Centrality Resilience in Complex Networks

Fariba Afrin Irany[1], Turja Kundu[1], Soumya Sarakar[2], Animesh Mukherjee[3], and Sanjukta Bhowmick[1]

[1] Department of Computer Science and Engineering, University of North Texas, Denton, Texas, USA
[2] Microsoft India, West Bengal, India
[3] Department of Computer Science and Engineering, IIT Kharagpur, West Bengal, India

Abstract. In this paper we study centrality resilience, that is, how well closeness and betweenness centralities are maintained under attacks. We propose efficient attack models to disrupt the rank of the top k centrality vertices. To develop our attack models, we extend the concept of rich clubs of influential vertices to the more general framework of scattered rich clubs–dense subgraphs of high-centrality vertices that are spread across the network. To improve computational efficiency, we use snowball sampling to identify these important substructures. Our results over real-world networks demonstrate that our algorithm can identify the single or scattered rich clubs efficiently and is more effective in disrupting the centrality rankings of the network, compared to other baseline methods.

Keywords: Attack Algorithm, Snowball Sampling, Scattered Rich Club

1 Introduction

Networks (or graphs) are mathematical models of complex systems of interacting entities that occur across diverse disciplines including cyber-security, bioinformatics, and mobile networks. The entities in the complex systems are represented as vertices, and their dyadic interactions as edges. Resiliency to attacks is an important property of networks. Most research focus on connectivity, that is how attacks can disconnect the network [4]. In this paper, we develop attack models to study network resilience with respect to path-based centralities. We term this type of resilience as *centrality resilience*. as opposed to the connectivity resilience of the earlier studies.

Motivation: Centrality resilience is a powerful tool for insidiously disrupting the functioning of a network without a drastic change to its structure.When one part of a network cannot communicate with the rest of the system, it is easy to infer that the cause is due to disconnectivity. Attack on centrality may not disconnect the network, but result in longer distances and more time to transmission when traversing the network. That the increased length of the distances is due to the change in the ranking of the high centrality vertices may not be immediately apparent until the centralities of the system are re-computed. Such techniques can be applied for both malicious (stealth attacks in cybersecurity, where the location of attack cannot be immediately known) and benign (reducing the load on bottlenecks, under limited resources) attacks.

Our goal is to develop edges attack models that will disrupt the centrality distribution of the network, i.e. high centrality nodes will no longer be of high centrality.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

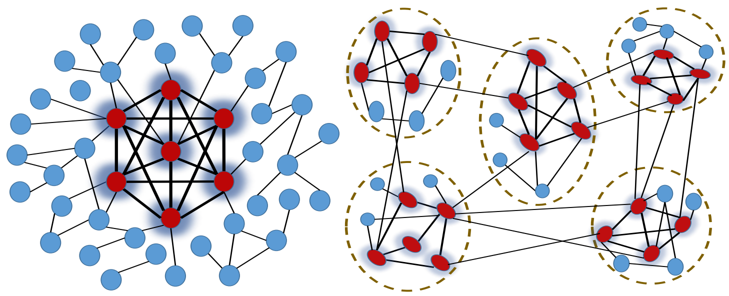

Fig. 1: Rich clubs in complex networks. (a) Network with a single rich club. (b) Network with rich clubs scattered across the network.

Key steps and contributions: Our steps to achieve this objective are as follows.

Step 1: Identify the structural properties that affect centrality resilience. (Section 2).

We identify substructures that affect centrality resilience based on the observation of [13] that path-based centralities form dense clusters or "rich clubs". *A rich club*, is an assortative subgraph, where all the vertices have high value of a vertex-based property p, in this case, betweenness or closeness centrality. Breaking these rich clubs will affect the ranking of the high centrality vertices. We extend the concept of a single rich club [13] to *scattered rich clubs*, i.e. clusters that contain high centrality vertices that may be spread across the network. Fig.1 compares the structure of networks with a single (a) and multiple scattered (b) rich club(s). Here, "high centrality vertices" refers to the union of the top-k (here set to 20) high betweenness and closeness vertices.

