ITU KALEIDOSCOPE
NANJING 2017
Challenges for a data-driven society

Fostering Smart City Development in Developing Nations: A Crime Series Data Analytics Approach

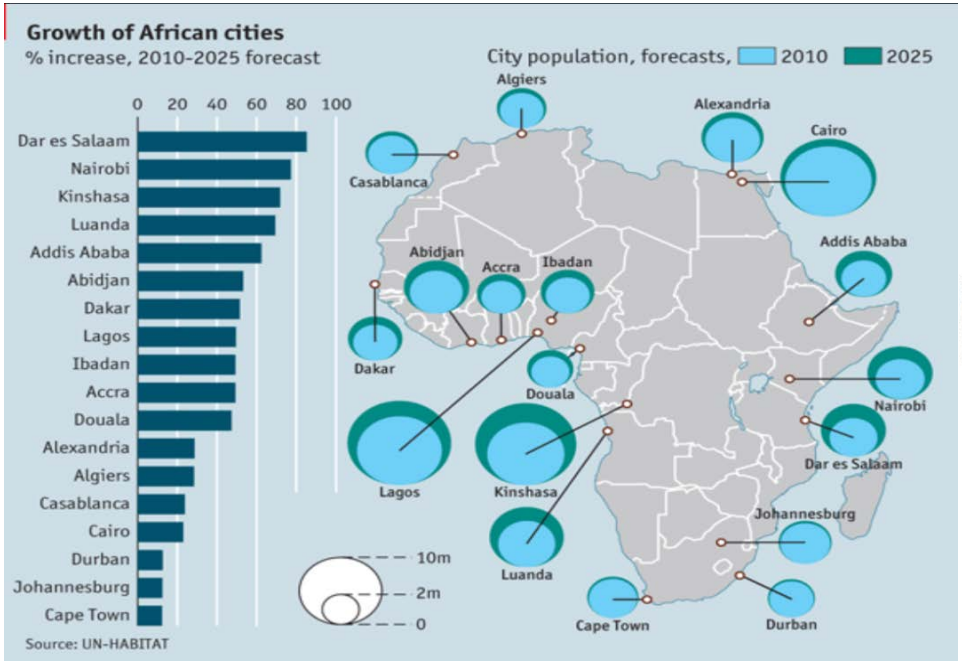
Omowunmi Isafiade
University of the Western Cape
oisafiade@uwc.ac.za

Nanjing, China
27-29 November 2017

Outline

- Introduction
  - Motivation and challenges
- Crime Control in South Africa
  - Current practice and gaps
- Proposed Intervention: CriClust System
  - Model formulation and design methodology
  - Results and discussion
- Conclusion and Outlook
- References
- Acknowledgements

