# Comparing the Sensitivity of the Spectral Graph and Core Resilience Model in Assessing SANReN's Network Resilience

Alexander White whtale015@myuct.ac.za Computer Science Department, University of Cape Town

#### **Abstract**

The South African National Research and Education Network (SAN-ReN) underpins teaching and research by connecting universities and research institutions nationwide, making the resilience of its physical topology critical to the continuity of service. This study evaluates the sensitivity of two complementary diagnostic approaches: the Spectral Graph Model, using algebraic connectivity to quantify sensitivity to fragmentation, and the Core Resilience Model, using Core Influence-Strength to assess layered redundancy. Both models are applied to multiple SANReN topologies under simulated node failures, with removal strategies ranging from random and targeted attacks to decile-based scenarios. Degradation curves are generated and statistically analysed to compare model performance. Results show that the Spectral Graph Model is significantly more sensitive in most scenarios, particularly under random failures, descending targeted attacks, and direct removal of core nodes, confirming its value as an early-warning indicator of fragmentation risk. By contrast, the Core Resilience Model is more sensitive when failures progress from the periphery inward, reflecting its focus on structural embedding. These findings demonstrate that the two models provide complementary perspectives on resilience, and together offer a richer basis for monitoring and strengthening research networks such as SANReN.

### 1 Introduction

With the emergence of the World Wide Web in the late 1990s, education and research communities quickly recognised that standard internet technologies could not handle the large-scale data transfers their work required [10, 19]. Building on early private university networks, countries began developing National Research and Education Networks (NRENs) to provide specialised connectivity for teaching and research [14]. Today, NRENs refer both to the physical network and the organisations that manage them [10]. Rather than building their own infrastructure, NRENs typically negotiate with ISPs to secure high-capacity, affordable bandwidth [14]. By aggregating demand across institutions, NRENs strengthen bargaining power and lower costs - a model shown to save costs for members significantly, as with ZAMREN in Zambia [14]. NRENs also interconnect via Regional Research and Education Networks (RENs), such as UbuntuNet in Southern Africa, creating continental and global backbones for collaboration [14].

The value of NRENs lies in enabling secure, high-speed data transfer that supports modern education and research [9, 19]. Students gain faster access to digital resources, while researchers can share large datasets, access costly scientific instruments remotely, and participate in international projects [19]. Video conferencing and

collaborative tools further extend teaching and research opportunities across borders [10, 14]. In South Africa, this role is fulfilled by SANReN, which connects over one million users across universities, research bodies, and science councils, linking them locally and internationally through high-speed terrestrial and undersea connections [7, 28].

Like other NRENs, SANReN faces ongoing risks that threaten the reliability of its services [2, 9]. The most immediate set of risks concerns network resilience. Distributed Denial of Service (DDoS) attacks frequently target SANReN's infrastructure, attempting to overwhelm it with traffic and disrupt access to critical academic platforms [3]. Such disruptions highlight a broader vulnerability: complex, highly connected networks like SANReN are difficult to secure against both random failures and coordinated, targeted attacks [12, 22]. A central problem is that while diagnostic models (like the Spectral Graph Model and Core Resilience Model) exist to measure resilience, it remains unclear whether they differ in sensitivity when testing SANReN's topology [2, 20]. For example, one model may detect risks related to connectivity collapse, while another highlights the structural redundancy of the network core. Without clarity on these differences, it is difficult to know which model provides more useful insights for strengthening SANReN against failures and attacks.

Ensuring the resilience of SANReN is critical because disruptions have consequences far beyond connectivity: they delay research projects, restrict student access to digital resources, and undermine national and international collaborations that depend on high-speed data sharing [2, 20]. As South Africa's NREN, SANReN supports over a million users across universities, science councils, and research institutes [7]. Failures or successful attacks can therefore jeopardise not only teaching and learning but also access to costly scientific infrastructure and participation in global research initiatives. By addressing vulnerabilities and identifying the most effective diagnostic models, it becomes possible to strengthen SAN-ReN's ability to withstand failures and targeted attacks, thereby safeguarding the continuity of academic and research activity.

This study focuses specifically on the problem of how best to evaluate SANReN's structural resilience. While risks such as DDoS disruptions and targeted node removals are well recognised, it remains unclear whether different diagnostic models highlight the same vulnerabilities or respond differently to attack scenarios. Choosing the wrong diagnostic perspective has practical consequences: it could leave structural weaknesses undetected, delay mitigation, and ultimately expose SANReN's one million users to avoidable

disruption. More broadly, clarifying how different resilience models perform in practice contributes to ongoing debates in network science about whether global connectivity metrics or core-based redundancy metrics offer the more reliable basis for strengthening critical infrastructure.

To address this gap, two graph-theoretic models are compared: the *Spectral Graph Model*, using algebraic connectivity ( $\lambda_2$ ) to detect fragmentation risk (i.e., the point at which the network separates into disconnected components), and the *Core Resilience Model*, using Core Influence-Strength to measure layered redundancy (i.e., the extent to which peripheral nodes provide backup paths into the network core). By simulating progressive node removals across diverse disruption scenarios (including random, descending targeted, ascending targeted, and decile-based failures) this study evaluates how the two models differ in diagnostic sensitivity.

The research question guiding this work is: How does the diagnostic performance of the Spectral Graph Model compare to that of the Core Resilience Model in accurately identifying resilience weaknesses and structural vulnerabilities in SANReN's physical topology under simulated failure scenarios?

From this question, the specific aims of the study are to: (1) **test whether**  $\lambda_2$  and Core Influence-Strength differ significantly in their sensitivity under controlled failure scenarios; and thus (2) **identify the conditions** under which the Spectral Graph and Core Resilience models produce divergent or convergent results, thereby evaluating whether they highlight the same or different structural vulnerabilities.

# 2 Background & Related Works

Network resilience refers to a network's ability to maintain suitable levels of connectivity and functionality when subjected to failures or targeted attacks [25]. **Global resilience** captures the preservation of overall connectivity during large-scale disruptions, while **structural resilience** describes how a network withstands more localised failures and fragmentation [26, 30].

Research specifically focused on SANReN remains sparse. Salie [22] highlighted high delays in Cape Town and Gqeberha due to circuitous routing and geographic isolation, but the work was descriptive and did not incorporate simulation, graph-theoretic metrics, or structural modelling. Beyond SANReN itself, however, two principal approaches have emerged for assessing network resilience: the well-established *Spectral Graph Model* and the more recent *Core Resilience Model* [4, 11, 16, 17].

The *Spectral Graph Model* provides a standard, globally focused view of resilience [4, 5]. Its metrics, such as algebraic connectivity, are widely implemented in *NetworkX* and extensively used in graph-theoretic studies of network robustness [1, 4, 17]. Algebraic connectivity ( $\lambda_2$ )—the second-smallest eigenvalue of the Laplacian spectrum—is particularly significant: if  $\lambda_2 > 0$  the graph is connected, while larger values indicate greater resilience to fragmentation [4, 24]. Conversely, low  $\lambda_2$  values signal structural weak points

and heightened risk of disconnection [4]. This makes  $\lambda_2$  a natural and established measure of global connectivity risk.

In contrast, the *Core Resilience Model*, proposed in Laishram's doctoral research [15, 16], focuses on redundancy and structural embedding revealed through *k*-core decomposition [16]. By characterising how deeply nodes are embedded in the core layers, this model emphasises resilience in terms of layered redundancy rather than global connectivity. Its key composite metric, *Core Influence-Strength*, integrates two diagnostics: **Core Strength**, which measures the stability of a node's *k*-core membership under neighbour loss, and **Core Influence**, which reflects how strongly a node supports higher-coreness neighbours [16]. Together, these yield a network-level indicator of whether structurally influential nodes are well-embedded or fragile [16].

By design, the Spectral Graph Model and the Core Resilience Model capture different dimensions of resilience:  $\lambda_2$  quantifies the risk of global fragmentation, while Core Influence-Strength highlights structural redundancy within the network core. Both models are theoretically compelling, yet they may diverge in sensitivity under failure conditions. Importantly, although the Spectral Graph Model is established and embedded in standard graph libraries, the Core Resilience Model is relatively new and has not yet been tested comparatively against spectral measures in empirical studies of SANReN or other networks [4, 6, 16].

This study addresses that gap. No prior work has systematically compared the diagnostic sensitivity of these two models under simulated node failures. Since sensitivity, the ability to detect and respond to structural collapse as it develops, is central to assessing resilience, it is essential to evaluate which model more effectively identifies weaknesses. Doing so provides insight not only for SANReN's topology management and risk mitigation strategies, but also for broader debates on whether redundancy-focused or connectivity-focused models should be prioritised in resilience diagnostics.

# 3 Experimental Design & Implementation

### 3.1 Metrics

To operationalise the comparison of the Spectral Graph Model and Core Resilience Model within this study, one representative metric was selected for each model: algebraic connectivity ( $\lambda_2$ ) for the Spectral Graph Model, and Core Influence-Strength for the Core Resilience Model. These were chosen because they capture complementary aspects of structural resilience: global connectivity versus layered redundancy. Their definitions and theoretical basis are discussed in Section 2: Background & Related Works.

### 3.2 Datasets

The study utilised topologies for 9 subgraphs of the SANReN network. These were provided by SANReN in .tgf files and represent the municipalities of Bloemfontein, Cape Town, Durban (eThekwini), East London, Johannesburg, Polokwane, Pretoria (Tshwane), Pietermaritzburg, and Vanderbijlpark. The .tgf files included the name and numbers of each node in a given topology as well as

the links between them. Nodes in each topology are identified by numbers.

#### 3.3 Methods

This study evaluates the comparative sensitivity of two resilience models, the *Spectral Graph Model* and the *Core Resilience Model*, under simulated node failures on the SANReN network. The overarching aims of the study are twofold: (1) to test whether algebraic connectivity ( $\lambda_2$ ) and Core Influence-Strength differ significantly in their sensitivity under controlled failure scenarios, and (2) to identify the conditions under which the two models converge or diverge in their diagnostic behaviour. To address these aims formally, the following hypotheses are tested under a variety of node-removal scenarios (experiments):

Null Hypothesis ( $H_0$ ): There is no statistically significant difference in the sensitivity of the two models.

# **Alternative Hypotheses:**

H<sub>1</sub>: The Spectral Graph Model is more sensitive, demonstrating earlier and steeper degradation in metric value under node removal.
 H<sub>2</sub>: The Core Resilience Model is more sensitive, demonstrating earlier and steeper degradation in metric value under node removal.

3.3.1 Tests. In this study, a test refers to a single simulation in which a specific topology is subjected to a defined sequence of node removals, with both resilience metrics recalculated after each removal.

For every test, the corresponding topology is ingested from a .tgf graph file and converted into a NetworkX graph object. The initial values of the chosen metrics are then calculated: algebraic connectivity ( $\lambda_2$ ) for the Spectral Graph Model and Core Influence-Strength for the Core Resilience Model. Algebraic connectivity ( $\lambda_2$ ) is computed directly from the Laplacian matrix of the graph using NetworkX's built-in spectral routines.

For Core Influence-Strength, no direct implementation exists in NetworkX, so it was calculated following Laishram's formulation [15, 16]. As outlined in the background research, the procedure involves: (i) obtaining each node's core number via k-core decomposition, (ii) computing Core Strength for all nodes, (iii) deriving Core Influence from the leading eigenvector of an influence matrix M that captures connectivity toward equal- or higher-coreness neighbours, and (iv) averaging the Core Strength of the most influential fraction of nodes  $S_f$  (those above the f-th percentile of Core Influence, with f=0.9 in this study) [15, 16]. This method ensures that Core Influence-Strength captures the resilience of the most structurally embedded nodes, making it a suitable counterpart to algebraic connectivity ( $\lambda_2$ ). A high-level pseudocode outline is given in Appendix B to support reproducibility based on the full definitions and justification provided in Laishram's original work [15].

