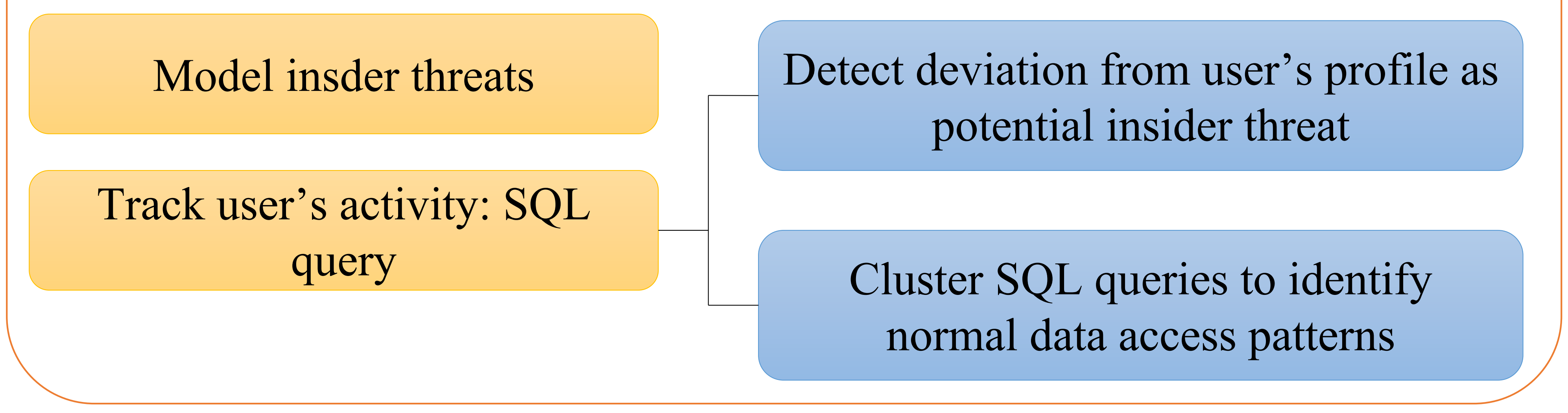
Data is Social: Exploiting Data Relationships to Detect Insider Attacks

Investigators: Oliver Kennedy, Varun Chandola and Shambhu Upadhyaya
PhD students: Gokhan Kul, Duc Thanh Anh Luong and Ting Xie
http://odin.cse.buffalo.edu/research/insider-threats/index.html