Contribution: We demonstrate that rich clubs of high centrality vertices can be spread across multiple clusters of a network and develop a metric to measure the *degree of scatteredness* based on the distribution of high centrality vertices across the clusters.

Step 2: Develop efficient algorithms to extract scattered rich clubs. (Section 3).

To compute the degree of scatteredness, we need to find the high centrality nodes as seeds and construct dense clusters around them. This approach is very computationally intensive. Further, several iterative steps are required to find the appropriate sets of nodes that form the clusters. We observe due to their high centrality, the vertices can reach out to many neighbors in a few hops, and thus the rich clubs have expander graph like properties. We use snowball sampling [11], which exploits the expander property of graphs, to identify the regions containing high centrality vertices.

Contribution: We develop an efficient algorithm to find scattered rich clubs using snowball sampling. We demonstrate that our sampling method can find most of the high centrality nodes and with much lower complexity than the naive method of finding high centrality vertices and then forming clusters.

Step 3: Develop attack models based on scattered rich clubs (Section 4).

Our final step is to develop attack models by removing edges that belong to the rich clubs. Attacks based on removing the vertices are equivalent to removing multiple edges. We therefore posit that edge removal is a more fine grained operation where the

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

attack is spread strategically across the networks. We use snowball sampling to find the single/scattered rich clubs, and then select edges to delete from these rich clubs.
Contribution: We develop sample-based attack strategies for disrupting the rank of high centrality vertices, based on single and scattered rich clubs. We quantify how much the ranking of the vertices have been perturbed, as per the Jaccard index. Our results show that our attack models are more effective than other baseline attack methods.

Related Work. Understanding network structure that governs robustness is an important research area. Adiga et. al. [2] studied the robustness of the top cores under sampling and in noisy networks and Laishram et. al. [8] developed core based approaches to increase network resilience. Here, we focus on resilience with respect to centrality.

There have been several papers on attack models that perturb the centrality of top k high centrality nodes. The attack models proposed in [7] target nodes with high importance and the average time complexity of the three models proposed in the paper is $O(E)$. A heuristic algorithm in [3] balances the centrality measures by link addition. [5] proposes an attack model targeting nodes with high eigenvector centrality.

Network	Type	Nodes	Edges	Avg Clus Co-eff	Max Core
dmela	Biological	7393	25569	0.01	11
euroraod	Infrastructure	1174	1417	0.016	2
HepPh	Citation	34546	420877	0.284	30
CondMat	Collaboration	23133	93439	0.63	25
as20000102	Autonomous System	6474	12572	0.25	12
caida	Autonomous System	26476	53383	0.21	22
HepTh	Citation	27770	352285	0.312	37
email-univ	Communication	1133	5451	0.22	11
AstroPh	Collaboration	18772	198050	0.63	56
grid-fission-yeast	Biological	2026	12637	0.221	34

Table 1: Test suite of networks, along with their properties.

2 SCATTERED RICH CLUBS

We evaluate our approach on ten real-world networks from SNAP(Stanford Large Network Dataset) [9] and Network repository [1], as given in Table 1. We show that high centrality vertices can be distributed across the network cores and propose a new metric, the *degree of scatteredness*, to quantify this distribution.

Distribution of rich clubs of high centrality vertices. In several networks, the high betweeness and closeness centrality vertices are located in the innermost cores [12,13]. Since the innermost cores form a dense subgraph, therefore this becomes a rich club. However, in many networks the high centrality vertices can be distributed across cores. This phenomenon requires a more general definition, that of *scattered rich clubs*.