After the initial metrics are calculated, nodes are removed sequentially in the specified order. Removals are *cumulative*: once a node is removed it remains absent for the rest of the test. After each removal, both metrics are recalculated. These values are normalised

relative to their initial values and expressed as percentages to construct degradation curves that show how resilience changes as more nodes are lost.

All experiments follow a *progressive cumulative* design. Rather than removing only a fixed set of nodes, each test is structured into stages at 5%, 10%, 15%, 30%, and 60% of the topology size (with truncation for very small graphs). This ensures that the degradation of each metric is observed at multiple depths of disruption, producing a sequence of intermediate states rather than a single before/after snapshot. The progressive design therefore increases the robustness of results by capturing sensitivity across different stages of failure.

The order in which nodes are removed is either random or targeted, depending on the test. For random experiments (A, D, G, J, M), this progressive cumulative design is repeated 20 times per topology (or restricted subset of a topology for top/bottom decile tests). In each repetition, nodes are selected via uniform random sampling without replacement to construct a staged removal sequence at 5%, 10%, 15%, 30%, and 60%. This repeated sampling not only captures variability across multiple randomisations, but also avoids the risk that results could be driven by a single arbitrary random seed [13, 18]. Instead, outcomes reflect the distribution of possible random failures, providing a more reliable basis for statistical comparison.

To conduct targeted node removal experiments, each topology's nodes were ranked in order of criticality and then subjected to the same progressive cumulative design. Critical nodes are those most likely to fragment the network or cause significant failures if removed. Rankings were derived by a different group member from the project group using a composite centrality score, computed as the average of betweenness, closeness, and degree centrality (each normalised to [0,1]). The higher the composite score, the more "critical" the node. These ordered lists formed the basis for descending, ascending, and decile-based progressive tests. Rankings for all topologies are included in Appendix A.

3.3.2 Experimentation Files. Experiments are defined via structured text files in which each line represents a distinct test. Each line contains the name of a topology and a list of nodes to remove in sequence. For example, a line such as cpt 5 6 4 3 refers to a test on the Cape Town topology where: the initial metrics for CPT are calculated; node 5 is removed and the metrics recalculated; node 6 is then removed (with node 5 already absent) and the metrics recalculated; node 4 is removed next, and so on. This ensures that each test traces a cumulative degradation path for that topology.

3.3.3 Area Under the Curve Calculation. For every degradation curve, the Area Under the Curve value is computed using numerical integration. A lower Area Under the Curve value corresponds to faster degradation and, by definition, higher sensitivity. Each test therefore results in a pair of Area Under the Curve values: one for the Spectral Graph Model and one for the Core Resilience Model. These values are stored for each test.

3.3.4 Statistical Analysis. After all tests in an experiment have been processed, the next step is to determine whether the two models differ significantly in their Area Under the Curve values. Each experiment therefore produces a distribution of paired Area Under the Curve differences (Spectral – Core), which is subjected to both hypothesis testing and uncertainty estimation.

To quantify uncertainty, 95% bootstrapped confidence intervals were generated for the mean Area Under the Curve difference using 10000 resamples. Bootstrapping is a non-parametric technique that resamples the observed data with replacement to construct an empirical distribution of the mean [8, 13]. It avoids assumptions of normality and provides robust error bounds even for small sample sizes, complementing formal hypothesis tests [8, 13].

For hypothesis testing, two candidate tests are appropriate: the paired t-test and the Wilcoxon signed-rank test [21, 27, 29]. Both assess whether paired differences are significantly different from zero, but they rest on different assumptions. The paired t-test requires the differences to be approximately normally distributed and is more powerful when this condition holds [21, 27]. The Wilcoxon signed-rank test, by contrast, is non-parametric and does not assume normality, but instead evaluates the median of the differences, making it more robust for skewed or non-Gaussian data [29].

A third test, the Shapiro–Wilk test, is therefore used to decide between the two. This test evaluates the null hypothesis that the paired differences are drawn from a normal distribution [23]. A non-significant result (p>0.05) indicates that the assumption of normality is reasonable [23], in which case the paired t-test is applied. If the result is significant ( $p\leq0.05$ ), normality cannot be assumed and the Wilcoxon signed-rank test is used instead.

The selected hypothesis test outputs both a test statistic and a p-value, which establish whether the null hypothesis ( $H_0$ : no difference in sensitivity) can be rejected. If p < 0.05, the sign of the difference indicates which model is more sensitive: the Spectral Graph Model ( $H_1$ ) or the Core Resilience Model ( $H_2$ ). Box-and-whisker plots are also generated to visualise the spread and central tendency of Area Under the Curve distributions for both models.

3.3.5 Validation. Reliability was strengthened through two complementary forms of validation: repetition and uncertainty estimation.

Firstly, repetition is built directly into the experimental design. Each experiment consists of multiple progressive cumulative tests (removals at 5%, 10%, 15%, 30%, and 60%) applied across all nine SANReN topologies. This ensures that outcomes are not dependent on a single removal sequence or graph structure [13, 18]. In addition, for random experiments (A, D, G, J, M), each topology was subjected to 20 independent random removal sequences. This repetition captures variance across multiple randomisations and prevents results from being an artefact of one arbitrary random seed [13, 18]. Together, these layers of repetition provide a broad

and diverse basis for statistical comparison.

Secondly, all results are accompanied by non-parametric error estimation. For every experiment, the distribution of paired Area Under the Curve differences was subjected to bootstrap resampling with 10000 iterations, generating 95% confidence intervals for the mean difference. Bootstrapping avoids strong distributional assumptions and quantifies the stability of the observed effects even for small samples [8, 13]. The fact that observed means consistently fell within their resampled confidence intervals confirms that the reported differences are not artefacts of single samples but reflect stable underlying trends.

3.3.6 Reproduction. All experiments are implemented in Python using the NetworkX and SciPy libraries. The codebase is modular and reproducible, with support for loading topologies, generating random tests, computing metrics, visualising degradation, and performing statistical testing. Experiment definitions are stored in plain-text files, with each file representing a full experimental setup and each line a specific node removal test. All experiment definitions used for this study are included in Appendix C.

3.3.7 Ethical Considerations. This study involved only simulated experiments on topology representations of SANReN's publicly documented topology. No human participants, personal data, or sensitive operational information were involved. Ethical risks were therefore minimal. Nevertheless, the research was conducted in line with good academic practice: all datasets were used for noncommercial, educational purposes, and results were reported transparently to avoid misrepresenting the resilience of SANReN or implying operational vulnerabilities beyond the scope of the study.

# 3.4 Experiments

To systematically operationalise the study's two aims: (1) testing whether algebraic connectivity ( $\lambda_2$ ) and Core Influence-Strength differ significantly in sensitivity under controlled failure scenarios, and (2) identifying the conditions under which the two models converge or diverge in their diagnostic behaviour, a structured suite of 15 experiments was designed. Each experiment consists of multiple tests, where a test involves removing a specific sequence of nodes from a topology. After each removal, the two resilience metrics are recalculated, the resulting degradation curves are generated, and the Area Under the Curve values are calculated as summary sensitivity measures. To quantify uncertainty, bootstrapped 95% confidence intervals for mean Area Under the Curve differences were also generated, complementing the formal hypothesis tests described in Section 3.3.4.

The experimental framework is organised along two primary dimensions: (i) the method of node removal (random versus targeted), and (ii) the scope of the topology set (full, small, or big). Additional experiments isolate the effects of removing only the top or bottom deciles of nodes by criticality. This systematic design enables direct evaluation of whether the models differ in sensitivity (addressing Aim 1) and clarifies the conditions under which their responses converge or diverge (addressing Aim 2). Aim 1 is addressed by calculating and statistically comparing the Area Under the Curve

**Table 1: Overview of Designed Experiments** 

Туре	ID	Description		
Full Topologies (All Graphs)		Random: Progressive removals across all topologies in random order.		
		<b>Descending Targeted (Most→Least)</b> : Progressive removals in descending order of node criticality across all topologies.		
		$\label{lem:ascending} \textbf{Ascending Targeted (Least} {\rightarrow} \textbf{Most)} \text{: Progressive removals in ascending order of node criticality across all topologies.}$		
	D	Random: Progressive removals restricted to small topologies in random order.		
Small Topologies (<17 nodes)		$\bf Descending\ Targeted:$ Progressive removals in descending order of node criticality within small topologies.		
		$\begin{tabular}{ll} \textbf{Ascending Targeted} : Progressive removals in ascending order of node criticality within small topologies. \end{tabular}$		
Big Topologies (≥17 nodes)		Random: Progressive removals restricted to big topologies in random order.		
		<b>Descending Targeted</b> : Progressive removals in descending order of node criticality within big topologies.		
	I	<b>Ascending Targeted</b> : Progressive removals in ascending order of node criticality within big topologies.		
Top 10% Critical Nodes (Single-Shot)		Random: One-time removal of the top 10% most critical nodes in random order.		
		Descending: One-time removal of the top 10% most critical nodes in descending order of importance.		
		${\bf Ascending:}$ One-time removal of the top 10% most critical nodes in ascending order of importance.		
Bottom 10% Least-Critical Nodes (Single-Shot)		<b>Random</b> : One-time removal of the bottom 10% least critical nodes in random order.		
		<b>Descending</b> : One-time removal of the bottom 10% least critical nodes in descending order of importance.		
	О	<b>Ascending</b> : One-time removal of the bottom 10% least critical nodes in ascending order of importance.		

values of  $\lambda_2$  and the Core Influence-Strength across all experiments, providing a direct test of whether one metric degrades more rapidly than the other. Aim 2 is addressed by structuring experiments to cover diverse disruption scenarios (random versus targeted removals, ascending versus descending orderings, and small versus large topologies) so that conditions of convergence or divergence between the models can be systematically identified. A full summary of Experiments A to O is provided in Table 1.

Progressive experiments (A–I) follow a cumulative design in which nodes are removed step by step, producing full attack or failure curves. To balance granularity with computational tractability, removals were staged at 5%, 10%, 15%, 30%, and 60% of each subset's nodes. This ladder provides finer resolution in the early regime (where robustness curves typically diverge most) while still extending to severe disruption at 60%. The cap at 60% avoids the saturation regime beyond which metrics often collapse to trivial values, and numerical procedures (e.g., matrix decompositions) may fail, particularly in smaller graphs. For very small topologies, not

all thresholds are achievable (e.g., 60% of 7 nodes exceeds the range of meaningful tests). In such cases, the progression is truncated to the largest valid cumulative removal count.

Random removals represent unpredictable disruptions such as incidental outages or hardware failures. Random experiments (A, D, G, J, M) are repeated 20 times per topology (or restricted subset for top/bottom decile tests). In each repetition, nodes are selected via uniform random sampling without replacement to construct a progressive removal sequence at the same staged thresholds (5%, 10%, 15%, 30%, 60%). This repetition captures variability across multiple randomisations and avoids dependence on a single random seed, yielding distributions of outcomes that better reflect the stochastic nature of random failures [13, 18].