Objective

Identify insider threats posed by malicious insiders within large organization

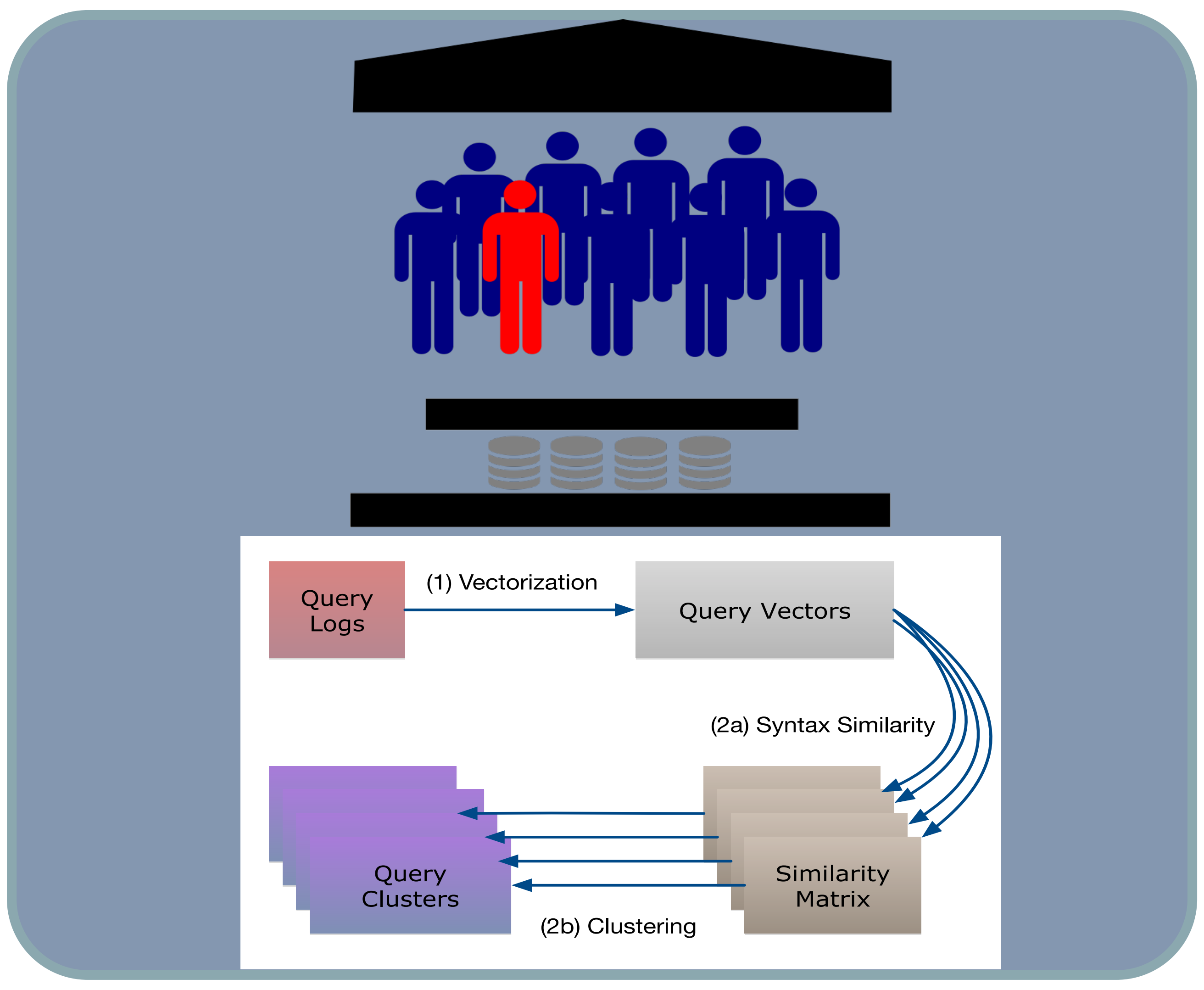
Approach


Model insder threats

Track user's activity: SQL query

Detect deviation from user's profile as potential insider threat

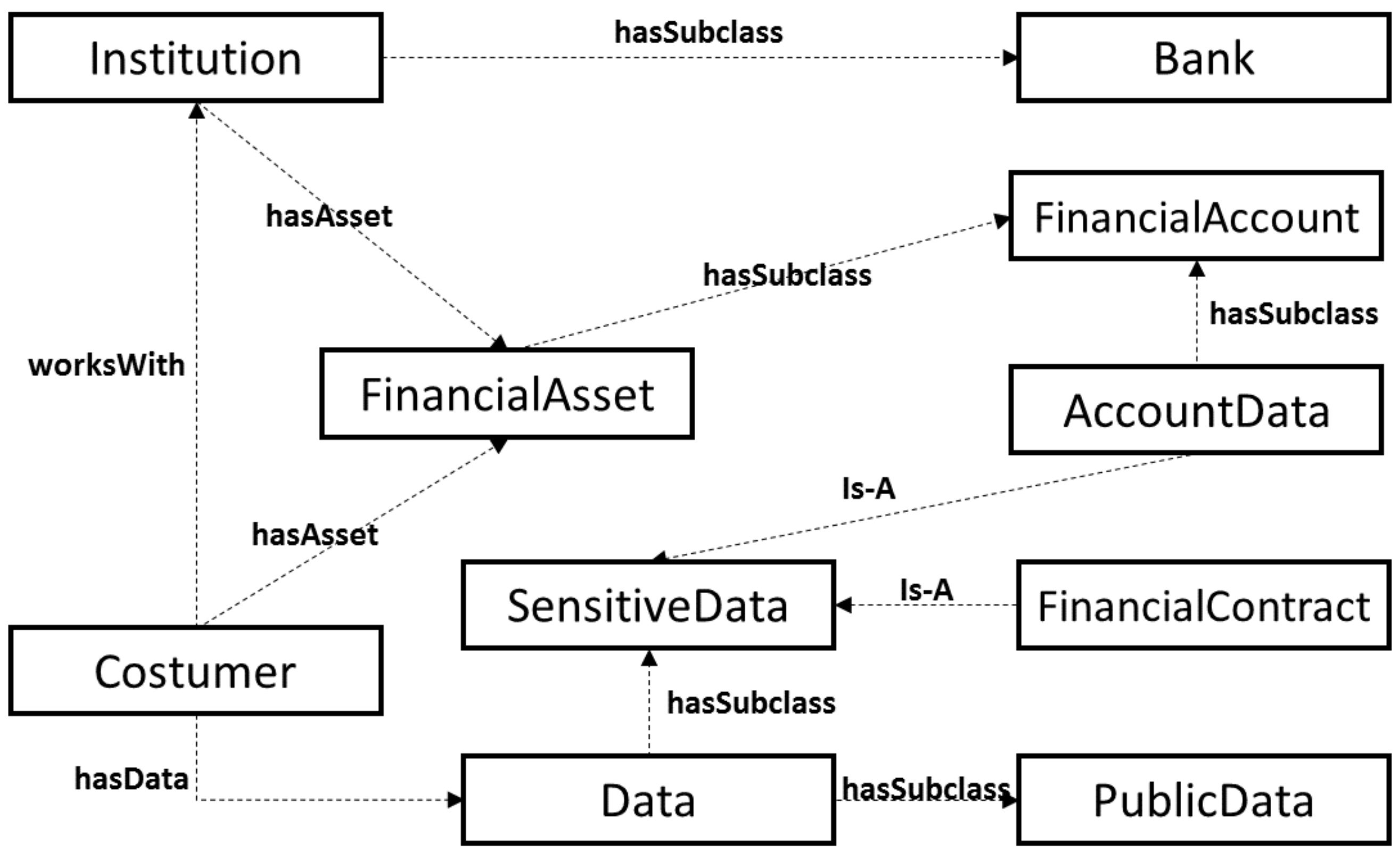
Cluster SQL queries to identify normal data access patterns