Rapid urbanisation: More people live in cities than rural areas



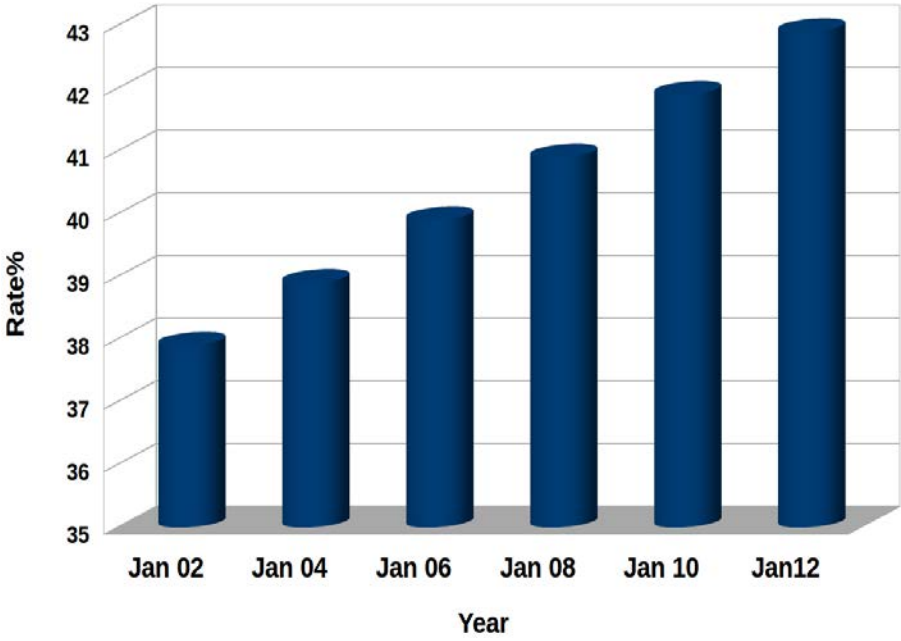
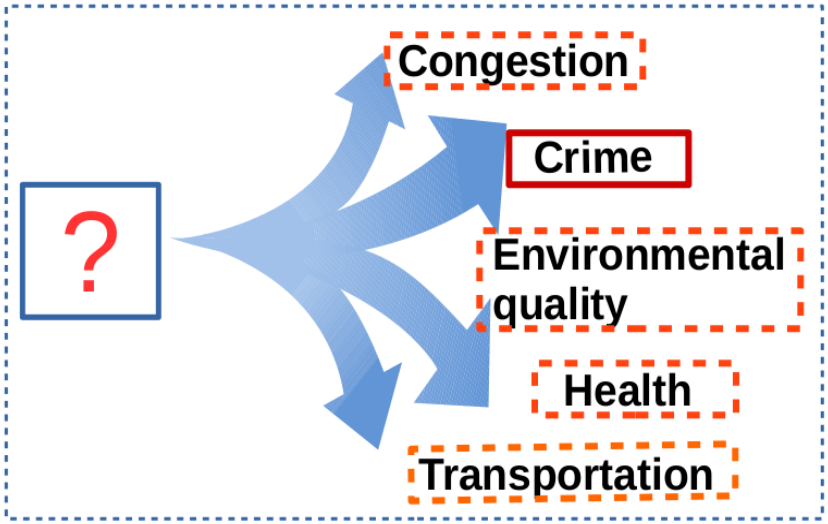


Fig: Growth of African cities

Fig: Trend in South Africa

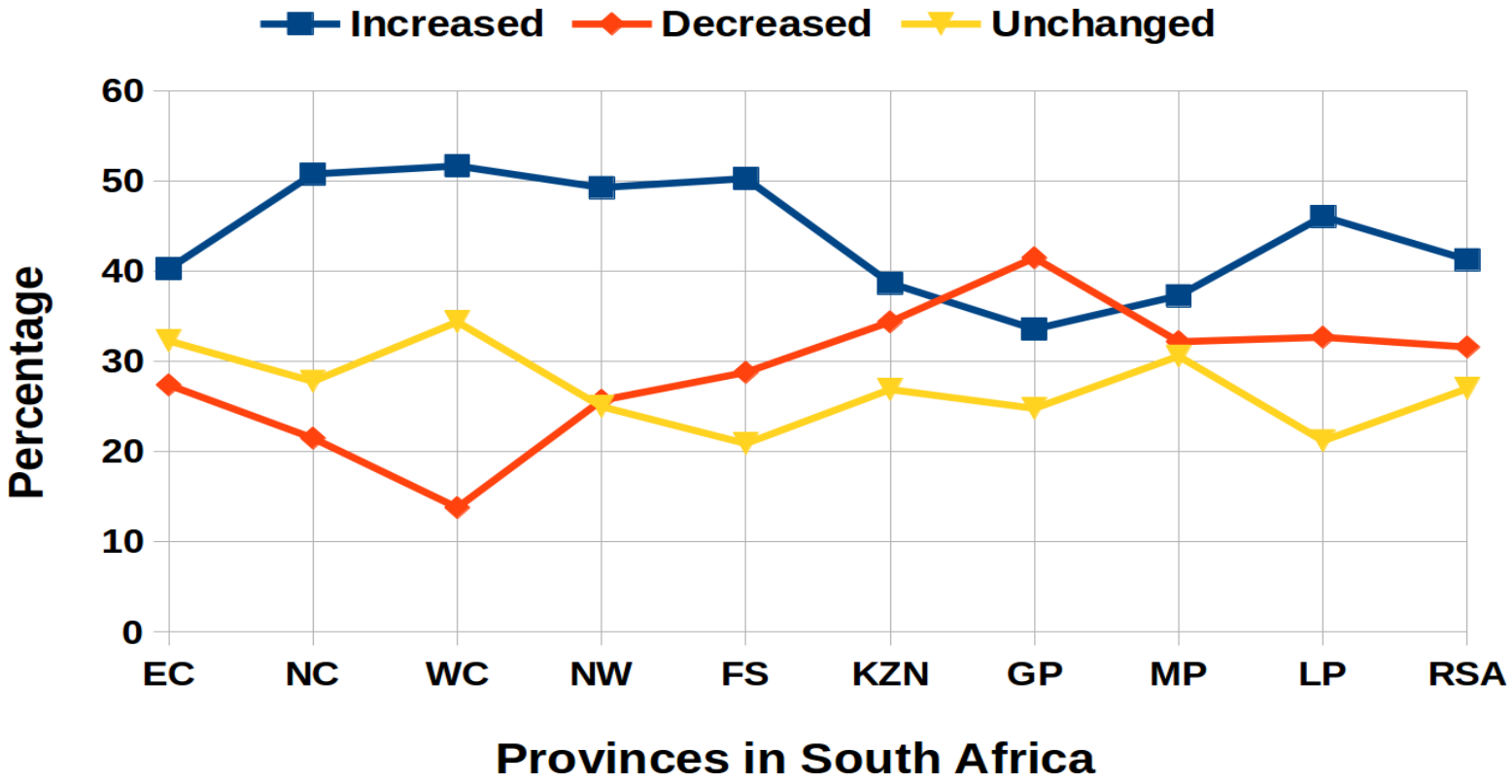
- More than 50% of the world's population lives in cities.
- Continued rural-urban migration forecasted up to year 2025.