Targeted removals (experiments B, C, E, F, H, I) simulate adversarial or cascading failures that spread through structurally significant nodes. In **descending targeted** experiments (B, E, H), nodes are

Table 2: Summary of Statistical Test Results for Experiments A-O

ID	Experiment	Test Applied	Mean Diff	95% CI	Outcome
A	Full Random	Wilcoxon (p = 0.0000)	-2.85	[-3.14, -2.57]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
В	Full Descending Targeted	Wilcoxon ( $p = 0.0000$ )	-3.29	[-4.22, -2.48]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
С	Full Ascending Targeted	Wilcoxon (p = 0.0000)	+3.69	[1.07, 7.89]	Reject $H_0$ ; $H_2$ favoured (CRM more sensitive)
D	Small Random	Wilcoxon ( $p = 0.0000$ )	-1.48	[-1.72, -1.26]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
E	Small Descending Targeted	Paired t-test ( $p = 0.0000$ )	-2.31	[-3.06, -1.61]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
F	Small Ascending Targeted	Wilcoxon (p = 0.0000)	+0.85	[0.38, 1.46]	Reject $H_0$ ; $H_2$ favoured (CRM more sensitive)
G	Big Random	Wilcoxon (p = 0.0000)	-4.18	[-4.64, -3.73]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
Н	Big Descending Targeted	Wilcoxon ( $p = 0.0000$ )	-4.04	[-5.43, -2.82]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
I	Big Ascending Targeted	Wilcoxon ( $p = 0.0023$ )	+5.42	[1.30, 11.62]	Reject $H_0$ ; $H_2$ favoured (CRM more sensitive)
J	Top 10% Random	Wilcoxon (p = 0.0000)	-1.51	[-1.67, -1.35]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
K	Top 10% Descending	Wilcoxon ( $p = 0.0039$ )	-1.61	[-2.17, -1.10]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
L	Top 10% Ascending	Wilcoxon ( $p = 0.0039$ )	-1.93	[-2.91, -1.07]	Reject $H_0$ ; $H_1$ favoured (SGM more sensitive)
M	Bottom 10% Random	Wilcoxon (p = 0.0000)	+0.15	[0.06, 0.24]	Reject $H_0$ ; $H_2$ favoured (CRM more sensitive)
N	Bottom 10% Descending	Wilcoxon ( $p = 0.4258$ )	-0.12	[-0.57, 0.15]	Fail to reject $H_0$ ; no significant difference
О	Bottom 10% Ascending	Paired t-test ( $p = 0.1598$ )	+0.37	[-0.04, 0.87]	Fail to reject $H_0$ ; no significant difference

removed from most to least critical; in **ascending targeted** experiments (C, F, I), the order is inverted, starting with peripheral nodes. Segmentation by topology size ensures that results are not biased toward either extreme. **Small** topologies are defined as having fewer than 17 nodes, while **big** topologies have 17 or more nodes. The value of 17 corresponds to the median size of the nine SANReN topologies, and was therefore chosen as a natural, datadriven threshold. This split creates two balanced groups. Using the median avoids researcher-imposed cut-offs and ensures that comparative analyses are not skewed toward only the smallest or largest networks in the dataset [13].

Experiments J–O isolate the impact of removing only the most or least critical nodes. The **Top 10**% group (experiments J–L) selects the most critical decile of nodes and removes them in either random order (experiment J), descending order (experiment K), or ascending order (experiment L). These represent highly disruptive, targeted failures limited to the network core. Conversely, the **Bottom 10**% group (experiments M–O) removes the least critical decile of nodes

under the same three orderings. These serve as a low-impact baseline to test whether the Spectral Graph Model and Core Resilience Model diverge even when failures affect only peripheral structures. Unlike the progressive experiments, these decile-based removals are evaluated as single-shot tests (with repetition for the random case), since no meaningful cumulative ladder exists within such restricted subsets.

Together, this framework captures a broad range of realistic and adversarial failure scenarios: unpredictable outages, progressive cascading failures, peripheral disturbances, and direct strikes at the network core. By combining repeated randomisations, staged progression, and targeted attacks, and by supplementing hypothesis testing with bootstrapped confidence intervals, the design provides a balanced and statistically robust basis for comparing the sensitivity of the Spectral Graph Model and Core Resilience Model.

#### 4 Results

This section presents the outcomes of all fifteen experiments (A-O), each designed to evaluate the sensitivity of the Spectral Graph Model and the Core Resilience Model under different node-removal scenarios. For each experiment, paired Area Under the Curve values were compared between models, and statistical tests were applied to determine whether observed differences were significant. Normality was assessed using the Shapiro-Wilk test, guiding the subsequent choice between a paired t-test and a Wilcoxon signed-rank test. A complete overview of the statistical outcomes, including test types, p-values, bootstrapped 95% confidence intervals, and favoured hypotheses, is provided in Table 2. Each row represents an experiment and is shaded as follows: unshaded rows identify experiments that found the Spectral Graph Model to be more sensitive, lightly shaded rows identify experiments that found neither to be more sensitive, and darkly shaded rows identify experiments that found the Core Resilience Model to be more sensitive. The results as output are included in Appendix D.

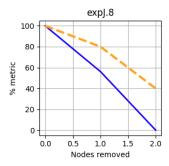


Figure 1: Degradation curve for Experiment J, Test 8.

The discussion below groups experiments by category (Full, Small, Big, Top 10%, and Bottom 10%), summarises general patterns within each, and illustrates the results with compact box-and-whisker plots. The left-hand, blue plot represents the Spectral Graph Model. The right-hand, orange plot represents the Core Resilience Model. On the left-hand side of each box-and-whisker plot, the vertical axis represents the Area Under the Curve values of the degradation curves. While the procedure for calculating Area Under the Curve values has already been described in detail, it is important to clarify its interpretation here. The Area Under the Curve is an aggregate measure of how quickly a network deteriorates under node removals: larger values indicate a more resilient network (slower decline), while smaller values indicate a more fragile network (faster decline). Although the values themselves are not absolute physical quantities, they provide a consistent and comparable index across experiments, making them a useful way to summarise and contrast resilience behaviours between different topologies and removal strategies. It is also worth noting that the bootstrapped 95% confidence intervals consistently captured the observed mean differences.

Figure 1 presents an example degradation curve generated from Test 8 of Experiment J. The solid blue line represents the Spectral Graph Model, while the dotted orange line represents the Core Resilience Model. Although such curves could be produced for every test in every experiment, this is not required for the analysis, as the Area Under the Curve values are already extracted and subjected to statistical testing. The example is included primarily to illustrate and contextualise what the Area Under the Curve values represent: a quicker drop-off indicates greater sensitivity, while a slower decline indicates greater resilience. This merely helps with understanding the results that follow.

Across all experiments (A–O), the observed mean differences between the Spectral Graph Model and the Core Resilience Model fell within the 95% bootstrap confidence intervals reported in Table 2. This consistency confirms that the confidence intervals provide accurate bounds for the effect sizes and that the statistical outcomes are robust to resampling variation.

4.0.1 Full Topologies (Experiments A–C). Under random removals (Experiment A) and descending targeted removals (Experiment B), the Spectral Graph Model is significantly more sensitive than the Core Resilience Model. When removals proceed from least to most critical (Experiment C), the Core Resilience Model is more sensitive. This indicates an order effect for deep cascading targeted failures: the Spectral Graph Model is more sensitive when descending, while the Core Resilience Model is more sensitive when ascending.

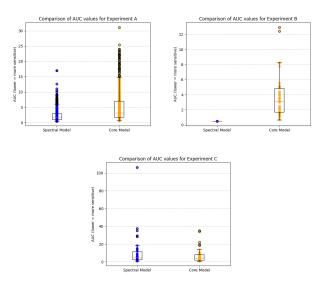


Figure 2: Full Topologies: (A) Random, (B) Descending Targeted, (C) Ascending Targeted.

4.0.2 Small Topologies (Experiments D–F). For small graphs (<17 nodes), the Spectral Graph Model is more sensitive under random (Experiment D) and descending targeted removals (Experiment E). With ascending targeted removals (Experiment F), the Core Resilience Model is more sensitive, showing that the ascending order effect persists in small networks.

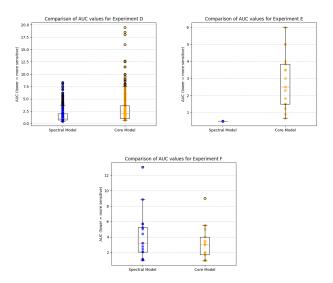


Figure 3: Small Topologies: (D) Random, (E) Descending Targeted, (F) Ascending Targeted.

4.0.3 Big Topologies (Experiments G–I). In big graphs (≥17 nodes), the Spectral Graph Model is more sensitive under random (Experiment G) and descending targeted removals (Experiment H). Under ascending targeted removals (Experiment I), the Core Resilience Model is more sensitive, mirroring the small-graph pattern and confirming a consistent order effect across sizes.

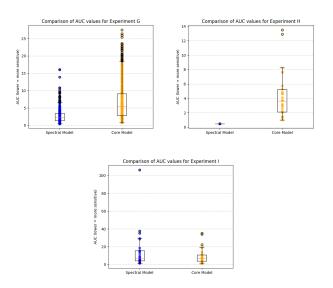


Figure 4: Big Topologies: (G) Random, (H) Descending Targeted, (I) Ascending Targeted.

4.0.4 Top 10% (Experiments J–L, single-shot). When removing the most-critical decile once the Spectral Graph Model is more sensitive across all orders: random (Experiment J), descending (Experiment

K), and ascending (Experiment L). Therefore, the Spectral Graph Model is more sensitive to direct strikes on core nodes. Unlike the previous results, the ascending order here does not result in the Core Resilience Model being more sensitive, suggesting that sensitivity depends not only on removal order but also on whether the sequence begins at the periphery or the core.

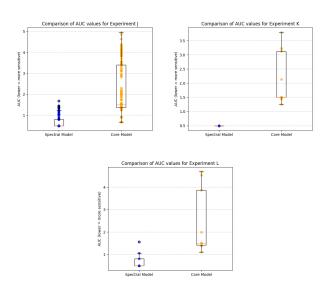


Figure 5: Top 10% (single-shot): (J) Random, (K) Descending, (L) Ascending.

4.0.5 Bottom 10% (Experiments M–O, single-shot). When only the least-critical decile of nodes is removed, results diverge. Under random removals (Experiment M), the Core Resilience Model is significantly more sensitive.

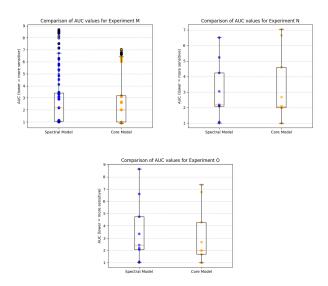


Figure 6: Bottom 10% (single-shot): (M) Random, (N) Descending, (O) Ascending.

However, under descending and ascending targeted removals (Experiments N and O), no statistically significant difference is observed between the models. When only the periphery is exposed to attacks (whether random or targeted), the observed differences in model sensitivity are less pronounced and do not follow the clearer patterns seen in core-directed failures (Experiments A - L).

### 5 Discussion

The suite of fifteen experiments (A–O) directly addresses the research question: How does the diagnostic performance of the Spectral Graph Model compare to that of the Core Resilience Model in identifying resilience weaknesses and structural vulnerabilities in SANReN's topology under simulated failure scenarios? By comparing algebraic connectivity ( $\lambda_2$ ) and Core Influence-Strength (across controlled disruption scenarios, the study tests whether the models differ in sensitivity and under what conditions they converge or diverge — the two aims set out in the introduction.

