Journal of the Royal Statistical Society Series A: Statistics in Society, 2026, 00, 1–29
https://doi.org/10.1093/jrsssa/qnag049
Published: 6 April 2026
Original article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society
OXFORD

A network-based distributional inference approach to model country-level cyber risk

Alessandro Spelta[1], **Silvia Facchinetti**[2], **Silvia Angela Osmetti**[3], **Claudia Tarantola**[2], **Codruta Mare**[4], **Andrea Mazzali**[1], **Lorenzo Merli**[5], **Maria Iannario**[6]

[1]Department of Economics and Managements, University of Pavia, Via San Felice 5, 27100 Pavia, Italy
[2]Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio 7, 20122 Milano, Italy
[3]Department of Statistical Science, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123 Milano, Italy
[4]Department of Statistics, Forecasts, Mathematics, 58-60, Teodor Mihali Str., 400591 and the Center for Interdisciplinary Data Science, Babes-Bolyai University, Cluj-Napoca, Romania
[5]Department of Computer Science and Statistics, Trinity College Dublin, 42A Pearse Street, D02 R123, Dublin, Ireland
[6]Department of Political Sciences, University of Naples Federico II, Via Rondinò 22, 80138 Napoli, Italy
Address for correspondence: Claudia Tarantola, Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio 7, 20122 Milano, Italy. Email: claudia.tarantola@unimi.it

Abstract

This article leverages Wasserstein Propagation in Social Network to propose a novel distributional framework for the inference of cyber risk across interconnected economic systems. Cyber attacks represent an increasing threat to global security and economic stability, making the assessment of cyber risk particularly challenging, especially in countries with limited data availability. Using a comprehensive dataset on worldwide cyber attacks along with a set of macroeconomic indicators, we estimate risk profiles for all countries, including those with sparse information. Our approach reveals critical interdependencies and vulnerabilities among countries, highlighting the interconnected nature of cyber risks. The findings demonstrate the value of social network analysis in modelling cyber risk and the related uncertainty within cyberspace. Furthermore, the results offer actionable insights to strengthen global cybersecurity policies and improve resilience against cyber threats.

Keywords cyber risk, optimal transport, social networks, Wasserstein propagation

1 Introduction

In an increasingly interconnected world, cyber attacks pose a significant threat to national security, economic stability, and societal well-being. As countries deepen their reliance on digital infrastructure, the frequency and severity of cyber incidents continue to escalate, impacting governments, businesses, and individuals alike. Understanding and mitigating cyber risks has become a priority for policymakers who must effectively allocate resources to strengthen cybersecurity, enhance resilience, and protect critical infrastructures.

Cyber risk is inherently complex and multifaceted, characterized by its dynamic nature, uncertainty, and the interplay of various economic, technological, and geopolitical factors. Measuring this risk is

Received: March 16, 2025. **Revised:** November 3, 2025. **Accepted:** March 5, 2026

challenging, particularly when data on the severity and likelihood of cyber attacks are incomplete or unavailable for certain countries. Without accurate risk assessments, policymakers struggle to design targeted interventions and allocate resources effectively to areas of greatest need sometimes leading to the worst approach in terms of dealing with uncertainty in cyberspace—ignoring it through inaction (van den Berg, 2024). Risk and uncertainty are two of the critical aspects inherently linked to digitalization and to cyberspace (Brantly, 2021), that are to be directly linked to cyber attacks. We do not know when a cyber attack will happen and what damage it will produce, but we know that there is always a risk for such an attack to appear. Any effort to model the factors influencing the likelihood of cyber attacks plays a vital role in mitigating the costs associated with uncertainty (Schulzke, 2018). The existing literature emphasizes the necessity and value of using network analysis to model, forecast, and evaluate cyber attacks (see Facchinetti et al., 2023; GhasemiGol et al., 2016). To address these challenges, we propose the following approach. For each country for which data on the severity distribution are available, we construct the empirical probability distribution of cyber attack severity. These country-level distributions are then propagated through a properly constructed country-level economic network to infer the severity distributions for countries with missing or incomplete data. This approach is innovative as it combines empirical distribution modelling with network-based information diffusion, enabling the estimation of cyber risk even in the absence of direct observations.

We analyse a dataset that contains information on more than 7,000 serious cyber attacks reported globally in 2023. The dataset was provided by Hackmanac, a Dubai-based organization specializing in monitoring global cyber threats. Additional information about Hackmanac can be found on their website: https://hackmanac.com/. In the dataset, attacks are categorized by severity, using an ordinal variable that quantifies the intensity of each attack. For more details on the data classification and previous analysis, see Andreolini et al. (2004), Facchinetti et al. (2023), and Facchinetti et al. (2024), among others. The dataset is integrated by macroeconomic information regarding the country involved.

By employing Wasserstein propagation (Solomon et al., 2015, 2014; Spelta et al., 2025), a method rooted in optimal transport theory, we can infer the cyber risk profiles of countries lacking detailed attack data. This approach not only fills critical data gaps but also provides a framework for assessing the diffusion of cyber risks across interconnected economies. From a policymaking perspective, this methodology is transformative. By understanding the distribution of the severity of cyber attacks for each country, decision makers can prioritize investments in cybersecurity infrastructure, foster international collaborations to mitigate cross-border risks, and implement evidence-based policies to protect economic and social systems, thus lowering uncertainty related to the field. Ultimately, this framework empowers policymakers with actionable insights to pre-emptively address vulnerabilities and build a more resilient global cybersecurity landscape.

Propagation techniques serve as fundamental tools in machine learning and data analysis, enabling the transfer of information between structured domains such as networks and graphs (Zhou et al., 2004). These methods have proven invaluable in various applications, from semisupervised learning to network analysis, where the goal is to extend partial information to the entire domain while preserving the underlying structural properties (Zhu et al., 2003). Traditional propagation approaches have focused mainly on scalar quantities, leveraging principles from graph theory, spectral analysis, and optimization (Belkin et al., 2004). These methods typically frame propagation as an energy minimization problem on graphs, where the objective is to ensure smooth transitions while respecting known values at the labelled nodes (Zhu & Ghahramani, 2002). For instance, harmonic function approaches model the problem through the lens of electric networks and random walks, leading to elegant solutions based on graph Laplacian systems (Kondor & Lafferty, 2002). Such scalar propagation methods have demonstrated remarkable success in tasks such as semisupervised classification and regression, offering both computational efficiency and theoretical guarantees (Bengio et al., 2006). However, many real-world phenomena naturally manifest as distributions rather than scalar values. Examples include traffic density patterns, climate data variations, and product rating profiles. In these scenarios, scalar propagation methods prove fundamentally inadequate as they fail to capture the rich structure inherent in distributional data. Attempting to apply scalar techniques to individual components of distributions (such as histogram bins) not only ignores crucial interdependencies but can also introduce artefacts like over smoothing or spurious multimodality (Solomon et al., 2014). To address these limitations, Wasserstein propagation has emerged as a principal framework

for extending propagation methods to distributional data (Solomon et al., 2015). This approach leverages optimal transport theory, which provides a natural geometry for comparing and interpolating probability distributions (Villani, 2008). Unlike traditional divergence measures such as Kullback–Leibler divergence, the Wasserstein distance inherently respects the underlying geometry of the probability space (Peyré & Cuturi, 2019). This property makes it particularly suitable for applications where distributions evolve smoothly over domains. The Wasserstein propagation framework reformulates the classical energy minimization problem by replacing scalar differences with Wasserstein distances (Solomon et al., 2014). This modification preserves distributional properties such as mean, variance, and multimodality during the propagation process. The resulting methodology has found successful applications in diverse fields, including semisupervised learning with probabilistic labels (Arjovsky et al., 2017), texture synthesis in computer graphics (Rabin et al., 2011), and modelling information diffusion in networks (Panaretos & Zemel, 2019).

In this work, differently from Solomon et al. (2014) which deals with the classical Dirichlet energy minimization principle, extending it to distributional data, we primarily focus on computing barycentre-based propagation and smoothing that adapts to the geometry of the domain. Wasserstein barycentres represent a sophisticated approach to averaging probability distributions while preserving their inherent geometric and probabilistic characteristics. At their core, these barycentres operate by minimizing the transport effort between distributions, following optimal transport principles, rather than simply averaging values pointwise. This fundamental property makes them especially valuable when working with distributional data rather than scalar values. The Wasserstein barycentre (Spelta, 2026; Spelta et al., 2025; Spelta & Pecora, 2024) serves as a central distribution that optimally balances its inputs, taking into account both their weighted contributions and spatial relationships. This approach overcomes simple blending by incorporating the alignment, spread, and shape of the input distributions, resulting in averages that are both mathematically rigorous and intuitively meaningful within the problem domain. In the context of propagation tasks, Wasserstein barycentres enable the natural spread of distributional data across various domains, such as graphs, meshes, or manifolds. This process ensures smooth transitions between distributions while preserving essential characteristics like variability and multimodality. The use of the Wasserstein metric provides stability to the propagation process, effectively preventing common issues such as over smoothing or the emergence of artificial features. This combination of properties makes Wasserstein barycentres particularly effective for applications requiring sophisticated distribution averaging and propagation while maintaining the integrity of the underlying data structure.

The article is structured as follows. Section 2 is devoted to the theoretical framework. It begins with a review of optimal transport theory and Wasserstein propagation for distributional data on networks. Section 3 provides a detailed description of the cyber risk dataset, including severity classification and construction of the countries' networks. Section 4 presents the results, starting with the construction and evaluation of network structures with particular emphasis on cyber risk propagation accuracy. It also examines the role of node centrality in cascade dynamics. Finally, Section 3 concludes the article with a brief summary and concluding remarks.

2 Methods

In this section, we present the methodological framework underlying the application of Wasserstein propagation to cyber risk modelling. The approach integrates concepts from optimal transport theory with network analysis to propagate cyber attack severity distributions across interconnected economic systems. The methodology is designed to address the challenge of estimating risk profiles for countries lacking direct data, leveraging information from countries with known distributions.