Insider threat modeling

- Created amd tested multiple threat models on the detection mechanism along with proving the complexity of the mechanism [P3][P4]
- Modeled an insider threat ontology in financial domain (see **Figure A**) for transforming anomaly detection into misuse detection [P3]

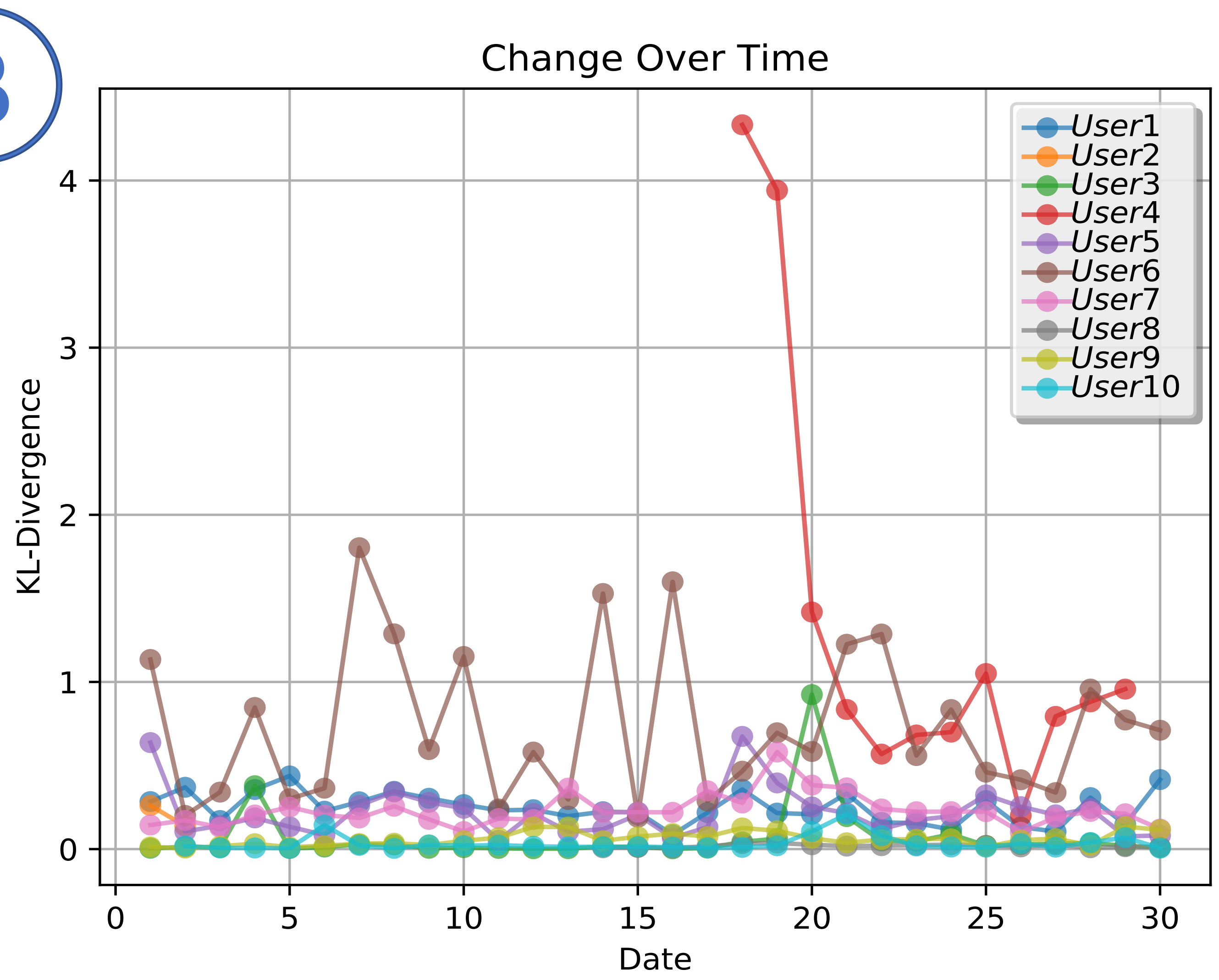
A



Database Activity Monitoring

- Construct user profiles by accumulating the extracting features for each user for a given period of time
- People who work in the same role in the organization can have different work habits, styles and priorities
- The expectation of behavior drift changes for different roles, and for different people as shown in **Figure B**
- The behavior patterns of tasks can change the temporal drift of a profile as can be seen in **Figure B**

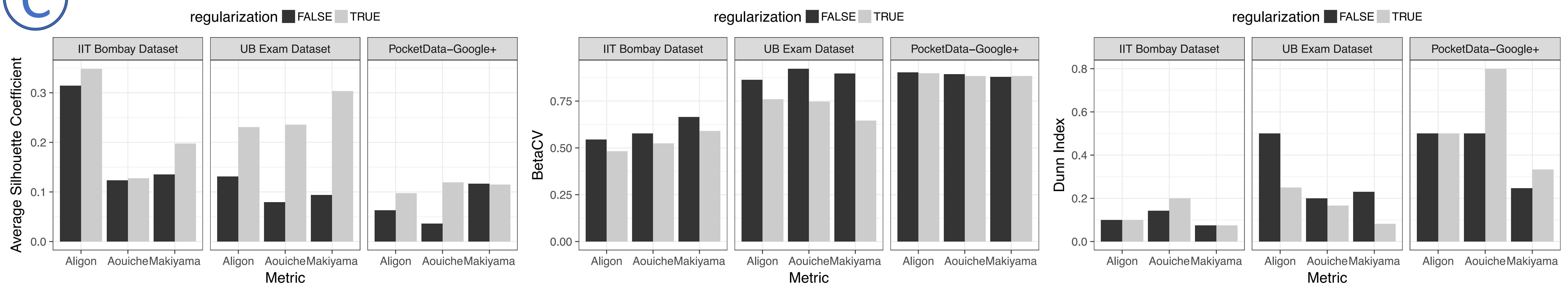
B



Clustering SQL queries

- Hierarchial clustering on SQL queries [P1]
- Extracted features from SQL queries: Makiyama [1], Aoiche [2], Aligon [3]
- Evaluated the feature extraction methods with three metrics: average Silhouette coefficients, BetaCV, Dunn Index [4] (see **Figure C**)
- Applied query-rewriting techniques (regularization) to improve the quality of features extracted (see **Figure C**) [P3][P5][P6]

C



(a) Average Silhouette Coefficient
(*larger* values are better)

(b) BetaCV
(*smaller* values are better)

(c) Dunn Index
(*larger* values are better)

Publications and Products