Scattered rich club: Given a graph $G(V, E)$, a vertex property f, and a threshold value, p, $scattered\ rich\ clubs$ are a *set of disjoint* subgraphs, $\{S_1, S_2, \ldots, S_n\}$, where $S_i(V_i, E_i)$, such that $V_1 \cap V_2 \cap \ldots \cap V_n = \phi$ and $\forall v \in (V_1 \cup V_2 \cup \ldots \cup V_n)$, $f(v) \geq p$.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

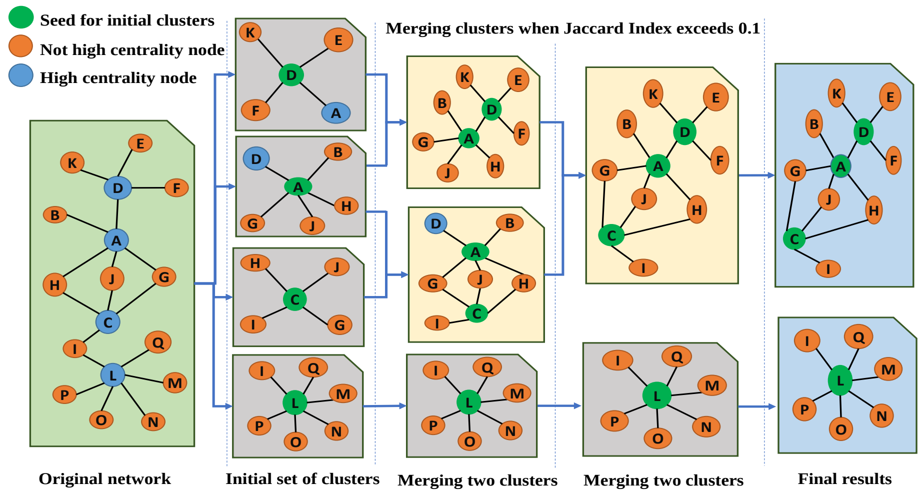

Fig. 2: Step-by-step illustration of the clustering algorithm.

The implicit expectation of a single subgraph in rich clubs is relaxed in scattered rich clubs. This generalizes the definition of rich clubs, because every network will have scattered rich clubs, even if the number of vertices in the rich clubs is one. Each scattered rich club includes at least one high-centrality node and may be augmented with a few neighboring nodes to form a dense, non-trivial cluster.

Identifying clusters forming the scattered rich clubs. The work of Estrada et. al. [6] shows that real-world networks fall in two categories. Either they have a dense core with sparsely connected periphery (single rich club) or are composed of subgraphs which are individually densely connected but have sparse inter-connection (scattered rich club).We identify scattered rich clubs using these steps, as shown in Figure 2.

First, we create the union of the top k (k=20) high betweenness (N_{hbc}) and closeness (N_{hcc}) centrality vertices of the network to create a unified set of high centrality vertices $N_{hc} = N_{hcc} \cup N_{hbc}$. *Second*, we form a cluster comprising of each node in N_{hc} and its neighbors. The total number of clusters formed be equal to the number of nodes ($|N_{hc}|$). *Third*, we merge overlapping clusters. For each cluster pair C_i, C_j, we compute the Jaccard Index as the ratio of their intersecting nodes to their total nodes. If it exceeds a given threshold (set to 0.1) we merge them as $C_i = C_i \cup C_j$ and repeat until no further merges are possible between any cluster pairs. This ensures that the maximum possible number of high-centrality nodes are put together in a cluster. At the end of these steps, we will have the disjoint clusters containing the high centrality vertices and their neighbors. If the network has a single rich club, then there will be one cluster, otherwise, there will be multiple clusters.

Quantifying the degree of scatteredness. In scattered rich clubs, the number of clusters, and the number of high centrality nodes in each cluster vary. We quantify this distribution using a new metric, *the degree of scatteredness*, as follows;

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

Let H be the total number of high centrality nodes and K the total number of clusters. Let H_i be the number of high centrality nodes in cluster C_i, ordered such that $H_i \geq H_j$ for $i < j$. For each cluster C_x, the ratio of the number of high-centrality vertices in the cluster to the number of clusters so far seen, $R_x = \frac{H_x}{x}$. The degree of scatteredness is then the mean of these ratios over the clusters; $\frac{1}{K}\left(\sum_{i=1}^{K} R_i\right)$

Using this formula, a single cluster will give the degree of scatterdness 1. When every cluster has one vertex, the value will be $\frac{1}{K}\left(\sum_{x=1}^{K} \frac{1}{x}\right)$, which will tend to zero as K becomes large. Table 2 shows the degree of scatteredness of our test networks. The more the high centrality nodes scatter into clusters, the value of scatteredness is lower.