2.1 Optimal transport theory

Optimal transport provides a robust mathematical framework for comparing probability distributions by calculating the minimal cost required to transform one distribution into another (Villani, 2008). Unlike traditional divergence measures, the Wasserstein distance incorporates structured

relationships within the data to provide a meaningful comparison between distributions. This makes it particularly suited for analysing scenarios where the data exhibit inherent order. In this study, we focus on probability distributions over the ranks of an ordinal variable representing cyber attack severity.

More formally, let two distributions be defined as $\mu = \sum_{i=1}^{l} u_i \delta_{x_i}$ and $\nu = \sum_{i=1}^{l} v_i \delta_{x_i}$, where $\{x_1, \ldots, x_l\}$ represents the support, and δ_{x_i} is the Dirac measure centred on x_i.[1] These distributions are fully described by the probability mass vectors $\mathbf{u} = (u_1, \ldots, u_l)$ and $\mathbf{v} = (v_1, \ldots, v_l)$, both defined over the simplex Σ_l.

The distance between μ and ν is defined as:

$$d(\mu, \nu) = \min_{\omega \in \mathcal{U}(\mu,\nu)} \langle \omega, \mathbf{C} \rangle, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. The transport plan $\omega \in \mathcal{U}(\mu, \nu)$ specifies the joint distribution between μ and ν over the product space $\{x_1, \ldots, x_l\} \times \{x_1, \ldots, x_l\}$. The transport polytope $\mathcal{U}(\mu, \nu)$ is defined as:

$$\mathcal{U}(\mu, \nu) = \left\{ \omega \in \mathbb{R}_+^{n \times n} \mid \omega \mathbf{1}_l = \mu, \omega^\top \mathbf{1}_l = \nu \right\}.$$

Here, $\mathbf{C} \in \mathbb{R}^{l \times l}$ represents the cost matrix, where each entry $C_{i,j}$ corresponds to the transport cost between x_i and x_j. When the cost is set as $C_{i,j} = \|x_i - x_j\|^2$, the resulting distance $d(\mu, \nu)$ is the ℓ_2-Wasserstein distance, denoted as $W(\mu, \nu)$.

The computation of $W(\mu, \nu)$ can be performed by solving the linear assignment problem (LAP) described in Equation (1), using specialized algorithms such as the Hungarian algorithm (Kuhn, 1955) or methods like the Earth Mover's Distance (Rubner et al., 1997) and Fast Transport (Pele & Werman, 2009). However, these methods are computationally expensive, as noted in Cuturi (2013). To alleviate this, an entropy-regularized variant of the problem, known as the Sinkhorn divergence, is introduced:

$$W_\epsilon(\mu, \nu) := \min_{\omega \in \mathcal{U}(\mu,\nu)} \langle \omega, \mathbf{C} \rangle + \epsilon H(\omega),$$

where $\epsilon > 0$ is a regularization parameter and $H(\omega) := \sum_{i,j} \omega_{i,j} \log \omega_{i,j}$ is the negative entropy.[2] This regularization makes $W_\epsilon(\mu, \nu)$ differentiable and solvable via Sinkhorn iterations (Cuturi, 2013).

Using the Gibbs kernel $\mathbf{K} := \exp\left(-\frac{\mathbf{C}}{\epsilon}\right) \in \mathbb{R}_+^{n \times n}$ and initializing $\mathbf{p}^{(0)} = \mathbf{1}_l$, the Sinkhorn iterations are given by:

$$\mathbf{q} \leftarrow \frac{\nu}{\mathbf{K}^\top \mathbf{p}}, \quad \mathbf{p} \leftarrow \frac{\mu}{\mathbf{K}\mathbf{q}}. \tag{2}$$

As shown by Cuturi (2013), this method converges to the optimal transport plan:

$$\omega = \mathrm{diag}(\mathbf{p}) \mathbf{K} \mathrm{diag}(\mathbf{q}).$$

2.1.1 Wasserstein barycentres

Before delving into the details of the propagation framework, it is useful to review the definition and computation of Wasserstein barycentres. Barycentre computation is formulated as an inverse problem that relies on the Sinkhorn operator, as outlined in Equation (2). The construction of a barycentre for a set of distributions depends on two key components: (i) a measure of distance between the

[1] For simplicity and without loss of generality, we assume a common support, although the method can be extended to disjoint supports.

[2] The usual convention $0 \log(0) = 0$ applies here.

distributions and (ii) a definition of the mean. In the context of optimal transport, the Wasserstein distance serves as the metric to quantify discrepancies between distributions, while the Fréchet mean provides the foundation for defining the barycentre.

In this study, the distributions examined represent the severity of cyber attacks across different countries. The barycentre provides a central distribution that optimally summarizes these severity profiles, accounting for both the variability and the ordinal nature of the severity levels. This makes Wasserstein barycentres particularly suitable for capturing the complex patterns in the propagation of cyber risks across nations.

The primary objective is to interpolate between a set of distributions $\{\beta_1, \ldots, \beta_M\} \in \mathbb{R}^I$ using a convex combination of weights $\lambda \in \Sigma_M$, which lie on the probability simplex. This approach yields a barycentre that balances the contributions of the input distributions.

Specifically, the barycentre β^* is computed by solving the following optimization problem:

$$\beta := \arg\min_{\beta \in \Sigma_I} \sum_{m=1}^{M} \lambda_m W_\epsilon(\beta, \beta_m), \quad \text{where } W_\epsilon(\beta, \beta_m) := \min_{\omega \in \mathcal{U}(\beta, \beta_m)} \langle \omega, \mathbf{C} \rangle - \varepsilon H(\omega), \tag{3}$$

and $W_\epsilon(\beta, \beta_i)$ is the entropy-regularized Wasserstein distance, as defined earlier. The transport plan ω belongs to the transport polytope $\mathcal{U}(\beta, \beta_m)$, and $\mathbf{C}$ represents the cost matrix. The optimization problem in Equation 3 can be solved efficiently using Sinkhorn iterations described in Equation 2.

Following Cuturi (2013), the barycentre β^* is computed iteratively using the Sinkhorn algorithm. The solution for the barycentre is expressed as:

$$\beta \overset{\text{def.}}{=} \prod_{m=1}^{M} \left(\mathbf{K}^\top \mathbf{p}_m\right)^{\lambda_m}, \quad \text{where } \begin{cases} \mathbf{q}_m \leftarrow \beta \oslash (\mathbf{K}^\top \mathbf{p}_m), \\ \mathbf{p}_m \leftarrow \beta_i \oslash (\mathbf{K}\mathbf{q}_m), \end{cases}$$

with initialization $\mathbf{p}_m^{(0)} = \mathbf{1}_I$ and $\mathbf{K} := \exp\left(-\frac{\mathbf{C}}{\varepsilon}\right) \in \mathbb{R}_+^{n \times n}$. This iterative approach refines the barycentre by alternating updates to the scaling factors $\mathbf{p}_m$ and $\mathbf{q}_m$, ensuring convergence to the optimal solution.

2.2 Wasserstein propagation framework

Building on the foundation of Wasserstein barycentres, we now present a generalized framework for distribution propagation across graph structures. This framework, known as Wasserstein propagation, extends the barycentre computation to handle more complex interpolation scenarios.

Let $G = (V, E)$ be a weighted graph where V is the set of nodes with size $|V| = N$ and E is the set of all connections among them (edges) with size $|E| = L$. The resulting network can be represented by its weighted adjacency matrix $\mathcal{S} \Leftrightarrow \mathbf{S}(G) = [s_{v,w}]_{1 \le i,j \le N}$.

Each vertex $v \in V$ is associated with a probability distribution $\beta_v \in \Sigma_I$. Given a subset of vertices $V_0 \subseteq V$ with known distributions, the goal is to interpolate distributions for the remaining vertices while preserving the geometric structure encoded by the graph.

The optimization problem can be formulated as:

$$\min_{(\beta_v)_{v \in V}} \sum_{(v,w) \in E} s_{v,w} W_\epsilon(\beta_v, \beta_w) \quad \text{subject to } \beta_v \text{ fixed for all } v \in V_0$$

where $s_{v,w}$ represents the edge weights and $W_\epsilon(\beta_v, \beta_w)$ is the entropy-regularized Wasserstein distance:

$$W_\epsilon(\beta_v, \beta_w) := \min_{\omega \in \mathcal{U}(\beta_v, \beta_w)} \langle \omega, \mathbf{C} \rangle - \varepsilon H(\omega).$$

The problem can be solved by optimizing over the transportation plans ω_e for each edge $e \in E$:

$$\begin{aligned} &\min_{\omega_e} \sum_{e\in E} s_{v,w}[\langle \omega_e, \mathbf{C}\rangle - \varepsilon H(\omega_e)] \\ \text{subject to}: \quad &\omega_e \mathbf{1} = \beta_v \quad \forall e = (v, w) \\ &\omega_e^\top \mathbf{1} = \beta_w \quad \forall e = (v, w) \\ &\beta_v \text{ fixed} \quad \forall v \in V_0 \end{aligned}$$

The solution is computed through iterative updates of scaling vectors. For vertices with fixed distributions ($v \in V_0$), the updates are given by:

$$\begin{aligned} \beta &= \beta_0(v) \\ \mathbf{q}_e &= \beta \oslash (\mathbf{K}^\top \mathbf{p}_e) \quad \text{for } e = (w, v) \\ \mathbf{p}_e &= \beta \oslash (\mathbf{K}\mathbf{q}_e) \quad \text{for } e = (v, w). \end{aligned}$$

For vertices with unknown distributions ($v \notin V_0$), the updates proceed as:

$$\begin{aligned} \alpha &= \sum_{e\in N(v)} s_e \\ \beta &= \mathbf{1} \\ \mathbf{d}_e &= \begin{cases} \mathbf{q}_e \otimes (\mathbf{K}^\top \mathbf{p}_e) & \text{for } e = (w, v) \\ \mathbf{p}_e \otimes (\mathbf{K}\mathbf{q}_e) & \text{for } e = (v, w) \end{cases} \\ \beta &= \beta \otimes \mathbf{d}_e^{s_e/\alpha} \\ \mathbf{q}_e &= \mathbf{q}_e \otimes \beta \oslash \mathbf{d}_e \quad \text{for } e = (w, v) \\ \mathbf{p}_e &= \mathbf{p}_e \otimes \beta \oslash \mathbf{d}_e \quad \text{for } e = (v, w), \end{aligned}$$

where $\mathbf{K} = \exp(-\mathbf{C}/\varepsilon)$ is the Gibbs kernel, $\otimes$ denotes element wise multiplication, and $\oslash$ denotes element wise division.