Across the majority of progressive experiments, the Spectral Graph Model emerged as more sensitive. In random removals (A, D, G), descending targeted removals (B, E, H), and all top-decile removals (J–L), the null hypothesis was consistently rejected in favour of  $H_1$ . These results conform to the theoretical expectation that  $\lambda_2$ , as a global connectivity metric, should respond sharply to disruptions that fragment or weaken the graph's connectedness. They confirm that  $\lambda_2$  declines more rapidly than Core Influence-Strength under random failures and direct attacks on critical nodes, satisfying Aim 1 by showing a statistically significant difference in sensitivity. In line with theory, the Spectral Graph Model therefore functions as a responsive indicator of fragmentation risk (i.e., the point at which the network separates into disconnected components and global connectivity is lost).

A contrasting trend appeared under some ascending targeted removals (C, F, I), where  $H_0$  was rejected in favour of  $H_2$ . In these cases, the Core Resilience Model was more sensitive, satisfying Aim 2 by identifying conditions under which the models diverge. This reversal is theoretically consistent: the Core Resilience Model captures redundancy and layered embedding, making it particularly attuned to failures that begin at the periphery and progressively erode support for the network core.

The decile experiments further clarify this dynamic. When only the top 10% of nodes are removed, the Spectral Graph Model is uniformly more sensitive, reflecting its focus on global connectivity and fragmentation risk. Experiment L provides an important nuance: although it is an ascending test, it still shows the Spectral Graph Model as more sensitive, suggesting that the Core Resilience Model only gains an advantage when failures begin sufficiently far out at the periphery. By contrast, the bottom-decile experiments produced mixed outcomes. Random removals from the least-critical 10% (M) favoured the Core Resilience Model, indicating that it can detect subtle weakening when peripheral redundancy is disrupted in an unstructured manner, while ordered removals of the same nodes (N and O) showed no significant differences between the

models. This divergence likely reflects both the distribution of failures across the periphery and the greater replication possible under random removals.

Taken together, the results suggest that the Core Resilience Model is more sensitive when removals *ascend deeply from the periphery*, but when only the periphery itself is targeted, there is no consistent evidence that either model offers a clearer diagnostic advantage.

The reliability of these results was strengthened by several validation measures. Each experiment contained multiple progressive tests (e.g., 5%, 10%, 15%, 30%, 60% removals), and random experiments were repeated twenty times per topology to capture variability across randomisations. This ensured that results were not artefacts of a single removal sequence. In addition, all mean Area Under the Curve differences were subjected to bootstrapped resampling with 10000 iterations, generating 95% confidence intervals. All observed means lay within their resampled bounds, confirming that reported effect sizes are stable and not dependent on a single sample. Together, replication and bootstrapping mitigate the natural limits of working with only nine topologies, although replication on a broader set of networks would further improve generality. All experiments were conducted on static topology snapshots. This approach provides clarity and control for isolating structural effects, but it does not capture dynamic processes such as adaptive routing, partial link degradation, or shifting traffic patterns that may interact with node failures in practice. As a result, the models' relative sensitivities may differ under real operational conditions, highlighting an avenue for future work using dynamic simulations.

The study's results demonstrate that the Spectral Graph Model and Core Resilience Model are not substitutes but complementary perspectives. The Spectral Graph Model excels as an early-warning metric, reacting quickly and predictably under a wide range of disruption scenarios. The Core Resilience Model, while less responsive overall, shows greater sensitivity when failures ascend deeply from the periphery and, in one case, under randomised peripheral losses. Together, their observed behaviours map directly onto their theoretical foundations: the Spectral Graph Model, via  $\lambda_2$ , provides insight into global connectivity collapse, while the Core Resilience Model, via Core Influence-Strength, highlights weaknesses in layered redundancy. These outcomes show that both aims of the study were achieved: Aim 1 was met by demonstrating significant differences in sensitivity, and Aim 2 was met by identifying the precise conditions of convergence and divergence. The results are therefore directly relevant to SANReN seeking actionable diagnostics and contribute to the literature on how spectral and core-based approaches complement one another in resilience assessment.

For SANReN, these findings provide clear guidance on resilience monitoring. The Spectral Graph Model's algebraic connectivity should be prioritised as a real-time indicator, since it reacts quickly to both random and targeted failures and can signal early fragmentation risks. The Core Resilience Model's Core Influence-Strength, while less responsive overall, adds value when used periodically to assess deeper structural redundancy, particularly under scenarios where disruptions originate in the periphery and erode support for

the core. Together, these metrics support a layered monitoring strategy:  $\lambda_2$  for fast detection of instability, and Core Influence-Strength for longer-term resilience planning. Implementing both perspectives in SANReN's diagnostic toolkit would improve early-warning capability while also informing investment decisions in reinforcing the network core and strengthening redundancy at the periphery.

### 6 Conclusion

This study set out to test whether the Spectral Graph Model and the Core Resilience Model differ in their diagnostic sensitivity, and to identify the conditions under which they converge or diverge. Both aims were achieved. The results show that the Spectral Graph Model, via algebraic connectivity ( $\lambda_2$ ), is significantly more sensitive under most random failures, descending targeted attacks, and direct removal of core nodes, confirming its value as an early-warning signal of fragmentation risk. Conversely, the Core Resilience Model, via Core Influence-Strength, is significantly more sensitive when failures progress from the periphery inward and, in one case, under random peripheral removals. This supports the theory that resilience is not one-dimensional but depends on both global connectivity and layered redundancy. Where only the bottom decile's nodes were removed in targeted orders, neither model dominated, clarifying the boundary conditions of their divergence.

Future work should extend the framework to dynamic and link-level failures, and test additional graph-theoretic indicators. These extensions would strengthen the reliability of resilience diagnostics and refine the practical recommendations for monitoring and improving research networks in resource-constrained environments.

The main outcome is clear: the two models are not alternatives but complementary. The Spectral Graph Model should be prioritised for real-time monitoring and proactive detection of instability, while the Core Resilience Model adds value in identifying weaknesses in structural embedding under bottom-up and peripheral failures. Together, they provide a richer and more actionable diagnostic toolkit for analysing the resilience of SANReN.

#### References

- Anirban Banerjee and Jürgen Jost. 2009. Spectral Characterization of Network Structures and Dynamics. *Dynamics On and Of Complex Networks* (2009), 117–132. https://doi.org/10.1007/978-0-8176-4751-3\_7
- [2] Sajitha Bashir. 2020. Connecting Africa's Universities to Affordable High-Speed Broadband Internet: What Will it Take? https://openknowledge.worldbank.org/entities/publication/c9dc3411-3b35-55b0-a6ba-da8b91d1fa90
- [3] Ivan Daniel Burke, Alan Herbert, and Roderick Mooi. 2018. Using network flow data to analyse distributed reflection denial of service (DRDoS) attacks, as observed on the South African national research and education network (SANReN): a postmortem analysis of the memcached attack on the SANReN. In Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (Port Elizabeth, South Africa) (SAICSIT '18). Association for Computing Machinery, New York, NY, USA, 164–170. https: //doi.org/10.1145/3278681.3278701
- [4] Egemen K. Çetinkaya, Mohammed J.F Alenazi, Justin P. Rohrer, and James P.G. Sterbenz. 2012. Topologyc connectivity analysis of internet infrastructure using graph spectra. 2012 IV International Congress on Ultra Modern Telecommunications and Control Systems (10 2012). https://doi.org/10.1109/icumt.2012.6459764
- [5] Egemen K. Çetinkaya, Mohammed J.F Alenazi, Yufei Cheng, Andrew M. Peck, and James P.G Sterbenz. 2014. A comparative analysis of geometric graph models

- for modelling backbone networks. Optical Switching and Networking 14 (06 2014), 95–106. https://doi.org/10.1016/j.osn.2014.05.001
- [6] Egemen K. Çetinkaya, Andrew M. Peck, and James P.G. Sterbenz. 2013. Flow robustness of multilevel networks. *Design of Reliable Communication Networks* (03 2013), 274–281.
- [7] Kevin Draai and Roderick Mooi. 2015. Implementing perfSONAR in the South African National Research and Education Network. In Proceedings of the Southern African Telecommunication Networks and Applications Conference (SATNAC). Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa. Available upon request or from SANReN documentation sources.
- [8] Bradley Efron and Robert J. Tibshirani. 1993. An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability, Vol. 57. Chapman and Hall/CRC, New York.
- [9] Michael Foley. 2016. The Role and Status of National Research and Education Networks (NRENs) in Africa. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/233231488314835003/the-role-and-status-of-national-research-and-education-networks-nrens-in-africa
- [10] Michael Foley. 2017. The Role and Status of National Research and Education Networks (NRENs) in Africa. SABER-ICT Technical Paper Series, No. 5. World Bank Group, Washington, D.C. http://documents.worldbank.org/curated/en/ 233231488314835003 World Bank Education, Technology & Innovation.
- [11] Linton C. Freeman. 1977. A set of measures of centrality based on betweenness. Sociometry 40, 1 (1977), 35–41.
- [12] Globus. 2025. Go beyond data Globus Compute. https://www.globus.org
- [13] Phillip Good. 2005. Permutation, Parametric and Bootstrap Tests of Hypotheses (3rd ed.). Springer, New York, NY.
- [14] B. Khunga and D. Kunda. 2017. Impact of NRENs in Universities The ZAMREN Experience. MANAS Journal of Engineering 5, 2 (2017), 13–23.
- [15] Ricky Laishram. 2020. The Resilience of k-Cores in Graphs. Ph.D. Dissertation. Syracuse University. https://surface.syr.edu/etd/1235 Dissertations - ALL, Paper 1235
- [16] Ricky Laishram, Ahmet Erdem Sariyüce, Tina Eliassi-Rad, Ali Pinar, and Sucheta Soundarajan. 2018. Measuring and Improving the Core Resilience of Networks. In Proceedings of the 2018 World Wide Web Conference (Lyon, France) (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 609–618. https://doi.org/10.1145/3178876.3186127
- [17] William Liu, Harsha Sirisena, Krzysztof Pawlikowski, and Allan McInnes. 2009. Utility of algebraic connectivity metric in topology design of survivable networks. UC Research Repository (University of Canterbury) 4 (10 2009), 131–138. https://doi.org/10.1109/drcn.2009.5340016
- [18] Mark Newman, 2010. Networks: An Introduction. Oxford University Press.
- [19] Izuchukwu Azuka Okafor, Smart Ikechukwu Mbagwu, Terkuma Chia, Zuwati Hasim, Echezona Ejike Udokanma, and Karthik Chandran. 2022. Institutionalizing Open Science in Africa: Limitations and Prospects. Frontiers in Research Metrics and Analytics 7 (04 2022). https://doi.org/10.3389/frma.2022.855198
- [20] Kasandra Pillay, Johann Hugo, Thuso Bogopa, Manqoba Shabalala, Thokozani Khwela, and Ajay Makan. 2024. SANReN's 100 Gbps Data Transfer Service: Transferring data fast! (11 2024), 765–769. https://doi.org/10.1109/scw63240. 2024.00109
- [21] Graeme D. Ruxton. 2006. The Unequal Variance t-test is an Underused Alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology* 17, 4 (2006), 688–690. https://doi.org/10.1093/beheco/ark016
- [22] L. Salie. 2021. An Analysis of Internet Traffic Flow in SANReN Using Active and Passive Measurements. Master's thesis. University of Cape Town, Faculty of Science, Department of Computer Science. http://hdl.handle.net/11427/36058 Accessed March 2025.
- [23] Samuel S. Shapiro and Martin B. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52, 3/4 (1965), 591–611. https://doi.org/10.2307/2333709
- [24] Tristan A. Shatto and Egemen K. Çetinkaya. 2017. Variations in graph energy: A measure for network resilience. 2017 9th International Workshop on Resilient Networks Design and Modeling (RNDM) (2017), 1–7. https://doi.org/10.1109/ RNDM.2017.8093019
- [25] James PG Sterbenz, David Hutchison, Egemen K Çetinkaya, Abdul Jabbar, Justin P Rohrer, Marcus Schöller, and Paul Smith. 2010. Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. Computer networks 54, 8 (2010), 1245–1265.
- [26] James PG Sterbenz, David Hutchison, Egemen K Çetinkaya, Abdul Jabbar, Justin P Rohrer, Marcus Schöller, and Paul Smith. 2010. Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. Computer Networks 54, 8 (2010), 1245–1265.
- [27] Student. 1908. The Probable Error of a Mean. Biometrika 6, 1 (1908), 1–25. https://doi.org/10.2307/2331554
- [28] Renier van Heerden and Ajay Makan. 2024. SA NREN Connectivity: SANREN + TENET. https://www.amlight.net/wp-content/uploads/2024/01/SA-NREN-RvHeerden-20240501.pdf Accessed: 2025-03-28.
- [29] Frank Wilcoxon. 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1, 6 (1945), 80–83. https://doi.org/10.2307/3001968

[30] Jian Wu, Mauricio Barahona, Yong-Jie Tan, and Hong-Zhong Deng. 2011. Spectral measure of structural robustness in complex networks. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 41, 6 (2011), 1244–1252.