Degree of scatteredness of Networks and high-centrality nodes prediction						
Dataset	**High-centrality nodes**	**Number of Clusters**	**Distribution of High Centrality Nodes**	**Degree of Scatteredness**	**Precision**	**Recall**
dmela	25	25	25(1)	0.152	0.07	0.88
euroroad	33	31	29(1), 2(2)	0.167	0.09	0.55
HepPh	28	21	17(1), 2(2), 1(4), 1(3)	0.293	0.04	0.96
CondMat	28	21	19(1), 1(2), 1(7)	0.362	0.14	0.79
as20000102	24	15	12(1), 2(5), 1(2)	0.402	0.70	0.79
caida	25	12	9(1), 1(11), 1(3), 1(2)	0.577	0.39	0.96
HepTh	26	10	6(1), 2(3), 1(2), 1(12)	0.609	0.15	0.77
email-univ	26	10	7(1), 2(2), 1(15)	0.683	0.15	0.85
AstroPh	31	9	6(1), 2(2), 1(21)	0.763	0.15	0.74
grid-fission-yeast	33	6	4(1), 1(26), 1(3)	0.862	0.30	0.45

Table 2: Degree of scatteredness and high-centrality nodes prediction via sampling. Multiplicity of clusters is shown as, K(M) = K clusters with M high-centrality nodes.

3 Identifying Scattered Rich Clubs Using Snowball Algorithm

While the steps in Section 2 can locate single and scattered rich clubs, in practice, finding high centrality vertices for large networks is computationally intensive. Moreover, in real-world applications, the entire network may not be available for analysis.

To address these challenges, we propose identifying the rich clubs by sampling the network, using snowball sampling. *Snowball sampling* was presented in [10], where the authors conjectured that samples with higher expansion factors are more likely to be representative of the community structure of the network. We posit that the rich clubs are good expanders, since the high centrality vertices embedded in them can, in a few hops, reach a wide set of vertices. Therefore, we modify the snowball sampling to find the high centrality vertices. The unique features of our algorithm are as follows;

We set the threshold, k, as 10% of the nodes in the network and used high-degree, high-clustering coefficient nodes as seed nodes. After each run of the sampling we obtain a sampled subgraph. We analyze the core periphery structure of this subgraph and designate the nodes in the innermost and second innermost core as high centrality nodes. We continue obtaining new snowball samples, until the set of high centrality

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

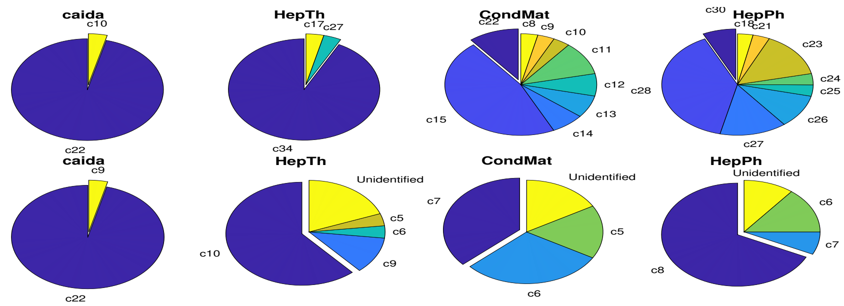

Fig. 3: Distribution of high-centrality nodes across network cores. Top: original networks. Bottom: networks sampled using the snowball in a single run with 10% of nodes.

nodes does not change, or a maximum threshold of runs (here set to 40) has been executed. At the end of this process, the sampled subgraphs are the *predicted rich clubs* and the nodes in the inner and second inner cores are the *predicted high centrality vertices.*

The algorithm's time complexity is dominated by the size of the subgraph S. Each edge in S is accessed once during snowball sampling and again during core-periphery computation. The complexity per sample is $O(T \cdot E(S))$; T is the maximum iterations.