The framework encompasses two important special cases. First, when G is a star graph with all vertices belonging to V_0 on the spokes and a vertex with unknown distribution at the centre, the problem reduces to the standard barycentre computation. Second, when G is a line graph with two fixed vertices in V_0 at the endpoints, the framework provides a solution for displacement interpolation. In these two settings, the algorithm converges to a local minimum of the objective function (Solomon et al., 2015). The quality of the solution depends on the entropy regularization parameter ε, the graph structure and edge weights, and the initialization of the scaling vectors. For practical applications, the solution provides a smooth interpolation of distributions that respects both the local structure encoded in the edge weights and the global topology of the graph.

3 Data description

3.1 Cyber risk data

Given the significant impact of cyber risk across countries, it is crucial to estimate the propagation of cyber attack distributions for countries lacking sufficient data to assess and manage this risk effectively. The Wasserstein propagation allows us to estimate the cyber attack severity distributions for countries with limited data, leveraging similarities inferred from nations with more comprehensive datasets. This method provides valuable insights into whether countries face similar levels of cyber threat, which may be influenced by factors such as shared technological infrastructures, regulatory environments, or economic characteristics. The findings of this propagation process can help guide national cybersecurity policies, promote international cooperation, strengthen collective defences, and inform targeted investments in critical sectors such as finance and infrastructure.

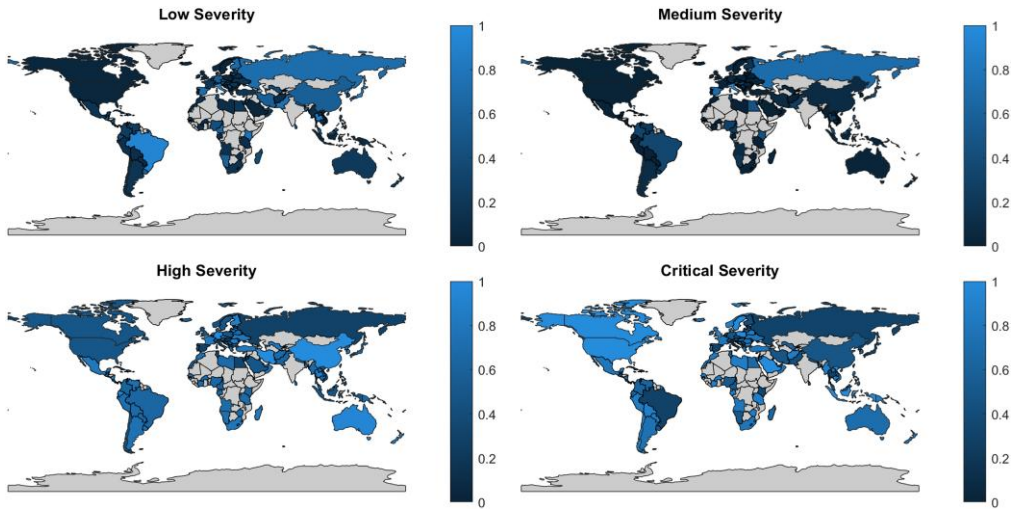

Figure 1 Severity distribution of cyber attack across countries. Lighter shades indicate higher probabilities, while grey areas denote countries where data is unavailable.

For this analysis, we use data from Hackmanac[3], a company that monitors the behaviour and evolution of real cyber threats worldwide. Our focus is on cyber attacks that occurred globally in 2023. The dataset includes 7059 daily observations of cyber attacks, each categorized by country that undergoals a cyber attack, and attack severity. The severity of an attack is assigned by experts on the basis of geopolitical, economic, and reputational and cost/opportunity impact. The severity variable classifies the gravity of attacks into four levels: low (Grade 1), medium (Grade 2), high (Grade 3), and critical (Grade 4). For each country, instead of using a single representative value of cyber attack severity, we consider the empirical probability distribution of the severity variable over its four categories.

To ensure the reliability of our analysis, we exclude countries with fewer than two observations, resulting in a final dataset of 123 countries. Using Wasserstein propagation, we estimate severity distributions for countries with insufficient data, enabling the identification of common risk profiles and intercountry dependencies.

Figure 1 presents a visual representation of the probability distribution of cyber attack severity levels across countries. It offers a global overview of cyber attack risk, classifying countries based on their likelihood of experiencing different severity levels: low, medium, high, and critical. Each map corresponds to a specific severity level, with shading intensity representing the probability of that severity in each country. Lighter shades of blue indicate higher probabilities, while grey areas denote countries where data are unavailable. In the low-severity map, countries such as Brazil, Russia, and parts of Southeast Asia show a higher likelihood of experiencing low-impact cyber attacks. This indicates that these regions may be less targeted by highly destructive attacks but still encounter minor disruptions. The medium severity map highlights regions such as Europe and Russia, where attacks are of moderate severity. This suggests vulnerabilities in the systems of these regions that could lead to noticeable disruptions. High-severity attacks are predominantly observed in countries like Australia, Germany, and China, which are often highly connected and technologically advanced, making them attractive targets for impactful cyber attacks. These regions exhibit a higher probability of attacks that can significantly disrupt operations or cause economic damage. Lastly, the critical severity map emphasizes regions such as the U.S., the U.K., and parts of Europe and the Middle East, where the most destructive cyber attacks are likely to occur. These areas face heightened risks, potentially due to their critical infrastructure, strategic importance, or high-value targets.

[3] https://hackmanac.com/

4 Network implementation and empirical results

In this section, we present the results of our analysis, structured around the following key steps. First, we construct a network of countries using macroeconomic indicators, employing four approaches: cosine (COS) similarity, minimum spanning tree (MST), k-nearest neighbours (KNN), and the salience (SAL) algorithm. Second, we evaluate the efficacy of these network structures in propagating cyber attack severity distributions. Third, we apply the propagation framework to estimate the distributions for countries lacking data and assess the robustness of the methodology. Finally, we analyse the role of network centrality in understanding cascade dynamics and highlight its implications for risk mitigation strategies.

4.1 Building the network

The network structure plays a pivotal role in Wasserstein propagation, as it defines the pathways through which risk distributions are transmitted across countries. In the context of cyber risk, economic, technological, or geopolitical connections between countries form the underlying framework that facilitates the diffusion of vulnerabilities and attack severity levels. This interconnectedness allows the propagation mechanism to account for both direct and indirect influences, capturing complex dependencies within the global risk landscape. By leveraging the network, Wasserstein propagation can model how cyber risk spreads and amplifies, providing critical insights into potential hotspots and cascading vulnerabilities. The network topology, including the strength and nature of connections, is therefore essential to accurately assess and mitigate cyber risks in regions with limited data availability.

Against this backdrop, we construct reliable network structures using macroeconomic variables from the 2022 Worldwide Governance Indicators (WGI), observed 1 year before the cyber attacks. These indicators include Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. Further details on the data sources, aggregation methodology, and interpretation of these indicators can be found in Kaufmann et al. (2010a). We rely on the WGI indicators for several reasons. First, it is well documented that the WGIs are highly correlated with the level of economic development of a country (Kaufmann et al., 2010b). This supports their use as proxies for institutional and socioeconomic conditions that are relevant to cyber risk. Second, they make a comprehensive assessment of the institutional quality and socioeconomic development, digitalization issues included. Third, the fact that they are by default standardized ensures methodological consistency across countries, and provide robust network construction. As such, the network approach that we employ is not sensitive to variables' measurement units. Fourth, the WGIs provide a wide country coverage. As one of the goals of our analysis is to extrapolate our results to the countries for which we have missing data in respect to cyber attacks, governance quality is a particularly meaningful proxy for cyber vulnerability.

To construct the network we proceed as follows. First, to evaluate the similarity between countries' behavioural features, the COS measure is employed (see Cardillo et al., 2023; Spelta et al., 2018). For two countries, with their macroeconomic characteristics represented by vectors $\mathbf{y}_v$ and $\mathbf{y}_w$, each containing the macroeconomic variables examined, the COS similarity is given by:

$$\cos(\mathbf{y}_v, \mathbf{y}_w) = \frac{\langle \mathbf{y}_v, \mathbf{y}_w \rangle}{\|\mathbf{y}_v\| \|\mathbf{y}_w\|},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\| \cdot \|$ represents the Euclidean norm. Using this metric, the distance between the countries is defined as:

$$\theta_{v,w} = 1 - \cos(\mathbf{y}_v, \mathbf{y}_w).$$

In this formulation, highly similar pairs of countries have smaller distances, while dissimilar pairs are associated with larger distances. The resulting distance matrix $\boldsymbol{\Theta}$ encodes these relationships for all pairs of countries.

The fully connected nature of the distance matrix $\boldsymbol{\Theta}$ can obscure meaningful patterns. To address this, we sparsify the matrix in three ways. We employ the MST, the KNN method, and the SAL approach of Grady et al. (2012).

The MST reduces complexity while preserving key structural information. Starting from the $N \times N$ matrix $\boldsymbol{\Theta}$, we use hierarchical clustering to identify the MST. At each step of the process, two clusters, l_v and l_w, are merged if they satisfy:

$$\theta_{l_v,l_w} = \min\{\theta_{\hat{v},\hat{w}}\},$$

where $\hat{v} \in l_v$ and $\hat{w} \in l_w$. This process is repeated until a single cluster is formed. The MST is the final tree structure that connects all nodes using $N-1$ edges while minimizing the total edge weight, and the respective adjacency matrix is indicated with $\mathbf{A}^{\mathrm{MST}}$. The threshold distance L is defined as the largest edge distance in the MST.