Smart City: Meeting the challenges of rapid urbanisatiion



- **Smart city**: using urban informatics and technology to improve the quality and efficiency of urban operation and services.

- Increase in crime anticipated with rapid urbanisation.
- Deterring crime is a top priority for realising a sustainable “safe and smart” city.
- The use of armed weapons is not sufficient to tackle crime.



- Fig: Distribution of crime across provinces in South Africa

Current Practice	Gaps/Limitations
- Random patrols at locations	• *Police: Citizen* ratio is 1:347 (288 police per 100,000 citizen)
- Manual means of data capture & processing (using excel software) - Accumulated data is transferred to provincial level for processing	• Limited technological tool for pattern detection • Delay in knowledge discovery (inaccuracies) • Mitigation practices hindered due to lack of domain experts and technological tools (e.g. Analyst's Notebook)
- CrimeHub statistics [Institute for Security Studies (ISS)]	• General background information – may not be actionable

Problem Statement: Challenges of squeezing crime to zero

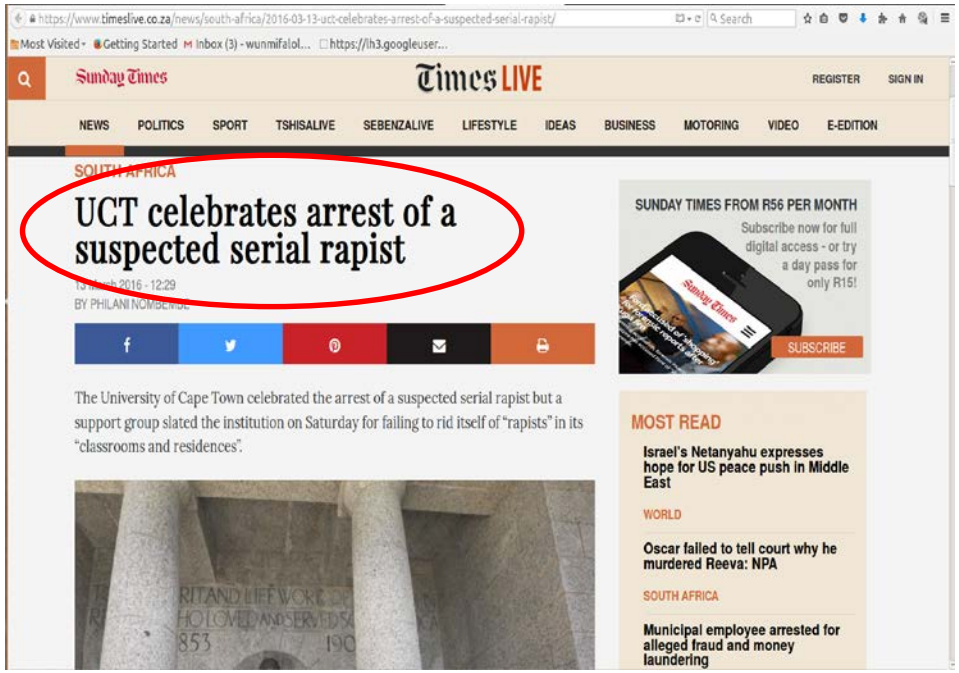
- Despite the vast resources allocated to crime, people still fall victim of crime
- Plethora of under-utilised crime reports archived by public safety.
- Manual means adopted at local stations is a huge constraint to effective policing in developing nations (e.g. South Africa).
- Need to promote knowledge-driven decision support for public safety improvement in developing nations.
- Crime series pattern (CSP) detection is less explored in developing nations

CriClust: Crime Series Pattern (CSP) detection

- Depiction of serial predator in related crime scenarios in a city.