Results and Discussion. Table 2 shows the effectiveness of the sampling algorithm in finding high centrality nodes using high degree and high clustering coefficient nodes as seed. Our ground truth is the union of the set of top 20 high betweenness centrality vertices and the top 20 high closeness centrality vertices. The precision values of the predicted set are generally low. This is because the total predicted set size can be higher than the ground truth. The recall, whether all the nodes in the ground truth were obtained is high, more that 0.70, for most of the networks with single and scattered rich clubs.

4 Attack Models for Disrupting Centrality

We develop the attack model, using the subgraphs obtained through snowball sampling, to remove edges such that the ranking of the high-centrality nodes is disrupted. As shown in Figure 3 graphs sampled using one instance of snowball sampling have most high-centrality nodes concentrated in the inner two cores. Further, the distribution of high centrality nodes of the sampled graph roughly mimic their distribution in the original graph. Thus, we deem a node to have a high core number if it is in the inner or second innermost core of the sampled subgraph, and select an edge for deletion if both its endpoints have a high core number in the sampled graph.

We test the centrality resilience by removing 2%, 4%, 6%, and 8% of the total edges. If we reach a limit of edges to choose, we relax the condition and select an edge if at least one of its endpoints has a high core number in the sampled graph. We stop the process if the required number of edges are removed or no more edges left for removal.

After edge removal, we identify the top 20 high betweenness and closeness centrality vertices of the perturbed network and compare these high centrality vertices with the ones in the original network using the Jaccard index. The closer the value is to 0,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

the more perturbed the network. The time complexity of the attack algorithm is $E(S)$, dominated by the cost of a single snowball sampling run that produces subgraph S.

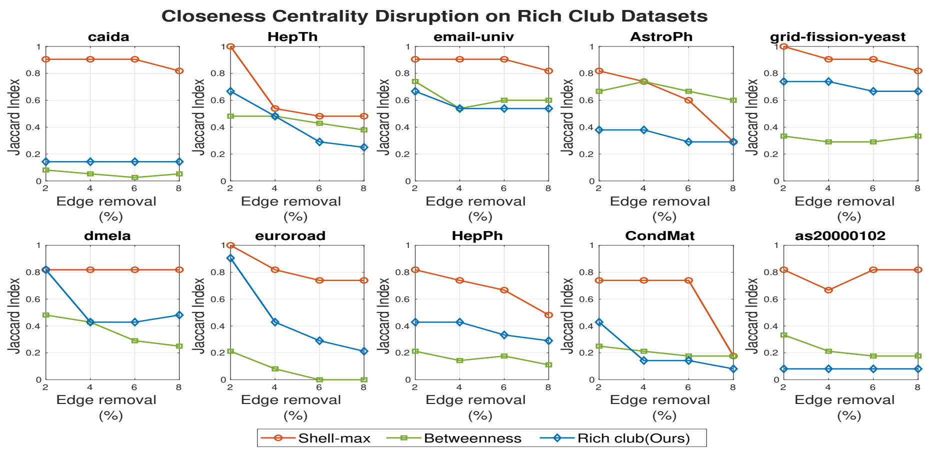

Fig. 4: Closeness centrality disruption under shell-max (red), betweenness (green), rich club (blue) attacks. Top: scatteredness > .5. Bottom: scatteredness < .5.

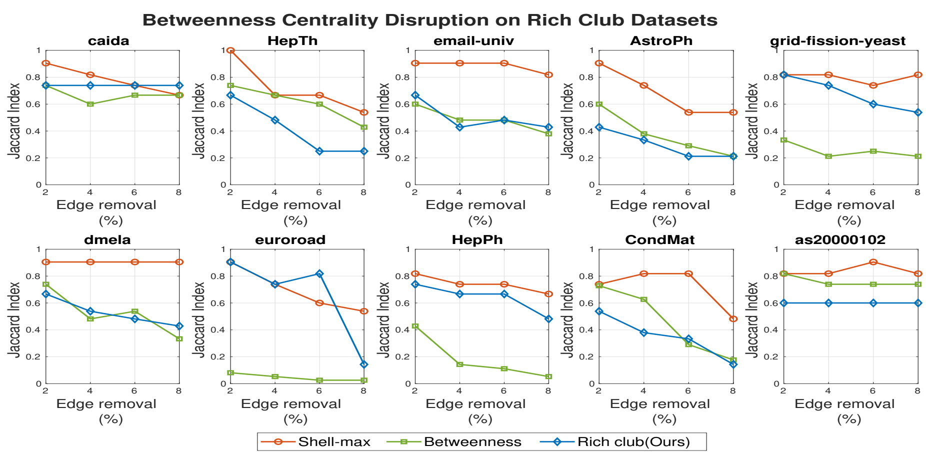

Fig. 5: Betweenness centrality disruption under shell-max, betweenness, rich club attacks. Top: scatteredness > .5. Bottom: scatteredness < .5.