To prune the network, the KNN approach is used to construct the adjacency matrix $\mathbf{A}^{\mathrm{KNN}}$. Starting from the pairwise distance matrix $\boldsymbol{\Theta}$, the procedure is as follows. For each node v, the distances $\{\theta_{v,w}\}_{j=1}^{N}$ are sorted in ascending order. The $k=2$ smallest distances, excluding $\theta_{v,v}$ (self-distance), identify the indices of the KNN of node v, denoted as $\mathcal{N}_v$. The adjacency matrix is then updated as:

$$a_{v,w}^{\mathrm{KNN}} = \begin{cases} \theta_{v,w}, & \text{if } w \in \mathcal{N}_v \text{ or } v \in \mathcal{N}_w, \\ 0, & \text{otherwise.} \end{cases}$$

This process ensures that the resulting adjacency matrix $\mathbf{A}^{\mathrm{KNN}}$ is symmetric and connects each node to its KNN.

Finally, for extracting meaningful structural features from the fully connected distance matrix $\boldsymbol{\Theta}$ we also introduce the concept of link SAL (Grady et al., 2012). Link SAL is based on the notion of effective proximity, a concept that captures the intuitive principle that strongly (weakly) coupled nodes are close to (distant from) each other. It also provides one way to define the length of a path that connects two terminal nodes (n_{Z_1}, n_{Z_k}) and consists of a sequence of $Z-1$ intermediate nodes n_{Z_i}, and connections $\theta_{n_{z_j} n_{z_j+1}} > 0$. The shortest path minimizes the total effective distance $\tau = \sum_{z_j=1}^{Z-1} \theta_{n_{z_j} n_{z_j+1}}$ and can be interpreted as the most efficient route between its terminal nodes. Given the notion of the shortest path that originates at node z_k and ends at node z_l, one can introduce the indicator function:

$$\iota_{v,w}(z_l, z_k) = \begin{cases} 1 & \text{if edge } v \to w \text{ is on the shortest path from } z_k \text{ to } z_l \\ 0 & \text{otherwise} \end{cases}.$$

A shortest path tree $\Upsilon(z_k)$ rooted at node z_k can be represented as a matrix with elements:

$$\upsilon_{v,w}(z_k) = \begin{cases} 1 & \text{if } \sum \iota_{v,w}(z_l, z_k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

and the SAL $a_{v,w}^{\mathrm{SAL}}$ of edge $v \to w$ is given by:

$$a_{v,w}^{\mathrm{SAL}} = \frac{1}{N} \sum_{z_k} \upsilon_{v,w}(z_k). \tag{4}$$

The SAL adjacency matrix $\boldsymbol{A}^{\mathrm{SAL}} = [a_{v,w}^{\mathrm{SAL}}]$ as defined by Equation 4 permits an intuitive definition of the skeleton of a network.

Since the propagation mechanism relies on similarity rather than distance, the final networks are generated by transforming distances into similarities using the reciprocal formula. In a nutshell, for the COS distance, we obtain the similarity matrix $s_{v,w}^{\mathrm{COS}} = \frac{1}{\theta_{v,w}}$, similarly, for the MST we have $s_{v,w}^{\mathrm{MST}} = \frac{1}{a_{v,w}^{\mathrm{MST}}}$, for the KNN $s_{v,w}^{\mathrm{KNN}} = \frac{1}{a_{v,w}^{\mathrm{KNN}}}$ and for the SAL network $s_{v,w}^{\mathrm{SAL}} = \frac{1}{a_{v,w}^{\mathrm{SAL}}}$.

Figure 2 provides a visualization of the distance matrix and of the clustering results using COS distance among countries. The left panel illustrates the COS distance matrix Θ, which highlights the pairwise dissimilarity between countries based on the given features. The red dashed lines indicate the partitioning derived from applying the k-means clustering algorithm with the silhouette criterion, which identified two optimal clusters out of a tested range of up to 10 clusters. The central and right panels present the bar chart of the distribution of the countries grouped into macroareas for each cluster, revealing a clear geographical pattern. Cluster 1 predominantly includes European countries, as evidenced by the high probability (of more than 0.5) associated with Europe. In contrast, Cluster 2 comprises a mix of Asian and African countries, along with smaller contributions from other regions. However, their individual probabilities are much lower than the one of Europe in Cluster 1. These findings suggest that COS similarity captures shared characteristics aligning with geographic and potentially socioeconomic or cultural features, underscoring the geographical structure present in the dataset. Figure 3 presents the three distinct network representations used to sparsify the fully connected (COS) matrix, namely the MST, the KNN, and the SAL, which has been symmetrized for visual purposes. In all three networks, the node colours are proportional to the degree, indicating the number of connections each node has. The MST network, shown in the left panel, exhibits a sparse structure typical of the tree networks, where all nodes are connected with the minimum total edge weight and no cycles are formed. The resulting topology is characterized by its tree-like structure, with limited branching and relatively few connections per node. The uniformity in node colour suggests that most nodes have a similar, low degree of connectivity, which is expected in this type of network where each node is connected only to the nearest node or a very limited number of others. The KNN network, depicted in the central panel, introduces a denser structure compared to the MST. In this representation, each node is connected to its nearest neighbours based on a predefined value of $k = 2$, which results in more local connections. The degree of the nodes in the KNN network shows greater variability than in the MST, as indicated by the more pronounced differences in node colour. Some nodes, particularly those located in dense clusters or in the centre of the network, possess a higher degree, acting as small hubs that connect to multiple other nodes, while others maintain fewer connections. The SAL network, displayed in the right panel, represents the most complex and interconnected structure of the three. This network incorporates both local and global patterns of connectivity, resulting in a highly dense and cohesive arrangement. The variation in node degree is reflected through the wide range of node colours, with several nodes exhibiting significantly higher degrees, which are likely functioning as central hubs within the network. These high-degree nodes are critical for maintaining the global structure of the network, while nodes with lower degrees are distributed more peripherally.

Table 1 provides a detailed summary of the structural metrics for the networks under investigation. The metrics offer insights into how the network topology and flow characteristics differ across these network types. The number of nodes (Nodes) remains fixed at 145 across all configurations, indicating

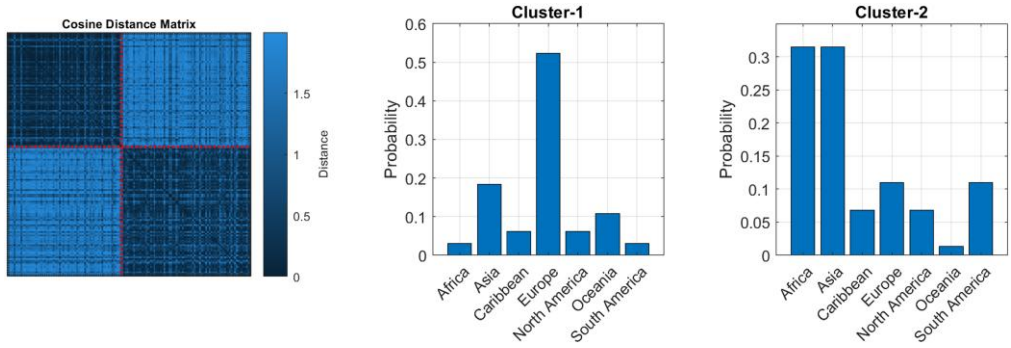

Figure 2 Cosine distance matrix (left) and k-means clustering results for countries, with the silhouette criterion identifying two clusters. The central and right panels display the distribution of countries grouped into macroareas for Cluster 1 and Cluster 2, respectively.

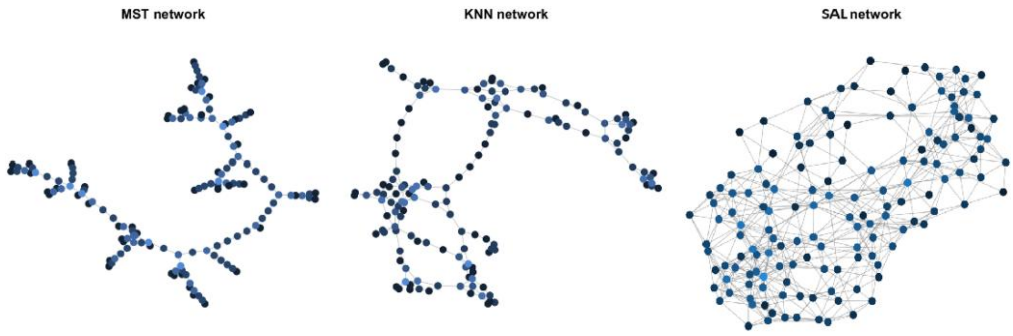

Figure 3 Comparison of network representations used to sparsify the fully connected COS matrix: MST, KNN ($k = 2$), and SAL networks.

Table 1. Structural metrics of networks derived for varying pruning algorithms: Cosine similarity ($\mathbf{S}^{COS}$), MST ($\mathbf{S}^{MST}$), KNN ($\mathbf{S}^{KNN}$), and SAL ($\mathbf{S}^{SAL}$). Each cell contains the computed values for the number of nodes, number of edges, density, average flow, diameter, average minimum path length, and clustering coefficient.