- Research shows that many crimes are due to repeat (serial) offenders: **crime series**.
- Crime series are crimes committed by same offender.
- If patterns are identified timeously police can prevent further recurrence.
- Several tools exist but mostly able to estimate background information.

Crime Series Pattern (CSP) Detection



- CriClust serves to assist in CSP detection using rape data.
- However, can extend to other forms of crime
- Issues around rape and sexual violence still an ongoing concern in South African communities.
- Hence, crucial to devise smart means of assisting police in developing nations

Phases in CriClust System

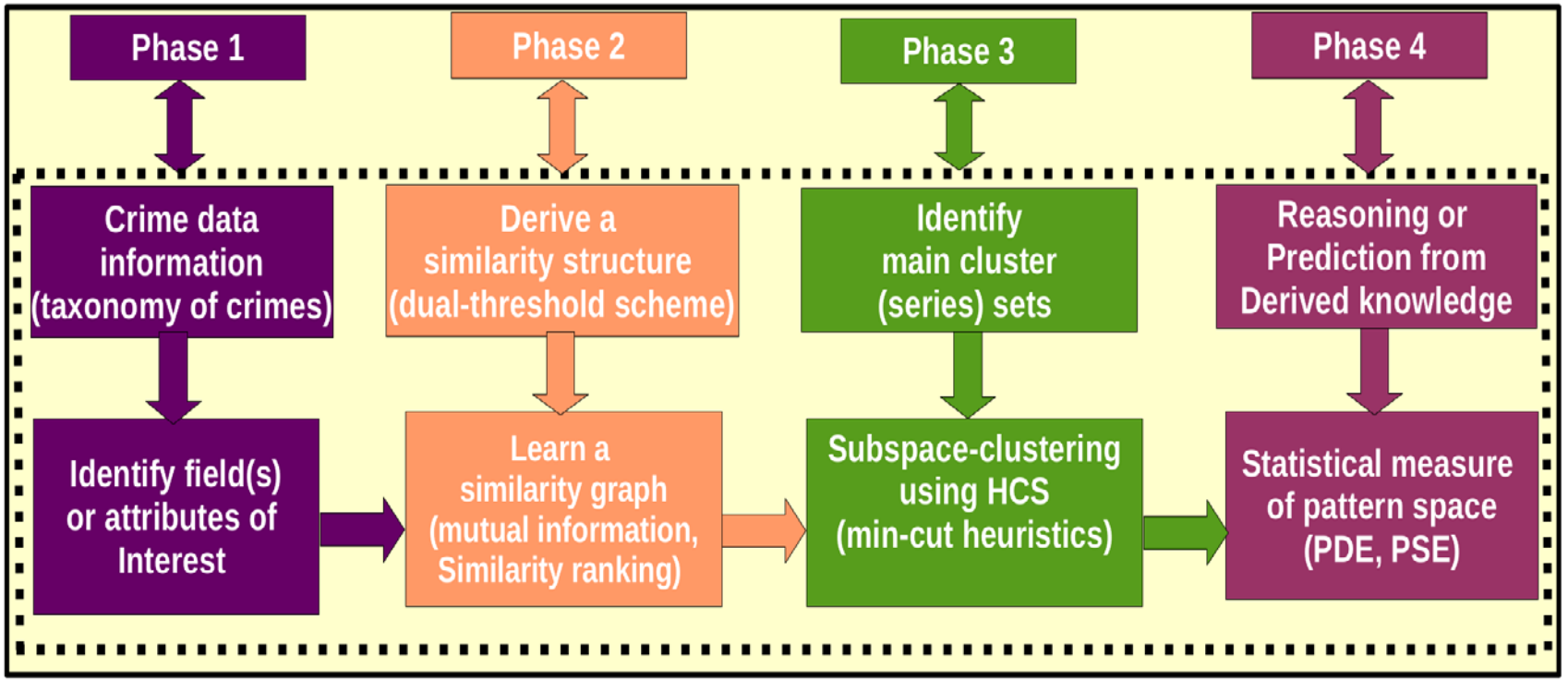


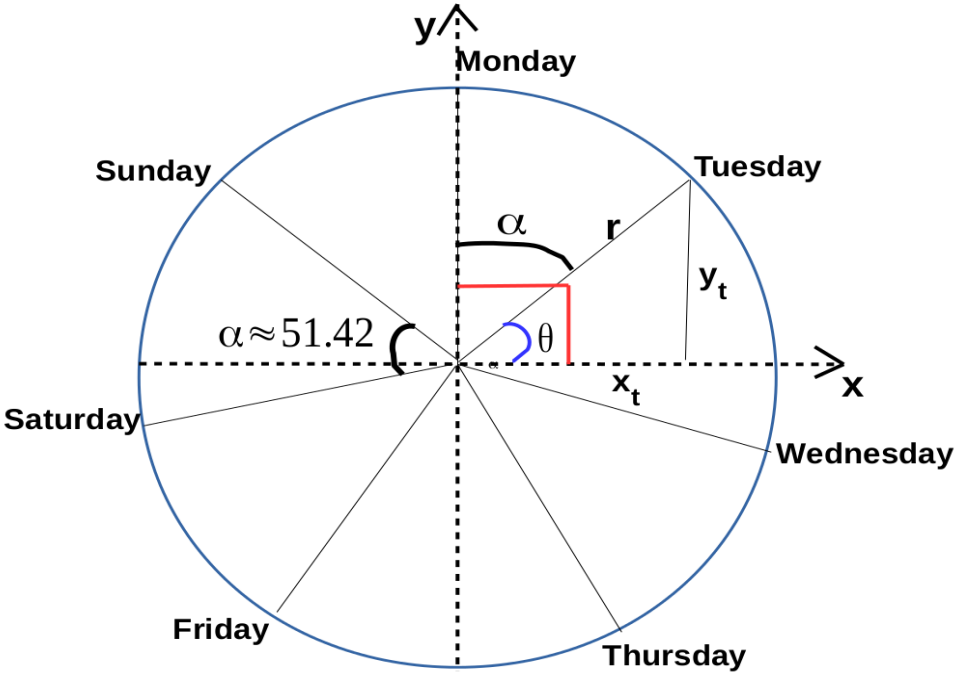
Fig: An overview of research phases in CriClust System

