reports that effectively convey findings to stakeholders is part of this. Making sure that data-driven decisions are founded on accurate interpretations requires effective visualization. Thus, by bridging the gap between technical analysis and real-world application, this layer makes it possible to make well-informed decisions about resource allocation, public health, urban planning, and other areas. The Hotspot Identification, Visualization, and Reporting modules are part of the Data Interpretation Layer, which receives the processed data. The Hotspot Identification module examines the calculated metrics to find noteworthy trends or regions of interest, such as densely populated areas or odd demographic trends, which are commonly known as "hotspots." The Visualization and Reporting module then receives these findings, creates visual representations such as dashboards, maps, or charts, and creates reports to effectively convey the insights. This final step completes the pipeline from raw data to meaningful outcomes by ensuring that the processed data is transformed into a format that stakeholders can easily understand and act upon. Applications in epidemiology, socioeconomic research, urban planning, and geospatial analytics can all benefit from this architecture. If you need assistance modifying or putting into practice a comparable architecture, please let me know.

IV. RESULTS AND ANALYSIS:

The initial step is to identify the regions with the highest communication activity, which can be done by making heatmaps. The CellID appears on the Y-axis, while the day is shown on the X-axis.

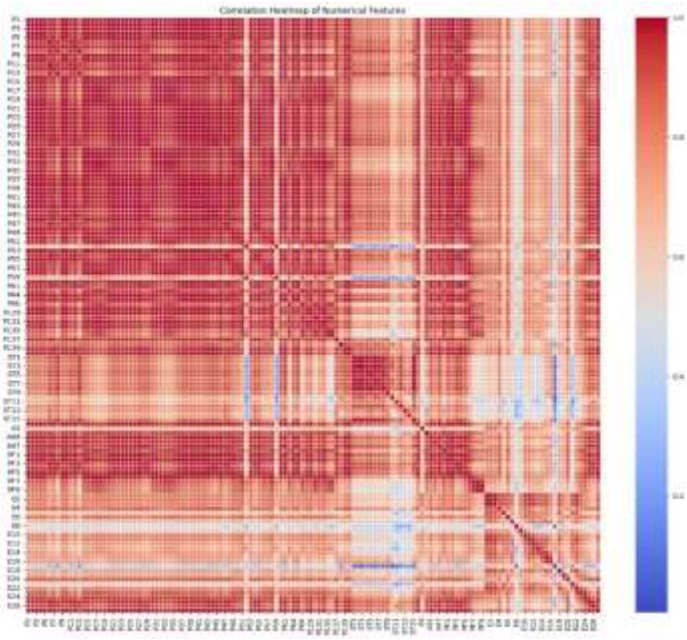

Fig 4.1 Heatmap

Heat maps provide a spatial visualization of hotspot intensity, making it easier to interpret high-activity areas in urban areas. Color-coding regions are used to highlight influential zones based on the sum of the centrality and similarity metrics scores. After all regions have been ranked, the heat map prominently displays the top 10 regions with the highest scores for convenient identification. The heat map can assist stakeholders, such as telecom providers and urban planners, in prioritizing areas for infrastructure development or service optimization. It enables comparative analysis across different time periods or datasets to identify pattern shifts and validate hotspot stability. Additionally, heat maps enhance stakeholder communication by translating complex analytical results into visually appealing images that non-technical audiences can comprehend. This encourages better departmental cooperation and speedier decision-making.