	$\mathbf{S}^{COS}$	$\mathbf{S}^{MST}$	$\mathbf{S}^{KNN}$	$\mathbf{S}^{SAL}$
Nodes	145	145	145	145
Edges	20,880	288	438	1,190
Density (%)	100	1.3793	2.0977	5.6992
Flow	5.5558	1.3803	1.7193	0.0069
Diameter	1.4373	7,017.4745	2,418.2135	0.5586
Av. path length	0.8657	2,520.7828	403.3630	0.0620
Clustering	2.3224	0	32.4549	0.0250

that the methodology primarily affects the edges rather than the vertices. The number of edges (Edges) varies significantly across the network types. The COS network is fully connected with 20,880 edges, while the MST network is the sparsest with just 288 edges. The KNN and SAL networks fall in between, with 438 and 1,190 edges, respectively. These differences in the number of edges highlight the varying levels of connectivity between the nodes for each network type. The density (Density %), which reflects the proportion of possible edges in the network, follows a similar trend. The COS network has a density of 100, reflecting its full connectivity. As network types become sparser, the density decreases, with values of 1.38, 2.10, and 5.70 for the MST, KNN, and SAL networks, respectively. The average flow (Flow) values, i.e. the sum of all edge weights, show a significant reduction across the network types. The COS network has the highest flow at 5.56, while the KNN and SAL networks have lower flows of 1.38 and 1.72, respectively. The MST network exhibits the lowest flow at 0.0069, indicating a significant reduction in edge weights for this sparse network. The diameter (Diameter), namely the longest of the shortest path, and the average path length (Av. Path Length), i.e. the average of all the shortest paths, show more variation. The COS network, being fully connected, has the smallest diameter (1.44) and average path length (0.87). As networks become sparser, the diameter increases dramatically, with the MST network having a diameter of over 2,000 and the KNN network reaching 403. The average path length follows a similar pattern, growing from 0.87 for the COS network to 2,520 for the MST network. The clustering coefficient (Clustering), i.e. the number of close triangle in the network, also varies significantly. While the MST network has a coefficient of 0. The KNN and SAL networks show intermediate values 32.45 and 0.025, respectively, indicating a shift in local connectivity as the networks become more sparse.

4.2 Assessing network structures for propagation of cyber risk

The choice of an appropriate network structure is critical when modelling the propagation of cyber risk across countries. Networks serve as conduits for information flow, capturing relationships among entities, and their topology directly influences the accuracy and reliability of propagation results. In the context of assessing country-level cyber risk, the selection and refinement of these networks are paramount to ensure that the modelled propagation reflects realistic interdependencies. To evaluate the efficacy of different network structures as propagation channels, we test how the COS similarity network, the MST network, the KNN graph, and the SAL network function as channel of propagation. For validation, we leverage a subset of countries with known attack severity distributions. To evaluate the robustness of the networks, we conduct an experiment in which 20% of the country-associated distributions are randomly removed. The remaining distributions are then propagated through the network to infer the missing ones. This process is repeated 20 times, each time removing a different random set of nodes.

The accuracy of each network structure is quantified by comparing the propagated distributions to the real ones, using the Wasserstein distance as a measure of similarity. This evaluation framework allows us to systematically assess which network structure captures the underlying relationships most effectively and facilitates accurate propagation of cyber risk. By analysing the performance of different network structures, we aim to identify the most suitable network topology for modelling cyber risk propagation, thereby enhancing the reliability of out-of-sample investigations in subsequent stages.

The boxplot reported in Figure 4 illustrates the Wasserstein distances between the propagated distributions and the real ones for different network structures: COS, MST, KNN, and SAL. The boxes represent the interquartile ranges (IQRs), with medians highlighted as horizontal lines, while the circles show individual data points. The results indicate that the COS network achieves the smallest median Wasserstein distance, demonstrating the most accurate propagation performance among the tested networks. The SAL network follows closely, with a slightly larger median error, but still a relatively compact IQR, suggesting consistent performance. The KNN network exhibits a larger spread and median error, while the MST network performs the least effectively, with the highest median error and the greatest variability in results. Although the SAL network does not achieve the smallest Wasserstein distance, it offers a compelling trade-off between accuracy and network complexity. With only 1,190 edges compared to the fully connected COS network's 20,880 edges, the SAL network achieves a significantly higher efficiency, making it the most suitable choice for the out-of-sample investigation. Its favourable ratio of error to the number of edges justifies its selection as the propagation channel in subsequent analyses. Results are robust across different percentage of omitted distributions as reported in Appendix A.

To assess the convergence behaviour of the propagation process, we analysed the iterative differences in the propagated distributions across model iterations for each network structure. At each iteration, the log-change of the distributions was computed to quantify the magnitude of differences between successive propagation steps. This approach provides insight into how quickly and effectively each network structure stabilizes the propagation dynamics and reaches convergence. The results presented in Figure 5 illustrate the convergence trajectories for the four types of networks that we employ. The *y*-axis represents the log-change in the distributions, while the *x*-axis corresponds to the number of iterations. The COS network displays the slowest but most consistent convergence pattern, with gradual reductions in log-change over the full iteration range, indicating a steady refinement of the propagated distributions. In contrast, the MST network exhibits the fastest initial convergence, with a sharp decline in log-change within the first 100 iterations, but it stabilizes prematurely, suggesting that its simpler topology limits further propagation dynamics. The KNN and SAL networks converge at intermediate rates, with the SAL network showing slightly more sustained improvements beyond the initial phases compared to KNN. These results highlight important trade-offs among the network structures. While the COS network ensures gradual and consistent refinement, its fully connected nature leads to slower convergence. However, rapid stabilization of the MST network may come at the cost of oversimplified dynamics. The SAL network demonstrates a favourable balance, achieving reasonably fast convergence while maintaining sustained improvements, reinforcing its suitability for efficient yet effective propagation modelling.

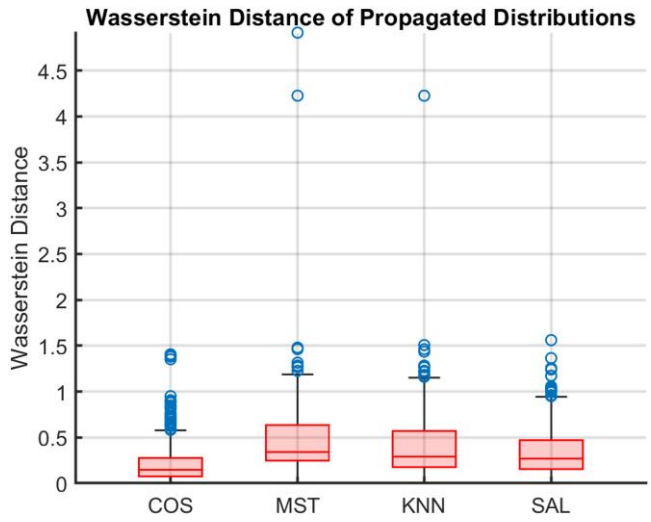

Figure 4 Boxplot of the Wasserstein distances between propagated and real distributions for different network structures: COS, MST, KNN, and SAL.

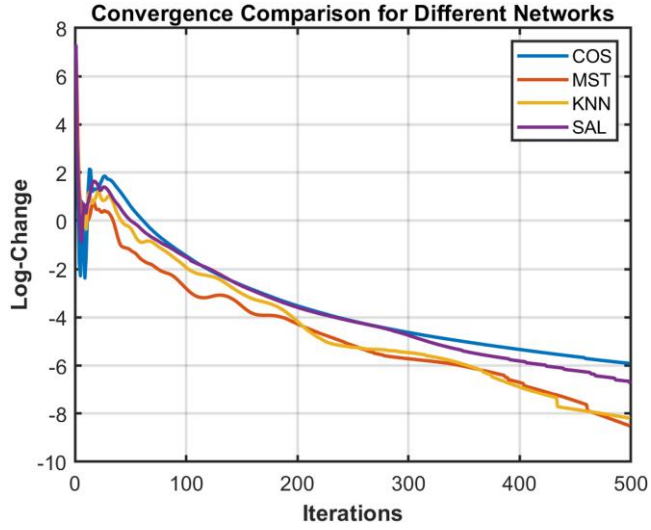

Figure 5 Convergence of the propagation algorithm for different network configuration: COS, MST, KNN, and SAL. The figure reports the log-change of the propagated distributions differences between successive propagation step.

4.3 Out-of-sample application

This analysis represents the out-of-sample application of our proposed methodology for cyber risk evaluation using Wasserstein propagation. Specifically, we focus on a subset of countries for which we lack direct data on cyber risk distributions, including Andorra (AD), Armenia (AM), Azerbaijan (AZ), Djibouti (DJ), Ethiopia (ET), Georgia (GE), Ghana (GH), India (IN), Iraq (IQ), Kazakhstan (KZ), Liberia (LR), Malta (MT), Marshall Islands (MH), Mongolia (MN), Montenegro (ME), Nauru (NR), Sudan (SD), Syria (SY), Togo (TG), Uganda (UG), Vanuatu (VU), and Zambia (ZM). These countries are critical for understanding the broader global cyber risk landscape and the propagation of cyber threats across regions.