CriClust: Problem definition

	A_1	A_2	$\cdots$	A_n
C_i	A_{1_i}	A_{2_i}	$\cdots$	A_{n_i}
C_j	A_{1_j}	A_{2_j}	$\cdots$	A_{n_j}
C_k	A_{1_k}	A_{2_k}	$\cdots$	A_{n_k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Let *C* be a set of crime objects, where each *i* in *C* is defined by a set of attributes *A*, our interest lies in crime objects that exhibit a coherent pattern on a subset of *A.*

Crime Series Pattern Detection: Algorithmic process



- $\alpha = \frac{360^0\ (in\ a\ circle)}{7\ (days\ in\ a\ week)}$
- α is the angle between each pair of days
- The 2-D component is relevant because a 1-D component will assume that Sunday is far from Monday.

CriClust: Learning the similarity graph

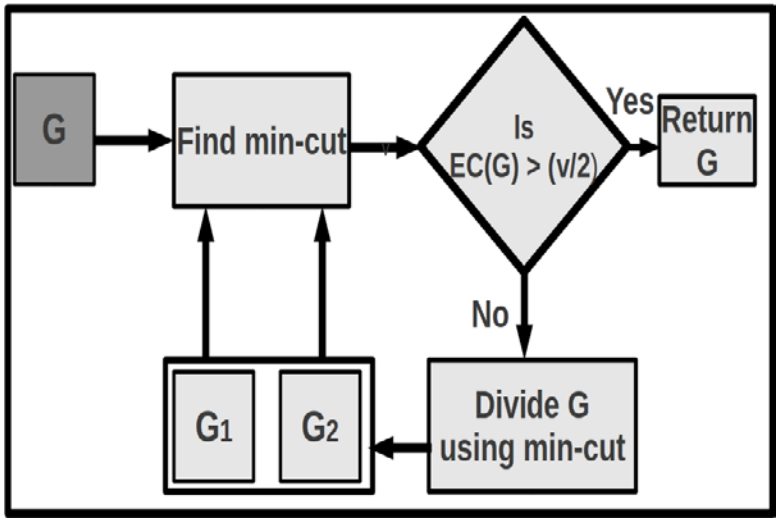


Fig: Flow of highly connected Subgraphs (HCS)