# **Supplementary Material (Appendix)**

# A Nodes Ranked by Criticality per Topology

**BFN** 

2 6 8 3 5 1 9 10 11 12 13 4 7 14 15 16 17

### **CPT**

3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 54 19 5 56 4 58 64 65 66 67 68 69 2 59 60 61 50 40 63 16 24 18 51 37 23 55 52 28 29 42 13 11 15 9 7 25 32 35 6 17 21 41 47 20 36 38 27 39

### **DUR**

1 15 8 4 13 19 11 22 21 23 7 16 5 6 12 2 24 25 26 27 17 20 9 10 3 14 18

### **ELS**

1 3 9 6 10 7 8 2 5 11 12 4 14 13

#### INB

 $1\ 3\ 10\ 16\ 24\ 31\ 21\ 27\ 23\ 11\ 20\ 25\ 28\ 22\ 8\ 26\ 2\ 5\ 32\ 14\ 15\ 19\ 30\ 4\ 6\ 7\ 9\ 12\ 13\ 29\ 18\ 17$ 

# POL

3152467

#### **PTA**

5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 45 36 30 26 57 33 34 35 10 42 46 18 6 7 59 29 2 28 39 31 50 52 38 3 20 48 55 37 54 17 24 16 53 27 15 14 12 13 49 47 4 25

# **PZB**

1658423910117

# VDP

15642378

# **B** Core Influence-Strength Pseudocode

The following pseudocode illustrates how Core Influence-Strength (CIS) is calculated in this study, based on Laishram's definitions.

### Algorithm 1 Core Influence-Strength (CIS)

```
Require: Graph G = (V, E), percentile f where 0 < f \le 1
  1: \kappa \leftarrow k-core numbers for all nodes (using NetworkX)
                                                                                                                                                              ▶ Core Strength
  2: for each node u \in V do
          CS[u] \leftarrow |\{v \in \Gamma(u) : \kappa(v) \ge \kappa(u)\}| - \kappa(u) + 1
  4: end for
                                                                                                                                                              ▶ Core Influence
  5: Build sparse matrix M
    for each edge (u, v) \in E do
          if \kappa(u) \leq \kappa(v) then
  7:
              denom \leftarrow |\{w \in \Gamma(v) : \kappa(w) \ge \kappa(v)\}|
  8:
               M[u,v] \leftarrow 1/denom
  9:
          end if
 10:
          if \kappa(v) \leq \kappa(u) then
 11:
               denom \leftarrow |\{w \in \Gamma(u) : \kappa(w) \ge \kappa(u)\}|
 12:
               M[v,u] \leftarrow 1/denom
 13:
          end if
 14:
 15: end for
 16: Add identity entries to diagonal of M
 17: r \leftarrow leading eigenvector of M
    for each u \in V do
          CI[u] \leftarrow r[u]
20: end for
                                                                                                                                                           ▶ CIS Aggregation
21: \tau \leftarrow f-th percentile of \{CI[u] : u \in V\}
22: S_f \leftarrow \{u \in V : CI[u] \ge \tau\}
23: CIS_f(G) \leftarrow \frac{1}{|S_f|} \sum_{u \in S_f} CS[u]
Ensure: CIS_f(G)
```

# **C** Experiment Definitions

### **Experiment A**

bfn 16 bfn 16 8 bfn 16 8 1 bfn 16 8 1 10 3 bfn 16 8 1 10 3 bfn 16 8 1 10 3 4 6 2 12 11 bfn 2 bfn 2 10 bfn 2 10 5 bfn 2 10 5 12 14 bfn 2 10 5 12 14 4 15 17 9 13 bfn 17 bfn 17 9 bfn 17 9 6 bfn 17 9 6 3 11 bfn 17 9 6 3 11 bfn 17 9 6 3 11 14 7 12 2 1 bfn 8 bfn 8 9 bfn 8 9 4 bfn 8 9 4 10 1 bfn 8 9 4 10 1 2 3 14 13 5 bfn 11 bfn 11 8 bfn 11 8 1 bfn 11 8 1 3 13 bfn 11 8 1 3 13 10 12 4 7 6 bfn 11 bfn 11 9 bfn 11 9 4 bfn 11 9 4 12 6 bfn 11 9 4 12 6 14 16 10 15 1 bfn 2 bfn 2 12 bfn 2 12 15 bfn 2 12 15 5 7 bfn 2 12 15 5 7 4 8 1 6 9 bfn 16 bfn 16 6 bfn 16 6 3 bfn 16 6 3 12 10 bfn 16 6 3 12 10 1 13 7 2 14 bfn 5 bfn 5 14 bfn 5 14 16 bfn 5 14 16 10 12 bfn 5 14 16 10 12 8 3 9 1 15 bfn 15 bfn 15 10 bfn 15 10 bfn 15 10 6 bfn 15 10 6 12 2 bfn 15 10 6 12 2 17 13 7 16 8 bfn 16 bfn 16 9 bfn 16 9 14 bfn 16 9 14 15 17 bfn 16 9 14 15 17 3 2 13 8 1 bfn 1 bfn 1 17 bfn 1 17 3 bfn 1 17 3 bfn 1 17 3 16 2 bfn 1 17 3 16 2 15 14 4 13 11 bfn 17 bfn 17 16 bfn 17 16 14 bfn 17 16 14 13 15 bfn 17 16 14 13 15 9 11 12 8 6 bfn 14 bfn 14 11 bfn 14 11 2 bfn 14 11 2 6 15 bfn 14 11 2 6 15 3 8 1 4 17 bfn 8 bfn 8 11 bfn 8 11 14 bfn 8 11 14 13 9 bfn 8 11 14 13 9 16 7 4 10 15 bfn 6 bfn 6 3 bfn 6 3 15 bfn 6 3 15 4 14 bfn 6 3 15 4 14 16 10 17 1 9 bfn 4 9 bfn 4 9 bfn 4 9 13 bfn 4 9 13 14 10 bfn 4 9 13 14 10 15 11 2 16 17 bfn 7 bfn 7 2 fn 7 2 14 bfn 7 2 14 13 11 bfn 7 2 14 13 11 ofn 7 2 14 10 1 bfn 8 7 fn 8 7 14 bfn 8 7 14 17 12 bfn 8 7 14 17 12 6 4 2 5 10 bfn 4 bfn 4 10 bfn 4 10 3 bfn 4 10 3 6 5 bfn 4 10 3 6 5 bfn 4 10 3 6 5 14 8 13 2 16 cpt 44 26 60 cpt 44 26 60 24 21 3 65 cpt 44 26 60 24 21 3 65 59 14 13 cpt 44 26 60 24 21 3 65 59 14 13 46 49 12 17 68 32 9 42 52 48 40 cpt 44 26 60 24 21 3 65 59 14 13 46 49 12 17 68 32 9 42 52 48 40 36 33 6 61 20 53 63 58 15 25 18 43 37 34 57 1 38 56 22 28 cpt 3 27 42 cpt 3 27 42 47 10 21 52 cpt 3 27 42 47 10 21 52 55 60 4 cpt 3 27 42 47 10 21 52 55 50 24 63 cpt 23 10 22 cpt 23 10 22 19 47 51 65 35 64 18 cpt 23 10 22 19 47 51 65 35 64 18 1 2 67 34 3 15 25 41 68 8 44 cpt 23 10 22 19 47 51 65 35 64 18 1 2 67 34 3 15 25 41 68 8 44 37 56 26 29 54 36 5 24 49 63 21 59 52 17 6 31 58 28 42 27 cpt 17 39 28 cpt 17 39 28 27 16 56 63 cpt 17 39 28 27 16 56 63 69 66 61 cpt 17 39 28 27 16 56 63 69 66 61 67 50 26 55 40 59 10 11 31 52 24 cpt 17 39 28 27 16 56 63 69 66 61 67 50 26 55 40 59 10 11 31 52 24 65 13 19 6 1 60 5 21 8 33 57 32 47 9 42 68 53 51 22 35 cpt 65 25 40 cpt 65 25 40 29 42 35 30 cpt 65 25 40 29 42 35 30 cpt 66 22 cpt 65 25 40 29 42 35 30 62 66 22 64 24 9 26 63 67 33 7 44 53 48 cpt 65 25 40 29 42 35 30 62 66 22 64 24 9 26 63 67 33 7 44 53 48 3 37 41 4 55 47 1 15 49 58 59 43 19 50 20 27 32 56 6 68 cpt 27 64 33 cpt 27 64 33 7 24 19 44 cpt 27 64 33 7 24 19 44 3 13 41 cpt 27 64 33 cpt 27 64 33 cpt 27 64 33 7 24 19 44 cpt 27 64 33 7 24 19 44 3 13 41 cpt 27 64 33 7 24 19 44 3 13 41 cpt 27 64 33 cpt 27 64 33 cpt 27 64 33 cpt 27 64 33 7 24 19 44 cpt 27 64 33 7 24 19 44 3 13 41 cpt 27 64 33 cpt 27 20 32 42 15 67 2 35 31 58 cpt 27 64 33 7 24 19 44 3 13 41 4 54 20 32 42 15 67 2 35 31 58 29 57 68 16 26 37 23 69 52 53 56 51 22 8 30 38 39 17 14 11 cpt 36 1 68 cpt 36 1 68 3 42 26 45 cpt 36 1 68 3 42 26 45 64 31 37 cpt 36 1 68 3 42 26 45 64 31 37 9 18 11 60 65 46 29 14 55 22 44 cpt 36 1 68 3 42 26 45 64 31 37 9 18 11 60 65 46 29 14 55 22 44 7 47 38 24 43 25 59 6 33 23 40 61 52 51 56 19 62 12 21 27 cpt 10 51 64 cpt 10 51 64 36 44 24 46 cpt 10 51 64 36 44 24 46 21 39 9 cpt 10 51 64 36 44 24 46 21 39 9 6 59 2 4 55 11 7 29 65 34 23 cpt 10 51 64 36 44 24 46 21 39 9 6 59 2 4 55 11 7 29 65 34