These countries are of particular relevance for several reasons. Many are located in regions that are experiencing rapid digital transformation or are critical players in regional geopolitics. Countries such as IN, IQ, and ET hold significant strategic value, influencing both regional stability and international trade. The increasing integration of digital infrastructure in these regions amplifies their vulnerability to cyber attacks, with potential ripple effects extending beyond their borders. Therefore, understanding the potential cyber risk in these nations is crucial for mitigating global cybersecurity threats. In addition to their geopolitical importance, some of these countries, like GH, UG, and ZM, represent emerging markets where digital adoption is accelerating. These regions are often more exposed to cyber risks due to the rapid adoption of technologies without the same level of investment in cybersecurity infrastructure seen in more developed nations. As digital penetration increases in these countries, they become more attractive targets for cybercriminals, and understanding their cyber risk profiles can help anticipate and mitigate potential threats. Furthermore, certain countries on this list, such as SY and SD, face additional risks due to political instability and conflict, factors that can significantly exacerbate vulnerability to cyber attacks. These conditions often result in underdeveloped cybersecurity frameworks, leaving critical infrastructure more exposed to malicious actors. The cyber risk in these regions could have profound geopolitical consequences, underlining the importance of including them in global cyber risk assessments. In this scenario, propagating cyber risk data for these countries is not only necessary to fill gaps in the existing risk landscape but also allows for a more accurate representation of the global cyber risk distribution. By applying our methodology to countries with incomplete data, we are able to simulate potential risks and forecast their impact on neighbouring and interconnected countries. This is essential for policy makers and international organizations to understand where vulnerabilities lie, especially in countries that are rapidly digitalizing or facing geopolitical instability. Ultimately, the relevance of this scenario lies in its ability to identify potential cyber risk in regions that have not been sufficiently studied, enabling a more comprehensive and proactive approach to global cybersecurity. Through Wasserstein propagation, we can better assess the dynamics of cyber risk propagation across countries, providing a powerful tool for informing both national and international cybersecurity policies. Figure 6 illustrates the distribution of the severity level of cyberattacks across countries for which direct data on cyber risk is not available. This analysis employs the proposed methodology based on Wasserstein propagation to estimate risk distributions, categorizing cyber risk into four levels: low, medium, high, and critical. Each country's

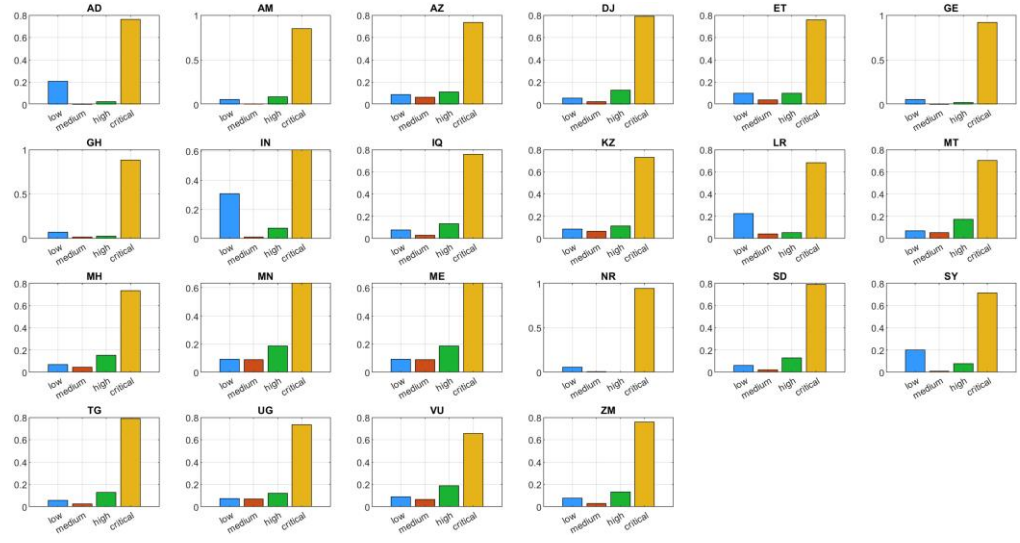

Figure 6 Cyber attack risk distribution across countries lacking direct data, categorized into four risk levels: low, medium, high, and critical. The results are derived using the proposed Wasserstein propagation methodology for cyber risk identification and the SAL network.

results are presented as individual bar charts, allowing for a comparative analysis of their respective risk profiles. A common trend across the countries is the predominance of the critical risk category, represented by the yellow bars. This category consistently accounts for the largest proportion of cyber attacks in most regions, indicating a significant vulnerability to severe cyber threats. Countries such as AD, DJ, ET, MT, NR, and SD demonstrate particularly high levels of critical risk, underscoring the severity of cyber threats in these areas. The high and medium risk categories (green and red bars) are generally minimal across most of the analysed countries. The low-risk category (blue bars) consistently accounts for the second largest proportion of threats in some particular countries such as IN, indicating a bimodal behaviour of cyber attacks distributions. Regionally, the results highlight distinct patterns linked to digital development and geopolitical factors. Emerging markets such as GH, UG, and ZM show particularly high proportions of critical risk, reflecting vulnerabilities associated with rapid digitalization and limited cybersecurity infrastructure. Similarly, countries experiencing political instability and conflict, such as SY and SD, exhibit elevated levels of critical risk, likely exacerbated by underdeveloped cybersecurity frameworks and increased exposure to malicious actors. Strategic geopolitical regions, including IN and IQ, display more diverse risk distributions, with notable proportions in the medium- and low-risk categories in addition to critical risks. This diversity highlights the complex cyber risk dynamics in countries that are both economically and geopolitically significant.

These findings underscore the uneven distribution of cyber risk across the analysed countries and are robust across different network specifications, as shown in Appendix B. The predominance of critical risks in many regions emphasizes the need for targeted cybersecurity interventions, particularly in countries undergoing rapid digital transformation or those affected by geopolitical instability. By extending cyber risk assessments to understudied regions, this analysis provides valuable insights that are essential for informing global cybersecurity policies. The results demonstrate the utility of the proposed methodology in filling data gaps and enabling a comprehensive understanding of global cyber risk distribution. Appendix D presents a simulation study designed to disentangle the effects of the propagation algorithm, network topology, and empirical data structure on the inferred risk distributions.

4.4 The role of node centrality in cascade dynamics

Node centrality is a fundamental concept in network science, capturing the importance of nodes in facilitating or mitigating cascading effects. In the context of cyber attack propagation, centrality plays a pivotal role in understanding how disruptions can spread through interconnected systems. Central nodes, characterized by their strategic positions in the network, either amplify or dampen the flow of shocks, making them critical targets for intervention. Countries with high centrality often serve as hubs or conduits, influencing the extent and speed of disruption propagation. As a result, identifying these nodes provides key insights into prioritizing cybersecurity policies and designing effective mitigation strategies.

In this scenario, we analyse the SAL network, where each country node is associated with a distribution of cyber attack severity levels. Previously, we employed a Wasserstein propagation technique to estimate these distributions for countries lacking empirical data. Building on this, we aim to study how centrality metrics, particularly the eigenvector centrality, influence the dynamics of shock propagation. Eigenvector centrality is a metric that quantifies the influence of a node in a network, considering not only the number of connections (degree) a node has but also the importance of the nodes to which it is connected. It is defined using the principal eigenvector of the network's adjacency matrix. The centrality score for a node is proportional to the sum of the centrality scores of its neighbours. Mathematically, the eigenvector centrality $\mathbf{e} = [e_1, e_2, \ldots, e_n]^\top$ satisfies:

$$\mathbf{S}^{Sal}\mathbf{e} = \lambda_{\max}\mathbf{e},$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{S}^{\mathrm{SAL}}$, and $\mathbf{e}$ is the corresponding eigenvector. The centrality score of node i is given by e_i, and it reflects how well-connected the node is, taking into account the importance of its neighbours.

Figure 7 show the normalized eigenvector centrality displayed on the map, indicating the relative importance of each country within the network. Countries with higher centrality are highlighted in yellow, while those with lower centrality are coloured dark blue. Grey areas represent countries for which data are not available. The visualization reveals distinct geographic patterns in eigenvector centrality, with regions of high influence and connectivity clearly concentrated in specific parts of the world. In particular, countries in North America, Western Europe, and Australia are prominently shaded in yellow, indicating their dominant positions in the network. In contrast, regions such as Sub-Saharan Africa and parts of South Asia exhibit a lower centrality, as evidenced by the prevalence of blue shades. These patterns suggest a clear disparity in connectivity and influence, with certain regions serving as central hubs, while others remain more peripheral in the network structure.

To explore the interplay between centrality and shock propagation, we propose a simulation framework to model the cascading effects of disruptions within the network. The framework incorporates the following steps. The algorithm begins by selecting a node with the highest (or lowest) centrality as the initial shock source. The shock modifies the node's risk profile, transforming its distribution of cyber attack severity into a uniform distribution, reflecting a complete redistribution of risk. At each propagation step, the algorithm identifies neighbouring nodes connected to the currently shocked nodes. The neighbours of a node i are given by:

$$N(i) = \{k \in V \mid s_{i,k}^{\mathrm{SAL}} > 0\},$$

where V is the set of all nodes, and $s_{i,k}^{\mathrm{SAL}}$ represents the strength or proximity of the connection between nodes i and k. The shock then propagates to these neighbours based on their relationship with the shocked nodes. The shock distribution of a newly affected node k is updated using the Wasserstein barycentre methodology. This approach computes the new distribution of the shocked node as a weighted average of the distributions of its neighbours, where at least one of them has received a shock in the previous iteration. The propagation process halts when no further neighbours remain unshocked. This level-wise diffusion ensures that the cascade evolves in a structured manner.

To quantify the impact of the propagation, we compute the Wasserstein distance between the original and updated distribution for each newly shocked node. The shock magnitude for a given step is the cumulative Wasserstein distance across all newly shocked nodes. In addition, we

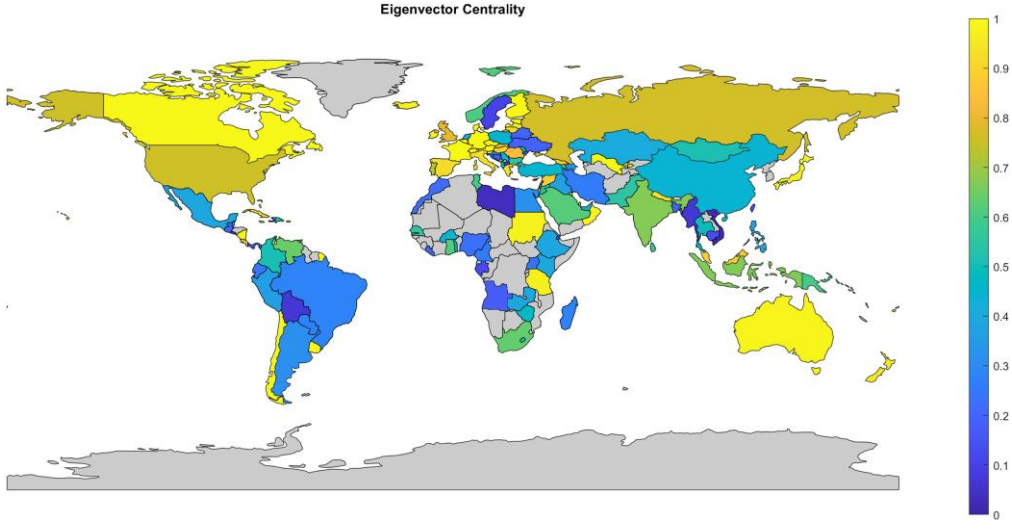

Figure 7 Normalized eigenvector centrality. The figure reports on the map the normalized eigenvector centrality of each country. Lighter shades indicate higher centrality, darker shades indicate lower centrality, and grey areas denote countries with unavailable data.

track the cumulative number of shocked nodes, providing insight into the extent of the cascade over time.