[P1] Kul, G., Luong, D., Xie, T., Coonan, P., Chandola, V., Kennedy, O., & Upadhyaya, S. (2016, April). Ettu: Analyzing query intents in corporate databases. WWW 2017 Companion.

[P2] Kul, G., & Upadhyaya, S. J. (2015). Towards a Cyber Ontology for Insider Threats in the Financial Sector. JoWUA, 6(4), 64-85.

[P3] Kul, G., Luong, D., Xie, T., Coonan, P., Chandola, V., Kennedy, O., & Upadhyaya, S. (2016). Summarizing Large Query Logs in Ettu. arXiv preprint arXiv:1608.01013.

[P4] Kul, G., Upadhyaya, S., & Hughes, A. (2017). Complexity of Insider Attacks to Databases. ACM CCS MIST 2017.

[P5] EttuBench – A SQL Query Similarity Metric Benchmark
https://github.com/UBOdin/EttuBench

[P6] The UB Exam Dataset
http://odin.cse.buffalo.edu/public_data/2016-UB-Exam-Queries.zip

References

[1] Makiyama, V. H., Raddick, J., & Santos, R. D. (2015, September). Text Mining Applied to SQL Queries: A Case Study for the SDSS SkyServer. In SIMBig (pp. 66-72).

[2] Aouiche, K., Jouve, P. E., & Darmont, J. (2006). Clustering-based materialized view selection in data warehouses. In Advances in Databases and Information Systems (pp. 81-95). Springer Berlin/Heidelberg.

[3] Aligon, J., Golfarelli, M., Marcel, P., Rizzi, S., & Turricchia, E. (2014). Similarity measures for OLAP sessions. Knowledge and information systems, 39(2), 463-489.

[4] Zaki, M. J., Meira Jr, W., & Meira, W. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.