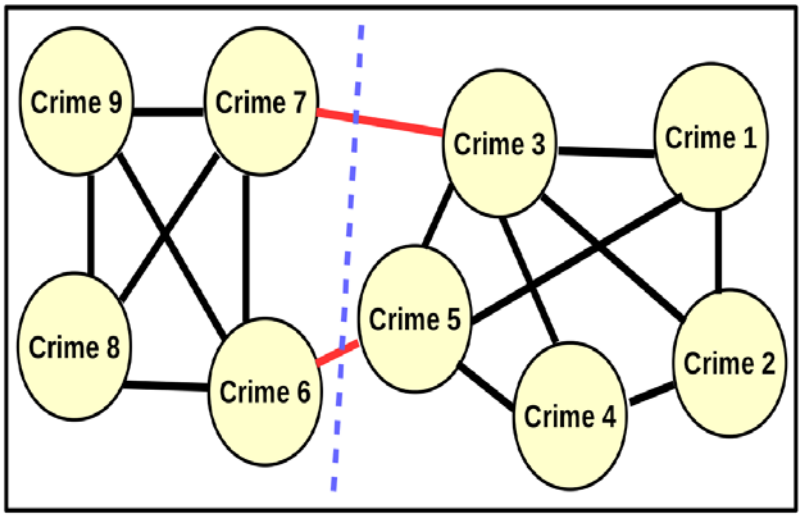
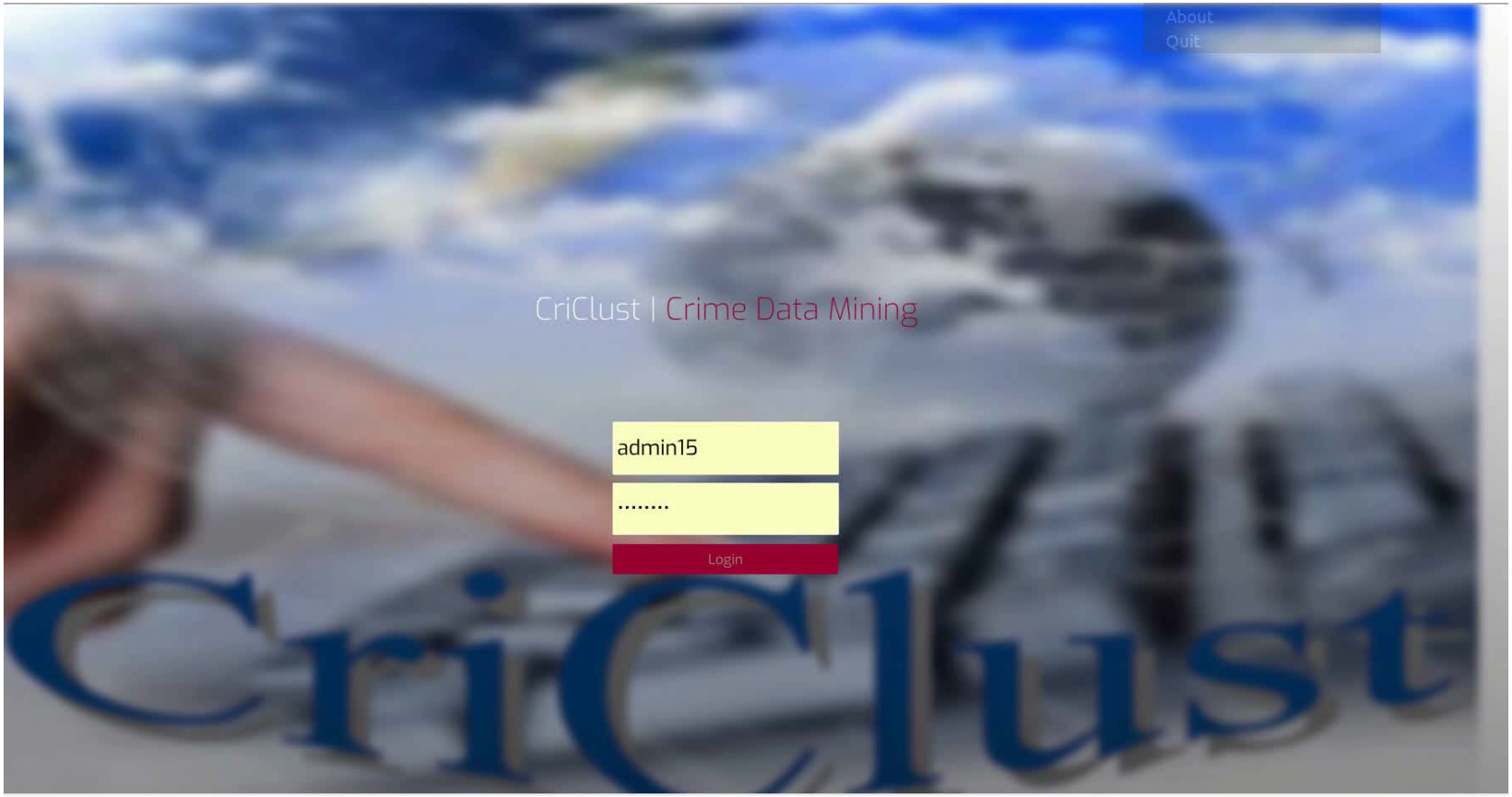