Results of the shock cascade dynamics are illustrated in Figure 8. The analysis highlights distinct propagation behaviours depending on the centrality of the target node, underlining the critical influence of network topology on shock transmission. When the target node is the one with the highest centrality (left panels), the propagation follows a multiphase pattern. Initially, there is a pronounced primary cascade and subsequent phases with diminishing effects following by a last wave of contagion near iteration 17. This extended propagation timeline results in a cumulative shock spread affecting approximately 70 nodes and a total Wasserstein distance change of 35. Such dynamics underscore the ability of highly central nodes to drive widespread influence through hierarchical relationships in the network. In contrast, targeting a low-centrality node (right panels) leads to delayed and subdued cascade effects. The propagation terminates at Step 2, affecting only three nodes with a cumulative Wasserstein distance of 3.5. The limited impact emphasizes the localized nature of shocks initiated from peripheral nodes, which lack substantial connections to propagate extensively. The stark contrast between these scenarios demonstrates the hierarchical structure of the network, where highly central nodes act as critical conduits for shock propagation. This finding suggests that mitigation strategies should prioritize high central nodes such as North America and European countries along with Australia, to effectively control systemic risk.

The results of the cascade dynamics analysis for other types of networks are reported in Appendix C. The findings reveal distinct spreading dynamics compared to the SAL network. In all cases, the propagation completes within only a few iterations, regardless of the targeted node. In the COS network, the propagation concludes in just two steps, with all nodes being affected. This rapid spread is attributed to the fully connected topology of the similarity network, which facilitates the shock transmission across the entire network almost instantly. For the MST and KNN networks, the number of propagation steps is similarly limited, involving only a few iterations. However, the key difference lies in the number of nodes affected by the end of the process. Unlike the COS network, both MST and KNN networks

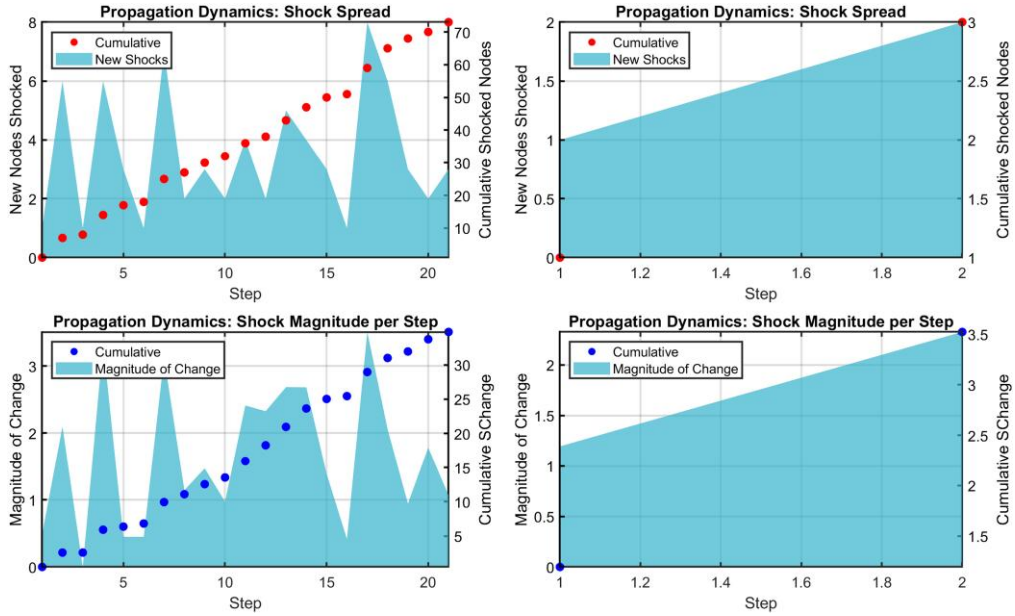

Figure 8 Propagation dynamics of shocks in the cyber attack distribution. Top left show propagation when targeting the highest centrality node; right panels represent targeting a low centrality node. Top panels illustrate shock spread dynamics: shaded areas indicate newly affected nodes at each step, while dots show cumulative affected nodes. Bottom panels display shock magnitude measured by Wasserstein distance: shaded areas represent the magnitude of distribution changes for newly affected nodes at each step, and dots show cumulative impact.

exhibit fewer affected nodes at the final step. This outcome is a direct result of their sparse connectivity, which restricts the pathways available for shock propagation.

We use eigenvector centrality because it captures global influence in the network by accounting for the importance of a node's neighbours. However, to address the concern that eigenvector centrality may inflate propagation in clustered regions, we complement our analysis with closeness centrality. The discussion and results are reported in Appendix E.

5 Conclusion

This study introduces a novel methodology for modelling the propagation of cyber risk across interconnected economic networks using Wasserstein propagation, a framework rooted in optimal transport theory. By leveraging a global dataset of cyber attacks and macroeconomic indicators, we demonstrated the ability of our procedure to estimate cyber attack severity distributions for countries lacking direct data, thus contributing to reducing uncertainty related to cyberspace and cyber attacks. This approach addresses critical gaps in understanding cyber risk distribution, offering actionable insights for policymakers and stakeholders.

Our findings highlight the effectiveness of the SAL network as a robust and efficient structure for propagating cyber risk, balancing accuracy and computational complexity. In addition, the role of central nodes in cascade dynamics emphasizes the importance of targeted interventions to mitigate systemic risks.

The methodology provides a scalable framework applicable to diverse contexts where risk propagation is a concern. By offering probabilistic assessments of cyber risk, even in data-sparse regions, it supports evidence-based decision-making and promotes resilience in global cybersecurity.

In the present work, we focus attention to severity distributions, as they provide direct information on tail risk and are less sensitive to heterogeneity in reporting across countries. As a possible extension, we could work with absolute frequency distributions of attacks that capture a different and complementary dimension of cyber risk, namely persistence/exposure. Working with absolute frequencies rather than empirical probability distributions would naturally lead to an unbalanced optimal transport formulation, allowing us to propagate both severity and exposure jointly. This is a recent extension of the Wasserstein framework that allows measures with different total masses to be compared (Chizat et al., 2018; Gallouët et al., 2021; Séjourné et al., 2021).

Future research will aim to advance this framework along several directions. First, additional data sources (i.e. socioeconomic, political, and governance indicators) will be integrated to enrich the analysis and enhance the robustness of risk assessments. Second, we plan to implement dynamic network structures, which would allow us to capture the temporal evolution of interdependencies and provide more realistic representations of systemic risks, also improving predictive capability also out of time. Third, we aim at extending the present work at sectoral-level cyber incidents once such data will become available. Sectoral granularity would allow us not only to make a global evaluation, at national level, but to go deeper to more complex information and capturing sector-specific interdependencies (e.g. financial linkages across countries) and interlayer connections representing cross-sectoral dependencies (e.g. energy and telecommunications) that would facilitate both the modelling of propagation of cyber risk across borders, and the way in which risks cascade across industries within and between countries.

Acknowledgments

We thank the Associate Editor and the referee for their valuable comments and constructive suggestions, which helped to improve the manuscript. We are also grateful to the experts at Hackmanac for providing access to the dataset used in this study.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

The authors acknowledge the support of the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance (acronym: DIGITAL, project no. 101119635). This work has been partially supported by the Italian Ministry of University and Research (MUR) through the project 'A geo-localized data framework for managing climate risks and designing policies to support sustainable investments' (No. 20229CWYXC) within the PRIN 2022 program, funded by the European Union-Next Generation EU. This work has also been partially supported by PRIN 2022 - CUP: E53C24002270006. The fifth author acknowledges financial support by a grant of the Ministry of Research, Innovation and Digitization in Romania, CNCS - UEFISCDI, project number PN-IV-P2-2.1-TE-2023-1317, within PNCDI IV. Additional support was provided by the University of Milan through the Early Career Development 2025 project CRiSP. The second and the fourth authors acknowledge financial support within the 'Fund for Departments of Excellence academic funding' provided by the Ministero dell'Università e della Ricerca to his department (CUP: G43C22004550001). The publication was also produced with funding from the Italian Ministry of University and Research under the Call for Proposals related to the scrolling of the final rankings of the PRIN 2022 call. Project 'Latent variable models and dimensionality reduction methods for complex data' (PI Prof. Paolo Giordani), Project No. 20224CRB9E, CUP C53C24000730006. Grant Assignment Decree No. 1401 adopted on 18.9.2024 by MUR.

Data availability

The data used in this study are available from the authors upon reasonable request.

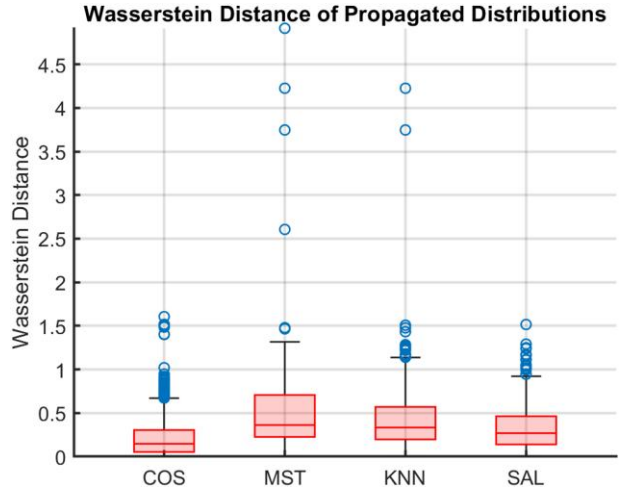

Figure A1 Boxplot of the Wasserstein distances between propagated and real distributions for different network structures: COS, MST, KNN, and SAL. In this experiment, we have removed the 30% of the distributions.