Results. We compare our scattered rich club attack method (i) Shell-Max, [14] that removes edges with high k-core value ; (ii) Betweenness centrality, where edges are removed according to the order of high edge betweenness centrality and use Jaccard Index to measure the change in top 20 high centrality vertices. The lower the Jaccard index the higher the disruption. Figures 4 and 5 show that in most of the cases our method outperforms the baseline methods. Our attack model, in general, creates higher

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5

disruption than Shell-Max, and is comparable to betweenness centrality based attacks. However, the time complexity of computing edge betweenness centrality is $O(VE)$, which is an order of magnitude higher than the scatterred rich club based method.

5 Conclusion

We demonstrate that rich club of central nodes can be scattered across the network as opposed to being concentrated at the core. We discuss the implications of scattered rich clubs in terms of network centrality resilience and develop a predictive approach for discovering the scattered rich clubs using snowball sampling. Scattered rich club provide more detailed insights into how change occurs in complex networks. In future, we will study how scattered rich clubs affect other properties such as communities, and whether they can be used to predict high centrality nodes in dynamic networks.

Acknowledgment: This work was supported by NSF award CCF-1956373.

References

1. Network data repository. https://networkrepository.com/index.php
2. Adiga, A., Vullikanti, A.K.S.: How robust is the core of a network? In: ECML-PKDD. pp. 541–556. Springer (2013)
3. Alenazi, M.J., Cetinkaya, E.K., Sterbenz, J.P.: Cost-constrained and centrality-balanced network design improvement. In: 2014 6th International Workshop on Reliable Networks Design and Modeling (RNDM). pp. 194–201. IEEE (2014)
4. Chan, H., Akoglu, L., Tong, H.: Make it or break it: Manipulating robustness in large networks. In: SDM. pp. 325–333. SIAM (2014)
5. Chiluka, N., Andrade, N., Gkorou, D., Pouwelse, J.: Personalizing eigentrust in the face of communities and centrality attack. In: 2012 IEEE 26th International Conference on Advanced Information Networking and Applications. pp. 503–510. IEEE (2012)
6. Estrada, E.: Topological structural classes of complex networks. Physical Review E **75**(1), 016103 (2007)
7. Georgiadis, G., Kirousis, L.: Lightweight centrality measures in networks under attack. Complexus **3**(1-3), 147–157 (2006)
8. Laishram, R., Sariyüce, A.E., Eliassi-Rad, T., Pinar, A., Soundarajan, S.: Measuring and improving the core resilience of networks. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web. pp. 609–618 (2018)
9. Leskovec, J.: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data/
10. Maiya, A.S., Berger-Wolf, T.Y.: Online sampling of high centrality individuals in social networks. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 91–98. Springer (2010)
11. Maiya, A.S., Berger-Wolf, T.Y.: Sampling community structure. In: Proceedings of the 19th international conference on World wide web. pp. 701–710 (2010)
12. Meyer, P., Siy, H., Bhowmick, S.: Identifying important classes of large software systems through k-core decomposition. Advances in Complex Systems **17** (2014)
13. Sarkar, S., Bhowmick, S., Mukherjee, A.: On rich clubs of path-based centralities in networks. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 567–576 (2018)
14. Sun, S., Liu, X., Wang, L., Xia, C.: New link attack strategies of complex networks based on k-core decomposition. IEEE Transactions on Circuits and Systems II: Express Briefs **67**(12), 3157–3161 (2020). https://doi.org/10.1109/TCSII.2020.2973668

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_5