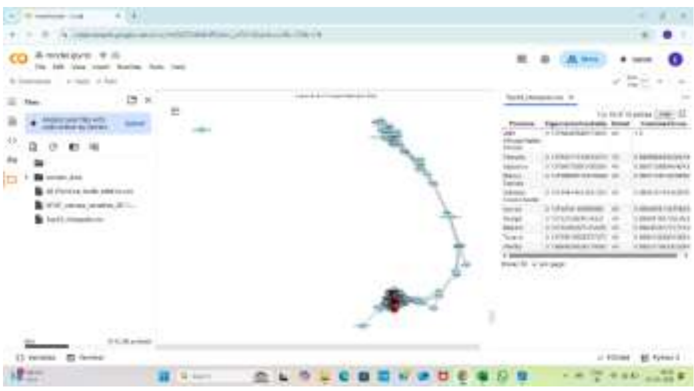

Fig 4.2 Graph with top-10 hotsspot highlights

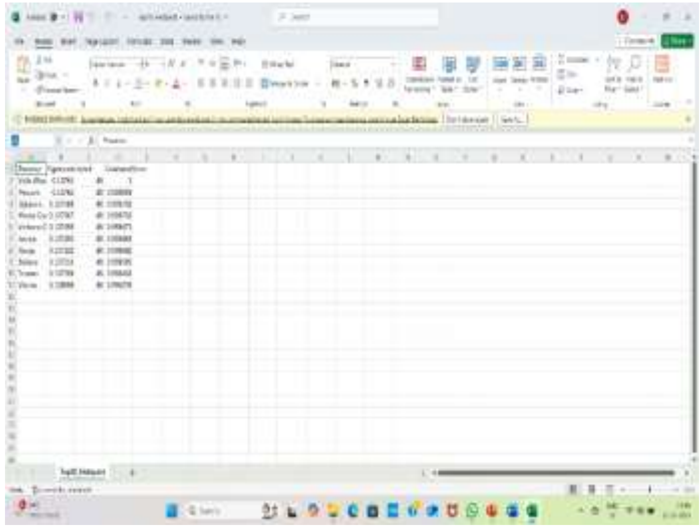

Fig 4.3 Top-10 hotspot identification using CPSS

Shows the top 10 urban hotspots determined by combining centrality and similarity scores in the CPSS model.

V. CONCLUSION:

In conclusion, the proposed work provides a new, trustworthy, and practical framework for locating city hotspots within the broader context of intelligent cyber-physical social systems. Graph-based analytics and socioeconomic modeling are used in the proposed CPSS system to bridge the cyber, physical, and social domains, laying the foundation for more creative, efficient, and sustainable urban management strategies. Future advancements will focus on adding real-time data streams, putting predictive modeling into practice, and expanding the system's capabilities to include cross-city comparative analytics and automated smart interventions. By facilitating data-driven decision-making, it establishes the groundwork for more intelligent, effective, and sustainable urban management techniques. The framework's future goals include integrating real-time data streams, creating predictive models, and implementing automated intelligent interventions.

VI. FUTURE SCOPE:

Future system enhancements will use state-of-the-art graph learning techniques and real-time data streams to improve the accuracy and responsiveness of urban hotspot detection. Adding dynamic sources such as real-time social media feeds, IoT sensor networks, mobile device location data, and public transportation usage statistics can significantly improve the system, which currently depends on static socioeconomic datasets. The model could adjust to quickly shifting urban conditions with these inputs. Furthermore, by replacing traditional Social Network Analysis metrics with state-of-the-art techniques like Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs), the system's ability to identify complex patterns and predict future urban activity hotspots based on shifting social and environmental trends can be enhanced.

VII.REFERENCES:

[1] S. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–

7336, 2007.
[2] M. Nanavati, D. Dixit, A. Gunawardhana, and A. Chakrabarti, “Analyzing the structure and evolution of massive telecom graphs,” *Bell Labs Technical Journal*, vol. 14, no. 3, pp. 209–227, 2009.
[3] Nattapon, K. Chamnongthai, and P. Chongstitvatana, "Preprocessing Methods for Call Detail Record Data in Telecommunication Networks," International Journal of Computer Applications, vol. 93, no. 14, pp. 20–25, 2014.
[4] "Social Network Analysis for Identifying Telecom Customer Churn," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 6, no. 1, pp. 149–156, 2015; A. Ahmad, J. J. Qureshi, and M. M. Rathore.
[5] D. Mededovic, A. Trifunovic, and M. Pavkovic, “Hotspot Detection in Mobile Networks Using Social Network Analysis,” *International Journal of Computer and Information Technology*, vol.5,no.6,pp.605–609,2016.
[6] S. Modarresi, D. H. Shih, and S. C. Hsu, "Graph-Theoretical Modeling of Home Networks for Smart Home Resilience," IEEE Systems Journal, vol. 13, no. 1, pp. 300–309, 2019.
[7] M. Seufert, P. Casas, and A. Blanc, “Passive crowdsourcing for mobile network hotspot monitoring,” Proceedings of IEEE INFOCOM Workshops, pp. 954–959, 2015.
[8] P. Peiyan, W. Jian, W. Xiaomin, and L. Qinglin, “Hoten: An Entropy-Based Metric for Hotspot Detection in Mobile Networks,” *Mobile Networks and Applications*, vol. 21, no. 1, pp. 133–145,2016.
[9] S. Brdar, S. Scepanovic, F. Soldo, and M. Marusic, “How Good is Good Enough? Re-evaluating the Use of Degree Distributions for Telecom Graph Analysis,” *Proceedings of the 23rd International World Wide Web Conference*, pp. 733–738, 2014.

[10] "City Hotspot Identification Using Smart Cyber-Physical Social Systems (CPSS)," International Conference on Smart Cities and Systems, 2022, R. Amin, A. Abbas, and M. M. Rathore.