Fig: A depiction of crime cluster detected by HCS

Overview: CriClust System Visualisation



CriClust: Scalability and trend of series observed

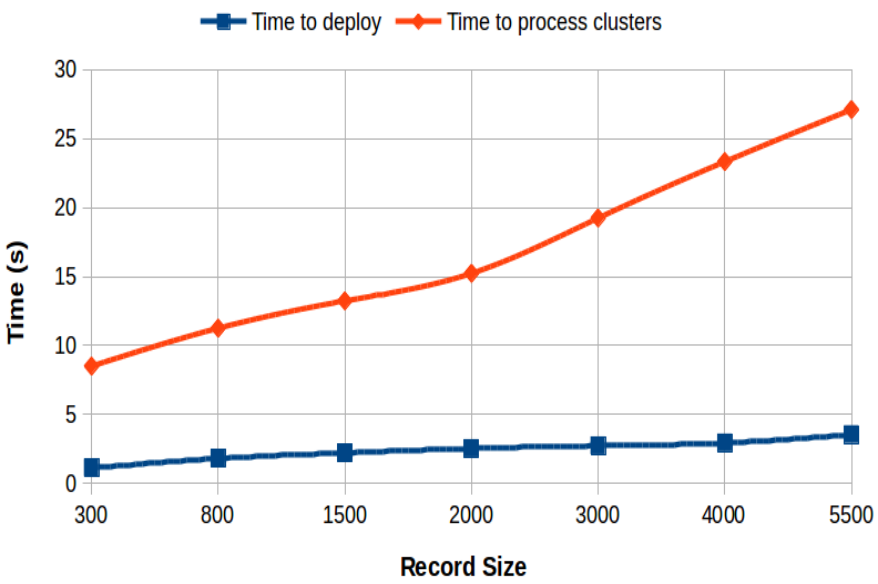


Fig: Scalability trend

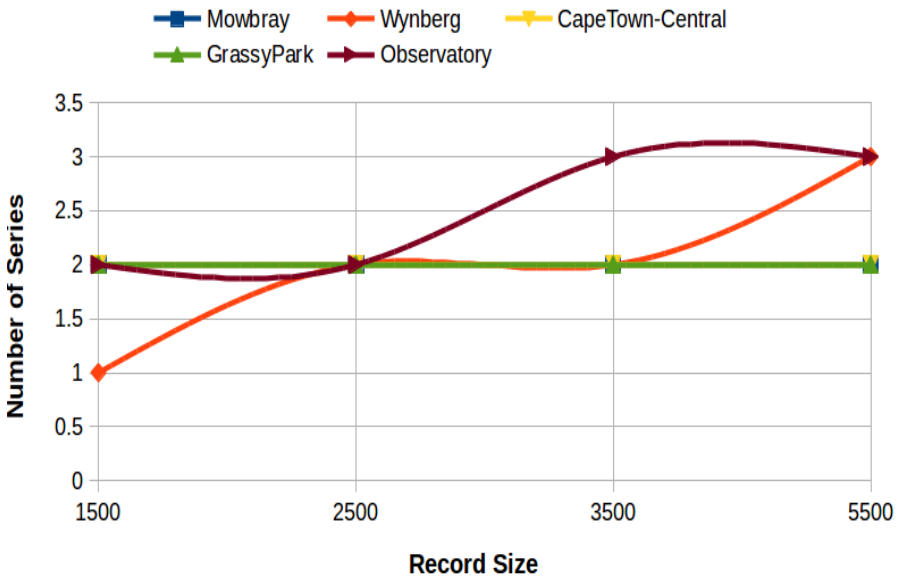


Fig: Trend of series observed across locations

Characterising features emerging for each series

S/N	Location	PDE(%)	Day	Time	Vic	Sus	VAge	SAge	SFr	Mot	MO	HCol	Mask	Sub-Ab
1	Mowbray	35 (S1)	1	1	1	1	0	0	1	0	1	1	0	0
2		65 (S2)	1	1	1	1	0	1	1	1	1	1	1	1
3	CapeTown	50 (S1)	0	1	1	1	0	0	1	1	1	1	0	0
4	Central	50 (S2)	1	1	1	1	0	1	1	0	1	0	0	0
5	Wynberg	40(S1)	1	1	1	1	0	1	1	0	1	1	0	0
6		34(S2)	1	1	1	0	1	0	0	1	1	0	1	1
7		26(S3)	0	1	0	1	0	1	1	0	1	1	0	0
8	Grassy-	21(S1)	1	1	1	0	0	0	0	1	1	1	0	1
9	Park	79(S2)	1	1	1	1	0	1	1	1	1	0	1	0

Systematic comparison of CriClust with existing research

S/N	Features	Crime Linkage **[5]**	Mining Rotten Core **[6]**	Serial Crime Pattern **[7]**	Crime Linkage **[8]**	**CriClust Model (our work)**
1	Exploratory basis	Crime linkage	Crime series detection	Serial criminal pattern detection	Crime linkage	Crime series detection
2	Crime explored	Breaking & entering crimes	Burglary (housebreaking)	Armed robberies	Burglary crimes	Sexual crime
3	Modelling approach	Statistical approach	Conventional optimisation	Neural Network (NN)	Bayesian Network (BN)	Dual threshold scheme & graphical model
4	Techniques used	Bayes factor, Hierarchical clustering	Integer linear programming, clustering, BFS	Cascaded network of Kohohen NN	Bayes Network	Geometric projection, HCS clustering
5	Empirical observation	Posterior odds, Bayes factor & *number of clusters*	Map location of series, pattern space, precision & recall	Percentage of predicted & actual patterns	Posterior probabilities & BN	Map (PDE, PSE) of series, *scalability,precision* & recall

Summary and Conclusion

- Challenge of crime is magnified in resource constraint settings.