### **Experiment B**

cpt 3 1 44 8 cpt 3 1 44 8 14 22 46 cpt 3 1 44 8 14 22 46 33 49 10 43 cpt 3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 cpt 3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 54 19 5 56 4 58 64 65 66 67 68 69 2 59 60 61 50 40 63 16 24 bfn 2 bfn 2 6 8 bfn 2 6 8 3 5 1 bfn 2 6 8 3 5 1 9 10 11 12 13 dur 1 15 dur 1 15 8 dur 1 15 8 dur 1 15 8 4 13 19 11 22 21 dur 1 15 8 4 13 19 11 22 21 23 7 16 5 6 12 2 24 els 1 els

1 3 els 1 3 9 els 1 3 9 6 10 els 1 3 9 6 10 7 8 2 5 jnb 1 3 jnb 1 3 10 16 jnb 1 3 10 16 24 jnb 1 3 10 16 24 31 21 27 23 11 jnb 1 3 10 16 24 31 21 27 23 11 20 25 28 22 8 26 2 5 32 14 pol 3 pol 3 1 pol 3 1 5 pta 5 1 51 23 pta 5 1 51 23 43 22 56 pta 5 1 51 23 43 22 56 21 44 11 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 45 36 30 26 57 33 34 35 10 42 46 18 6 7 59 29 2 28 pzb 1 pzb 1 6 pzb 1 6 5 8 pzb 1 6 5 8 4 2 3 vdp 1 vdp 1 5 vdp 1 5 6 vdp 1 5 6 4

# **Experiment C**

cpt 39 27 38 36 cpt 39 27 38 36 20 47 41 cpt 39 27 38 36 20 47 41 21 17 6 35 cpt 39 27 38 36 20 47 41 21 17 6 35 32 25 7 9 15 11 13 42 29 28 cpt 39 27 38 36 20 47 41 21 17 6 35 32 25 7 9 15 11 13 42 29 28 cpt 39 27 38 36 20 47 41 21 17 6 35 32 25 7 9 15 11 13 42 29 28 52 55 23 37 51 18 24 16 63 40 50 61 60 59 2 69 68 67 66 65 64 bfn 17 bfn 17 16 bfn 17 16 15 bfn 17 16 15 14 7 4 bfn 17 16 15 14 7 4 13 12 11 10 9 dur 18 14 dur 18 14 3 dur 18 14 3 10 9 dur 18 14 3 10 9 20 17 27 26 dur 18 14 3 10 9 20 17 27 26 25 24 2 12 6 5 16 7 els 13 els 13 14 els 13 14 4 els 13 14 4 12 11 els 13 14 4 12 11 5 2 8 7 jnb 17 18 jnb 17 18 29 13 jnb 17 18 29 13 12 9 7 6 4 30 jnb 17 18 29 13 12 9 7 6 4 30 19 15 14 32 5 2 26 8 22 28 pol 7 pol 7 6 pol 7 6 4 pta 25 4 47 49 pta 25 4 47 49 13 12 14 pta 25 4 47 49 13 12 14 15 27 53 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 vdp 8 vdp 8 7 vdp 8 7 3 vdp 8 7 3 2

# **Experiment D**

els 6 els 6 els 6 3 14 9 els 6 3 14 9 els 6 3 14 9 2 10 7 5 els 11 els 11 10 els 11 10 els 11 10 14 4 els 11 10 14 4 3 6 1 8 els 5 els 5 els 5 8 4 6 els 5 8 4 6 els 5 8 4 6 10 3 13 2 els 12 els 12 els 12 8 els 12 8 els 12 8 10 11 els 12 8 10 11 4 2 14 1 els 12 els 12 els 12 14 7 1 els 12 14 7 1 els 12 14 7 1 9 13 5 10 els 11 els 11 els 11 1 els 11 1 6 5 els 11 1 6 5 4 14 3 8 els 4 els 4 els 4 10 els 4 10 3 11 els 4 10 3 11 2 9 6 12 els 3 els 3 els 3 11 els 3 11 13 14 els 3 11 13 14 8 4 10 6 els 10 els 10 els 10 4 els 10 4 els 10 4 8 2 els 10 4 8 2 9 12 1 6 els 7 els 7 els 7 els 7 6 13 10 els 7 6 13 10 14 4 12 5 els 8 els 8 els 8 6 els 8 6 5 7 els 8 6 5 7 14 2 10 3 els 6 els 6 els 6 8 els 6 8 1 12 els 6 8 1 12 13 2 4 5 els 4 els 4 els 4 9 els 4 9 1 13 els 4 9 1 13 14 12 7 11 els 10 els 10 els 10 7 els 10 7 13 4 els 10 7 13 4 9 11 3 8 els 13 els 13 els 13 12 els 13 12 els 13 12 10 7 els 13 12 10 7 9 2 3 14 els 11 els 11 els 11 4 els 11 4 1 8 els 11 4 1 8 3 14 5 12 els 9 els 9 els 9 els 9 2 els 9 2 4 3 els 9 2 4 3 7 13 12 5 els 3 els 3 els 3 11 els 3 11 6 9 els 3 11 6 9 els 3 11 6 9 14 5 1 13 els 13 els 13 els 13 11 els 13 11 10 7 els 13 11 10 7 12 9 2 1 els 9 els 9 els 9 3 els 9 3 els 9 3 13 12 els 9 3 13 12 2 11 10 1 pol 1 pol 1 pol 1 pol 1 4 pol 1 4 6 3 pol 7 pol 7 pol 7 4 pol 7 4 2 1 pol 3 pol 3 pol 3 pol 3 5 pol 3 5 7 4 pol 4 pol 4 pol 4 pol 4 7 1 5 pol 6 pol 6 pol 6 pol 6 4 pol 6 4 5 7 pol 6 pol 6 pol 6 pol 6 1 pol 6 1 5 3 pol 2 pol 2 pol 2 pol 2 3 pol 2 3 4 6 pol 1 pol 1 pol 1 pol 1 2 pol 1 2 7 6 pol 1 pol 1 pol 1 pol 1 pol 1 7 pol 1 7 5 4 pol 6 pol 6 pol 6 pol 6 4 pol 6 4 7 3 pol 3 pol 3 pol 3 pol 3 7 pol 3 7 6 5 pol 3 pol 3 pol 3 pol 3 6 pol 3 6 pol 3 6 1 4 pol 4 pol 4 pol 4 pol 4 1 pol 4 1 5 7 pol 3 pol 3 pol 3 pol 3 5 pol 3 5 4 1 pol 4 pol 4 pol 4 pol 4 7 pol 4 7 5 3 pol 3 pol 3 pol 3 pol 3 pol 3 2 pol 3 2 1 4 pol 6 pol 6 pol 6 pol 6 pol 6 2 pol 6 2 7 3 pol 7 pol 7 pol 7 pol 7 1 4 2 pol 6 pol 6 pol 6 pol 6 5 3 4 pol 6 pol 6 pol 6 pol 6 4 pol 6 4 5 7 pzb 9 pzb 9 pzb 9 1 pzb 9 1 5 pzb 9 1 5 pzb 9 1 5 3 8 10 2 pzb 4 pzb 4 pzb 4 3 pzb 4 3 8 pzb 4 3 8 5 6 2 7 pzb 11 pzb 11 pzb 12 pzb 4 pzb 4 pzb 4 pzb 4 3 8 pzb 4 3 8 pzb 4 3 8 5 6 2 7 pzb 11 pzb 12 pzb 13 pzb 14 pzb 14 pzb 14 pzb 14 pzb 15 pzb 14 pzb 14 pzb 15 pzb 15 pzb 15 pzb 15 pzb 15 pzb 15 pzb 16 pzb 17 pzb 18 p 11 1 pzb 11 1 5 pzb 11 1 5 10 4 7 3 pzb 11 pzb 11 pzb 11 pzb 11 5 pzb 11 5 6 pzb 11 5 6 pzb 1 5 6 1 8 3 4 pzb 1 pzb 1 8 pzb 1 8 pzb 1 8 11 pzb 1 8 11 9 10 4 5 pzb 1 pzb 1 pzb 1 7 pzb 1 7 6 pzb 1 7 6 5 10 9 4 pzb 1 pzb 1 pzb 1 pzb 1 8 pzb 1 8 2 pzb 1 8 2 9 5 9 6 pzb 1 pzb 1 pzb 1 pzb 1 10 6 pzb 1 10 6 8 2 4 7 pzb 10 pzb 10 pzb 10 6 pzb 10 6 2 pzb 10 6 2 pzb 10 6 2 9 1 5 8 pzb 11 pzb 11 pzb 11 pzb 11 9 pzb 11 9 7 pzb 11 9 7 1 4 6 5 pzb 1 pzb 1 pzb 1 6 pzb 1 6 4 pzb 1 6 4 7 5 10 8 pzb 8 pzb 8 pzb 8 5 pzb 8 5 11 pzb 8 5 11 9 7 6 4 pzb 2 pzb 2 pzb 2 4 pzb 2 4 8 pzb 2 4 8 11 3 1 5 pzb 4 pzb 4 11 pzb 4 11 10 pzb 4 11 10 pzb 4 11 10 3 2 97 pzb 6 pzb 6 pzb 6 4 pzb 6 4 7 pzb 6 4 7 pzb 6 4 7 8 2 1 9 pzb 4 pzb 4 pzb 4 11 pzb 4 11 2 pzb 4 11 2 pzb 4 12 1 5 9 7 pzb 1 pzb 1 pzb 1 3 pzb 1 3 9 pzb 1 3 9 2 8 4 7 pzb 9 pzb 9 pzb 9 7 pzb 9 7 5 pzb 9 7 5 pzb 9 7 5 1 11 8 2 pzb 2 pzb 2 pzb 2 pzb 2 9 pzb 2 9 7 pzb 2 9 7 1 10 11 6 pzb 7 pzb 7 pzb 7 3 pzb 7 3 4 pzb 7 3 4 2 9 1 11 vdp 8 vdp 8 vdp 8 vdp 8 5 vdp 8 5 vdp 8 5 7 3 4 vdp 5 vdp 5 vdp 5 vdp 5 1 vdp 5 1 vdp 5 1 6 4 7 vdp 8 vdp 8 vdp 8 vdp 8 3 vdp 8 3 5 1 6 vdp 4 vdp 4 vdp 4 vdp 4 3 vdp 4 3 5 6 8 vdp 6 vdp 6 vdp 6 vdp 6 4 vdp 6 4 8 1 2 vdp 5 vdp 5 vdp 5 vdp 5 4 vdp 5 4 odp 7 vdp 7 vdp 7 vdp 7 vdp 7 vdp 7 6 vdp 7 6 1 5 3 vdp 8 vdp 8 vdp 8 vdp 8 7 vdp 8 7 2 4 1 vdp 3 vdp 3 vdp 3 vdp 3 vdp 3 6 vdp 3 6 8 1 4 vdp 1 vdp 1 vdp 1 vdp 1 4 vdp 1 4 7 2 8 vdp 4 vdp 4 vdp 4 vdp 4 1 vdp 4 1 3 8 6 vdp 1 vdp 1 vdp 1 vdp 1 6 vdp 1 6 7 4 3 vdp 8 vdp 8 vdp 8 vdp 8 6 vdp 8 6 2 4 5 vdp 1 vdp 1 vdp 1 vdp 1 7 vdp 1 7 4 6 2 vdp 5 vdp 5 vdp 5 vdp 5 4 vdp 5 4 2 6 8 vdp 3 vdp 3 vdp 3 vdp 3 7 vdp 3 7 2 8 5 vdp 1 vdp 1 vdp 1 vdp 1 3 vdp 1 3 7 5 4 vdp 2 vdp 2 vdp 2 vdp 2 3 vdp 2 3 6 5 1 vdp 7 vdp 7 vdp 7 vdp 7 1 vdp 7 1 8 3 6 vdp 1 vdp 1 vdp 1 vdp 1 3 vdp 1 3 4 6 2

# **Experiment E**

els 1 els 1 3 els 1 3 9 els 1 3 9 6 10 els 1 3 9 6 10 7 8 2 5 pol 3 pol 3 1 pol 3 1 5 pol 3 1 5 2 pzb 1 pzb 1 6 pzb 1 6 5 8 pzb 1 6 5 8 4 2 3 vdp 1 vdp 1 5 vdp 1 5 6 vdp 1 5 6 4

# **Experiment F**

els 13 els 13 14 els 13 14 4 els 13 14 4 12 11 els 13 14 4 12 11 5 2 8 7 pol 7 pol 7 6 pol 7 6 4 pol 7 6 4 2 pzb 7 pzb 7 11 pzb 7 11 10 9 pzb 7 11 10 9 3 2 4 vdp 8 vdp 8 7 vdp 8 7 3 vdp

# Experiment G

bfn 16 bfn 16 10 bfn 16 10 17 bfn 16 10 17 3 15 12 7 9 13 1 bfn 9 bfn 9 4 bfn 9 4 6 bfn 9 4 6 12 5 bfn 9 4 6 12 5 7 2 17 13 14 bfn 16 bfn 16 14 5 bfn 16 14 5 9 15 bfn 16 14 5 9 15 8 10 13 3 4 bfn 13 8 bfn 13 8 2 bfn 13 8 2 4 16 bfn 13 8 2 4 16 14 3 6 12 1 bfn 6 bfn 6 13 bfn 6 13 3 15 4 bfn 6 13 3 15 4 7 10 17 2 5 bfn 10 bfn 10 4 bfn 10 4 8 bfn 10 4 8 6 3 bfn 10 4 8 6 3 15 9 13 14 7 bfn 1 bfn 1 6 bfn 1 6 2 bfn 1 6 2 5 4 bfn 1 6 2 5 4 15 3 10 8 7 bfn 2 bfn 2 3 15 bfn 2 3 13 7 6 bfn 2 3 13 7 6 16 8 4 12 1 bfn 9 bfn 9 8 bfn 9 8 4 6 2 bfn 9 8 4 6 2 bfn 9 8 4 6 2 15 5 3 11 7 bfn 13 bfn 13 3 bfn 13 3 2 bfn 13 3 2 10 8 bfn 13 3 2 10 8 7 17 1 16 15 bfn 2 8 bfn 2 8 16 bfn 2 8 16 15 7 bfn 2 8 16 15 7 5 9 12 10 11 bfn 15 bfn 15 17 bfn 15 17 13 bfn 15 17 13 8 16 bfn 15 17 13 8 16 5 9 6 2 4 bfn 4 8 7 bfn 4 8 7 2 17 bfn 4 8 7 2 17 6 14 10 12

15 10 jnb 30 25 jnb 30 25 29 jnb 30 25 29 6 14 jnb 30 25 29 6 14 jnb 30 25 29 6 14 8 28 3 13 12 jnb 30 25 29 6 14 8 28 3 13 12 27 1 32 24 4 11 26 7 15 jnb 31 19 jnb 31 19 24 jnb 31 19 24 17 22 jnb 31 19 24 17 22 29 14 7 6 26 jnb 31 19 24 17 22 29 14 7 6 26 2 12 3 10 23 16 4 15 25 jnb 18 29 jnb 18 29 10 jnb 18 29 10 11 32 jnb 18 29 10 11 32 5 17 22 13 24 jnb 18 29 10 11 32 5 17 22 13 24 jnb 18 29 10 11 32 5 17 22 13 24 30 27 9 15 7 12 25 19 6 jnb 23 26 jnb 23 26 22 jnb 23 26 22 1 19 jnb 23 26 22 1 19 28 5 27 2 21 jnb 23 26 22 1 19 28 5 27 2 21 31 20 15 8 12 14 6 9 17 jnb 30 4 jnb 30 4 20 jnb 30 4 20 10 26 jnb 30 4 20 10 26 32 18 19 1 22 inb 30 4 20 10 26 32 18 19 1 22 12 14 23 11 15 13 7 16 5 inb 20 27 inb 20 27 15 inb 20 27 15 29 14 inb 20 27 15 29 14 4 19 12 9 26 inb 20 27 15 29 14 4 19 12 9 26 23 3 2 7 32 21 16 5 8 jnb 14 16 jnb 14 16 21 jnb 14 16 21 3 13 jnb 14 16 21 3 13 9 2 19 7 1 jnb 14 16 21 3 13 9 2 19 7 1 29 31 28 22 11 24 30 18 23 jnb 3 21 jnb 3 21 20 jnb 3 21 20 10 8 jnb 3 21 20 10 8 jnb 3 21 20 10 8 7 17 16 5 31 jnb 3 21 20 10 8 7 17 16 5 31 23 19 28 14 29 13 24 2 27 jnb 1 2 jnb 1 2 18 jnb 1 2 18 17 14 jnb 1 2 18 17 14 13 6 32 8 25 jnb 1 2 18 17 14 13 6 32 8 25 3 nb 1 2 18 17 14 13 6 32 8 25 30 26 28 24 12 21 5 19 27 jnb 20 29 jnb 20 29 24 jnb 20 29 24 21 32 jnb 20 29 24 21 32 23 10 31 22 30 jnb 20 29 24 21 32 23 10 31 22 30 27 18 6 19 14 8 5 15 2 jnb 20 17 jnb 20 17 18 jnb 20 17 18 14 4 jnb 20 17 18 14 4 28 12 25 3 32 jnb 20 17 18 14 4 28 12 25 3 32 19 21 7 1 6 8 5 10 24 jnb 6 11 jnb 6 11 7 jnb 6 11 7 25 17 jnb 6 11 7 25 17 3 1 12 5 28 jnb 6 11 7 25 17 3 1 12 5 28 15 22 14 16 8 20 9 29 27 jnb 13 25 jnb 13 25 23 jnb 13 25 23 jnb 13 25 23 7 3 jnb 13 25 23 7 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 5 30 27 31 8 jnb 13 25 23 37 3 3 jnb 13 25 23 37 3 3 3 3 20 27 31 8 jnb 13 25 23 37 3 3 3 3 3 20 27 31 8 jnb 13 25 23 37 3 3 3 3 3 3 20 27 31 8 jnb 13 25 23 37 3 3 3 3 3 3 20 27 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 28 1 16 4 6 17 20 jnb 1 14 jnb 1 14 23 jnb 1 14 23 19 6 jnb 1 14 23 19 6 18 10 9 5 11 jnb 1 14 23 19 6 18 10 9 5 11 8 16 3 24 4 7 30 27 21 pta 46 33 60 pta 46 33 60 23 36 41 pta 46 33 60 23 36 41 12 38 19 pta 46 33 60 23 36 41 12 38 19 pta 46 33 60 23 36 41 12 38 19 pta 46 33 60 23 36 41 12 38 19 pta 46 33 60 23 36 41 12 38 19 25 1 10 40 34 46 2 4 39 26 5 32 52 7 61 41 45 pta 22 57 48 10 40 34 46 2 4 39 26 5 32 52 7 61 41 45 51 8 9 18 49 36 47 42 25 60 53 11 35 14 15 38 21 19 27 pta 58 42 59 pta 58 42 59 60 40 10 pta 58 42 59 60 40 10 44 53 41 pta 58 42 59 60 40 10 44 53 41 13 7 11 35 46 39 54 32 25 pta 58 42 59 60 40 10 44 53 41 13 7 11 35 46 39 54 32 25 55 33 43 23 56 26 17 21 14 45 2 29 36 1 37 30 8 47 49 pta 36 43 12 pta 36 43 12 61 3 1 pta 42 pta 36 43 12 61 3 1 27 33 42 30 51 4 46 59 17 22 6 55 pta 36 43 12 61 3 1 27 33 42 30 51 4 46 59 17 22 6 55 25 19 49 31 8 2 39 53 28 56 48 7 34 18 35 11 23 32 10 pta 12 20 33 pta 12 20 33 35 49 16 pta 12 20 33 35 49 16 pta 12 20 33 35 49 16 18 60 37 pta 12 20 33 35 49 16 18 60 37 58 14 44 38 50 59 7 17 11 pta 12 20 33 35 49 16 18 60 37 58 14 44 38 50 59 7 17 11 34 41 4 57 39 30 22 1 48 19 42 13 6 23 32 9 61 40 45 pta 43 5 21 pta 43 5 21 56 55 25 pta 43 5 21 56 55 25 61 26 15 pta 43 5 21 56 55 25 61 26 15 7 40 37 41 48 27 38 57 20 pta 43 5 21 56 55 25 61 26 15 7 40 37 41 48 27 38 57 20 pta 43 5 21 56 55 25 61 26 15 7 40 37 41 48 27 38 57 20 42 8 60 23 36 19 3 18 59 17 46 9 13 54 50 58 14 22 16 pta 21 60 46 pta 21 60 46 54 5 51 pta 21 60 46 54 5 51 8 3 18 pta 21 60 46 54 5 51 8 3 18 61 24 22 55 39 13 59 50 23 pta 21 60 46 54 5 51 8 3 18 61 24 22 55 39 13 59 50 23 36 53 41 48 28 57 6 47 10 15 49 25 43 27 35 17 4 19 1 pta 49 34 22 pta 49 34 22 3 31 18 pta 49 34 22 3 31 18 16 61 57 pta 49 34 22 3 31 18 16 61 57 59 14 40 27 38 6 29 42 35 pta 49 34 22 3 31 18 16 61 57 59 14 40 27 38 6 29 42 35 4 30 10 2 46 28 32 44 13 53 58 7 33 23 56 5 41 50 15 pta 56 29 49 pta 56 29 49 23 20 53 pta 56 29 49 23 20 53 30 54 41 pta 56 29 49 23 20 53 30 54 41 6 11 42 39 15 40 46 24 25 pta 56 29 49 23 20 53 30 54 41 6 11 42 39 15 40 46 24 25 13 7 58 3 44 59 12 33 35 4 19 1 47 16 34 60 21 5 48 pta 10 6 5 pta 10 6 5 44 1 2 pta 10 6 5 44 1 2 28 60 35 pta 10 6 5 44 1 2 28 60 35 pta 10 6 5 44 1 2 28 60 35 15 11 54 41 40 42 34 30 8 pta 10 6 5 44 1 2 28 60 35 15 11 54 41 40 42 34 30 8 pta 10 6 5 44 1 2 28 60 35 15 11 54 41 40 42 34 30 8 27 3 38 48 51 21 23 55 53 13 47 39 17 24 32 20 59 4 58 pta 48 55 49 pta 48 55 49 51 45 40 pta 48 55 49 51 45 40 44 47 46 pta 48 55 49 51 45 40 44 7 46 25 9 34 60 27 17 31 58 10 pta 48 55 49 51 45 40 44 47 46 25 9 34 60 27 17 31 58 10 43 4 3 28 42 52 53 19 8 32 7 38 54 16 14 59 15 18 2 pta 32 58 12 pta 32 58 12 21 42 10 pta 32 58 12 21 42 10 25 53 2 pta 32 58 12 21 42 10 25 53 2 pta 32 58 12 21 42 10 25 53 2 20 35 47 30 28 44 59 11 57 pta 32 58 12 21 42 10 25 53 2 20 35 47 30 28 44 59 11 57 49 43 51 7 14 24 50 41 38 8 18 48 3 31 56 52 54 6 60 pta 11 50 46 pta 11 50 46 48 58 17 pta 11 50 46 48 58 17 43 31 22 pta 11 50 46 48 58 17 43 31 22 51 53 55 19 61 33 42 18 23 pta 11 50 46 48 58 17 43 31 22 51 53 55 19 61 33 42 18 23 3 14 45 44 28 26 38 32 52 10 47 25 16 35 39 20 13 2 12 pta 38 52 11 pta 38 52 11 10 37 39 25 20 23 pta 38 52 11 10 37 39 25 20 23 16 22 31 29 58 26 42 44 48 pta 38 52 11 10 37 39 25 20 23 16 22 31 29 58 26 42 44 48 24 2 55 41 34 8 4 50 9 53 33 61 43 45 46 32 59 27 21 pta 48 20 59 pta 48 20 59 13 47 14 pta 48 20 59 13 47 14 7 58 41 pta 48 20 59 13 47 14 7 58 41 30 31 42 21 29 4 25 46 50 pta 48 20 59 13 47 14 7 58 41 30 31 42 21 29 4 25 46 50 55 15 49 28 22 54 5 3 53 37 45 44 43 17 61 8 9 27 23 pta 23 7 54 pta 23 7 54 31 38 18 pta 23 7 54 31 38 18 60 48 10 pta 23 7 54 31 38 18 60 48 10 21 61 50 29 56 40 8 37 42 pta 23 7 54 31 38 18 60 48 10 21 61 50 29 56 40 8 37 42 2 12 52 45 46 39 59 53 49 58 15 11 32 24 22 13 55 41 4 pta 51 36 45 pta 51 36 pt pta 51 36 45 41 4 26 9 18 19 pta 51 36 45 41 4 26 9 18 19 53 38 47 43 28 27 31 17 16 pta 51 36 45 41 4 26 9 18 19 53 38 47 43 28 27 31 17 16 pta 51 36 45 41 4 26 9 18 19 53 38 47 43 28 27 31 17 16 34 21 5 40 23 54 8 32 57 6 12 58 11 61 15 3 2 55 42 pta 14 21 32 pta 14 21 32 11 12 31 pta 14 21 32 11 12 31 24 59 26 pta 14 21 32 11 12 31 24 59 26 pta 14 21 32 11 12 31 24 59 26 pta 14 21 32 11 12 31 24 59 26 9 44 37 36 53 22 47 55 50 pta 14 21 32 11 12 31 24 59 26 9 44 37 36 53 22 47 55 50 54 20 13 16 60 4 40 45 8 15 51 33 23 1 39 27 56 61 38 pta 8 3 40 pta 8 3 40 7 35 41 pta 8 3 40 7 35 41 30 18 9 pta 8 3 40 7 35 41 30 18 9 pta 8 3 40 7 35 41 30 18 9 2 26 49 43 10 12 27 6 17 pta 8 3 40 7 35 41 30 18 9 2 26 49 43 10 12 27 6 17 25 4 5 55 11 22 31 45 14 46 61 60 32 1 39 29 15 38 13 pta 60 6 47 pta 60 6 47 15 37 46 pta 60 6 47 15 37 46 40 16 44 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 46 40 16 40 pta 60 6 47 15 37 40 pta 60 60 60 47 15 40 pta 60 60 60 47 15 40 pta 60 60 60 40 8 55 32 27 61 10 pta 60 6 47 15 37 46 40 16 44 18 1 48 8 55 32 27 61 10 31 3 43 9 49 39 14 52 56 21 45 29 50 30 4 7 26 58 23