- Police need to be empowered with context-aware and cost-effective technologies for effective policing.
- Crime series detection is less explored in developing nations
- CriClust serves to assist in crime series identification, using a dual threshold mechanism and geometric projection.
- CriClust is not a panacea but can assist with underperformance in policing.
- CriClust is to be considered for deployment with the police, and there is an ongoing collaboration with an NGO on community policing.

References

1) A. Milgram: *Why smart statistics are the key to fighting crime*. TED Talk, Retrieved from: https://www.ted.com/talks, 2013.
2) International Association of Crime Analysts (IACA): *Crime Pattern Definitions for Tactical Analysis. Standards, Methods, and Technology* (SMT) Committee White Paper, 2011.
3) S. Writer : *South African Police force vs the world*, Business Tech, 2015
4) F. Húffner and C. Komusiewicz and M. Sorge: *Finding Highly Connected Subgraphs*, 41st International Conference on Current Trends in Theory and Practice of Computer Science, Pec pod Sněžkou, Czech republic pp 1-20, 2015.

References

5) M. D. Porter: *A statistical approach to crime linkage*. arxiv e-prints. pp. 1-33, 2014.

6) T.Wang, C. Rudin, D.Wagner, and R. Sevieri*, Finding patterns with a rotten core: Data mining for crime series with core sets*. Big Data, vol. 3, no. 1, pp. 3-21, March 2015.

7) K. Dahbur and T. Muscarello*. Classification system for serial criminal pattern.* Artificial Intelligence and Law, 11(4):251-269, 2003.

8) J. Zoete, M. Sjerps, D. Lagnado, and N. Fenton: *Modelling crime linkage with Bayesian networks.* International Journal of Science and Justice: Elsevier Ireland, 7(5):1-9, 2014

Acknowledgements

SOUTH AFRICAN POLICE SERVICE

RESPICE PROSPICE
UNIVERSITY of the WESTERN CAPE

NRF National Research Foundation

Thank You