### **Experiment H**

cpt 3 1 44 8 cpt 3 1 44 8 14 22 46 cpt 3 1 44 8 14 22 46 33 49 10 43 cpt 3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 cpt 3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 cpt 3 1 44 8 14 22 46 33 49 10 43 26 12 57 62 48 34 45 53 30 31 54 19 5 56 4 58 64 65 66 67 68 69 2 59 60 61 50 40 63 16 24 bfn 2 bfn 2 6 bfn 2 6 8 bfn 2 6 8 3 5 1 bfn 2 6 8 3 5 1 9 10 11 12 13 dur 1 15 dur 1 15 8 dur 1 15 8 4 13 dur 1 15 8 4 13 19 11 22 21 dur 1 15 8 4 13 19 11 22 21 23 7 16 5 6 12 2 24 jnb 1 3 jnb 1 3 10 16 jnb 1 3 10 16 24 jnb 1 3 10 16 24 31 21 27 23 11 jnb 1 3 10 16 24 31 21 27 23 11 20 25 28 22 8 26 2 5 32 14 pta 5 1 51 23 43 22 56 pta 5 1 51 23 43 22 56 21 44 11 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 pta 5 1 51 23 43 22 56 21 44 11 9 41 60 61 32 19 58 40 8 45 36 30 26 57 33 34 35 10 42 46 18 6 7 59 29 2 28

### **Experiment I**

cpt 39 27 38 36 cpt 39 27 38 36 20 47 41 cpt 39 27 38 36 20 47 41 21 17 6 35 cpt 39 27 38 36 20 47 41 21 17 6 35 32 25 7 9 15 11 13 42 29 28 cpt 39 27 38 36 20 47 41 21 17 6 35 32 25 7 9 15 11 13 42 29 28 52 55 23 37 51 18 24 16 63 40 50 61 60 59 2 69 68 67 66 65 64 bfn 17 bfn 17 16 bfn 17 16 bfn 17 16 15 14 7 4 bfn 17 16 15 14 7 4 13 12 11 10 9 dur 18 14 dur 18 14 3 dur 18 14 3 10 9 dur 18 14 3 10 9 20 17 27 26 dur 18 14 3 10 9 20 17

27 26 25 24 2 12 6 5 16 7 jnb 17 18 jnb 17 18 29 13 jnb 17 18 29 13 12 jnb 17 18 29 13 12 9 7 6 4 30 jnb 17 18 29 13 12 9 7 6 4 30 19 15 14 32 5 2 26 8 22 28 pta 25 4 47 49 pta 25 4 47 49 13 12 14 pta 25 4 47 49 13 12 14 15 27 53 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 pta 25 4 47 49 13 12 14 15 27 53 16 24 17 54 37 55 48 20 3 85 25 03 139 28 2 29 59 7 6 18 46 42 10 35 34 33

# **Experiment J**

bfn 2 6 bfn 6 2 bfn 6 2 bfn 6 2 bfn 2 6 bfn 2 6 bfn 6 2 bfn 6

# **Experiment K**

bfn 2 6 cpt 3 1 44 8 14 22 46 dur 1 15 8 els 1 3 jnb 1 3 10 16 pol 3 pta 5 1 51 23 43 22 56 pzb 1 6 vdp 1

# **Experiment L**

bfn 6 2 cpt 46 22 14 8 44 1 3 dur 8 15 1 els 3 1 jnb 16 10 3 1 pol 3 pta 56 22 43 23 51 1 5 pzb 6 1 vdp 1

# **Experiment M**

bfn 17 16 bfn 16 17 bfn 17 16 bfn 17 16 bfn 17 16 bfn 16 17 bfn 16 17 bfn 16 17 bfn 17 16 bfn 17 18 29 20 47 47 13 14 20 47 47 47 18 29 47 47 14 20 47 38 36 47 47 41 20 47 38 36 41 27 39 ext 47 29 18 18 13 ext 18 14 ext 18 3 14 dur 14 3 18 dur 14 3 18 dur 14 3 18 dur 14 18 3 ext 18 13 ext 18 17 ext 18 20 17 ext 18

### **Experiment N**

bfn 16 17 cpt 41 47 20 36 38 27 39 dur 3 14 18 els 14 13 jnb 13 29 18 17 pol 7 pta 14 12 13 49 47 4 25 pzb 11 7 vdp 8

# **Experiment O**

bfn 17 16 cpt 39 27 38 36 20 47 41 dur 18 14 3 els 13 14 jnb 17 18 29 13 pol 7 pta 25 4 47 49 13 12 14 pzb 7 11 vdp 8

### D Statistical Experiment Output

**expA:** Mean difference (SGM–CRM) = -2.85, 95% CI [-3.14, -2.57]. Shapiro–Wilk: stat = 0.7564, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 43253.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expB:** Mean difference = -3.29, 95% CI [-4.22, -2.48]. Shapiro–Wilk: stat = 0.7967, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expC:** Mean difference = +3.69, 95% CI [1.07, 7.89]. Shapiro–Wilk: stat = 0.3763, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 102.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_2$ : CRM more sensitive.

**expD:** Mean difference = -1.48, 95% CI [-1.72, -1.26]. Shapiro–Wilk: stat = 0.8032, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 12110.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expE:** Mean difference = -2.31, 95% CI [-3.06, -1.61]. Shapiro–Wilk: stat = 0.9416, p = 0.3371. Differences look normally distributed  $\rightarrow$  Paired t-test: t = -5.9993, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expF:** Mean difference = +0.85, 95% CI [0.38, 1.46]. Shapiro–Wilk: stat = 0.7143, p = 0.0002. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_2$ : CRM more sensitive.

**expG:** Mean difference = -4.18, 95% CI [-4.64, -3.73]. Shapiro–Wilk: stat = 0.8152, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 6445.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expH:** Mean difference = -4.04, 95% CI [-5.43, -2.82]. Shapiro–Wilk: stat = 0.8217, p = 0.0005. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expI:** Mean difference = +5.42, 95% CI [1.30, 11.62]. Shapiro–Wilk: stat = 0.4752, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 53.0000, p = 0.0023. Reject  $H_0$  in favour of  $H_2$ : CRM more sensitive.

**expJ:** Mean difference = -1.51, 95% CI [-1.67, -1.35]. Shapiro–Wilk: stat = 0.8739, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expK:** Mean difference = -1.61, 95% CI [-2.17, -1.10]. Shapiro–Wilk: stat = 0.7840, p = 0.0134. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0039. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expL:** Mean difference = -1.93, 95% CI [-2.91, -1.07]. Shapiro–Wilk: stat = 0.8297, p = 0.0443. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 0.0000, p = 0.0039. Reject  $H_0$  in favour of  $H_1$ : SGM more sensitive.

**expM:** Mean difference = +0.15, 95% CI [0.06, 0.24]. Shapiro–Wilk: stat = 0.7953, p = 0.0000. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 3314.0000, p = 0.0000. Reject  $H_0$  in favour of  $H_2$ : CRM more sensitive.

expN: Mean difference = -0.12, 95% CI [-0.57, 0.15]. Shapiro-Wilk: stat = 0.5978, p = 0.0001. Differences NOT normally distributed  $\rightarrow$  Wilcoxon signed-rank test: statistic = 15.0000, p = 0.4258. Fail to reject  $H_0 \rightarrow$  No significant difference.

**expO:** Mean difference = +0.37, 95% CI [-0.04, 0.87]. Shapiro-Wilk: stat = 0.9121, p = 0.3308. Differences look normally distributed  $\rightarrow$  Paired t-test: t = 1.5499, p = 0.1598. Fail to reject  $H_0 \rightarrow$  No significant difference.