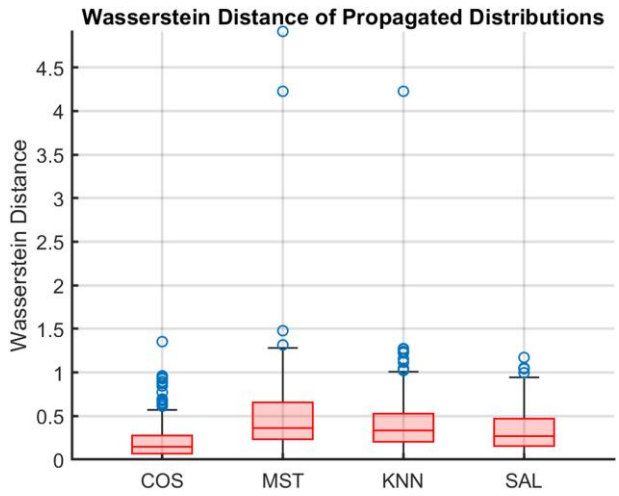

Figure A2 Boxplot of the Wasserstein distances between propagated and real distributions for different network structures: COS, MST, KNN, and SAL. In this experiment, we have removed the 10% of the distributions.

Appendix A. Robustness analysis

We report results on the sensitivity analysis related to the choice of the SAL network for propagation scenarios. In the manuscript, we performed an experiment in which 20% of the distributions associated with these countries are randomly removed for 20 iterations. Here, we replicate the experiment with different thresholds. We remove 30% of the distributions in the first scenario and 10% in the second. As Figures A1 and A2 suggest, the COS similarity network consistently achieves the smallest median Wasserstein distance, followed by the SAL graph, the KNN and the MST.

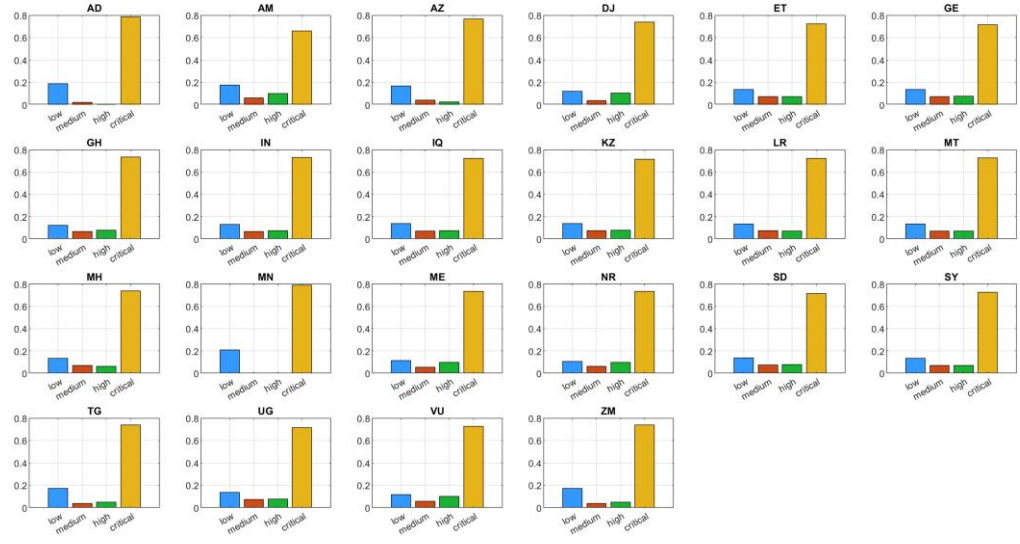

Figure B1 Cyber attack risk distribution across countries lacking direct data, categorized into four risk levels: low, medium, high, and critical. The results are derived using the proposed Wasserstein propagation methodology for cyber risk identification and the COS network.

Appendix B. Out-of-sample: propagation with different networks

To validate the robustness of our cyber risk propagation methodology, we extended our analysis across multiple network topologies beyond the original Wasserstein approach. We systematically investigated the risk distribution using COS (Figure B1), MST (Figure B2), and KNN (Figure B3) networks,

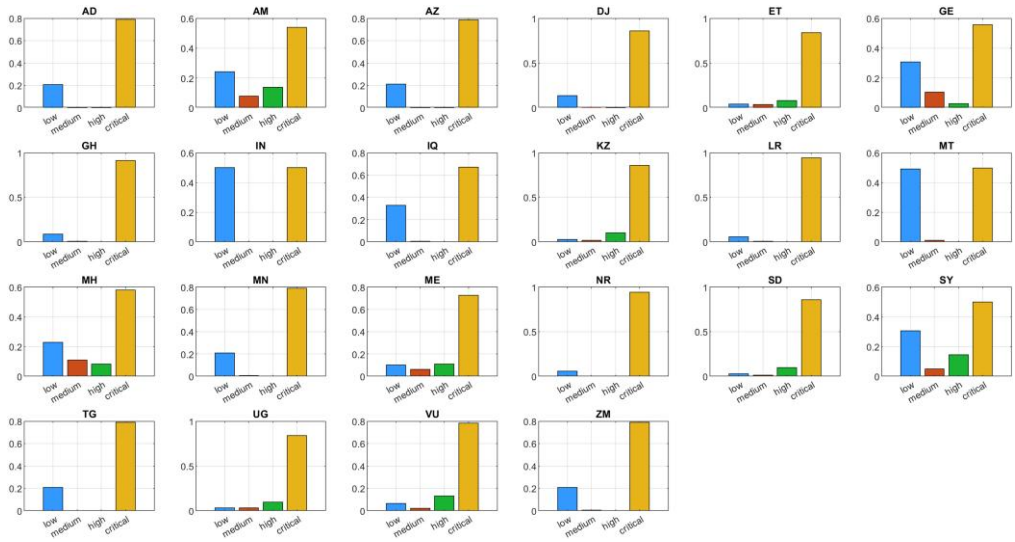

Figure B2 Cyber attack risk distribution across countries lacking direct data, categorized into four risk levels: low, medium, high, and critical. The results are derived using the proposed Wasserstein propagation methodology for cyber risk identification and the MST network.

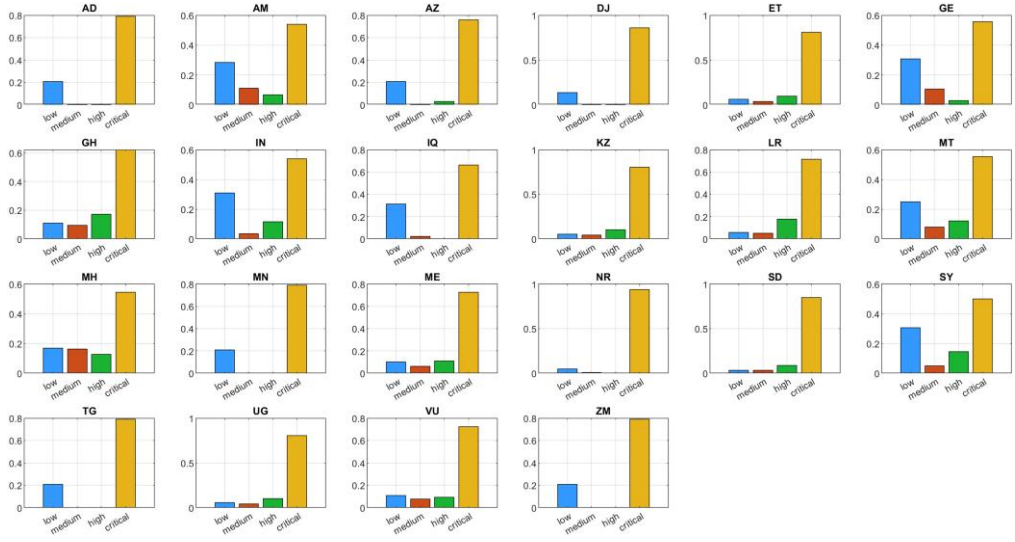

Figure B3 Cyber attack risk distribution across countries lacking direct data, categorized into four risk levels: low, medium, high, and critical. The results are derived using the proposed Wasserstein propagation methodology for cyber risk identification and the KNN network.

revealing consistent patterns that reinforce the reliability of our initial findings. Across these diverse network specifications, we observe a remarkably consistent prevalence of critical risk levels, which underscores the structural vulnerability of the studied countries regardless of the underlying network connectivity model.

The experiments with alternative network topologies demonstrate that our methodology is not merely contingent on a single network representation, but captures fundamental cyber risk dynamics. The COS network, which measures the angular similarity between risk vectors, replicated the high proportion of critical risks, particularly in regions experiencing rapid digital transformation and geopolitical complexity. Similarly, the MST network, which captures the most essential connectivity relationships, and the KNN network, which identifies local risk neighbourhoods, both confirmed the initial observations of critical risk dominance.

In particular, the bimodal distribution pattern, characterized by significant proportions of critical and low-risk categories, remains evident across these network specifications. This persistent bimodality, especially observable in countries like India, suggests an inherent complexity in cyber risk landscapes that transcends specific network modelling approaches. This consistency provides strong empirical support for the proposed Wasserstein propagation methodology and enhances confidence in our ability to assess cyber risk in data-sparse regions.

These comprehensive network topology experiments not only validate our initial findings but also demonstrate the adaptability and reliability of our approach in capturing nuanced cyber risk distributions. By showing robustness across different network representations, we provide a more comprehensive and trustworthy framework for understanding cyber risk propagation in geopolitically diverse and digitally evolving environments.

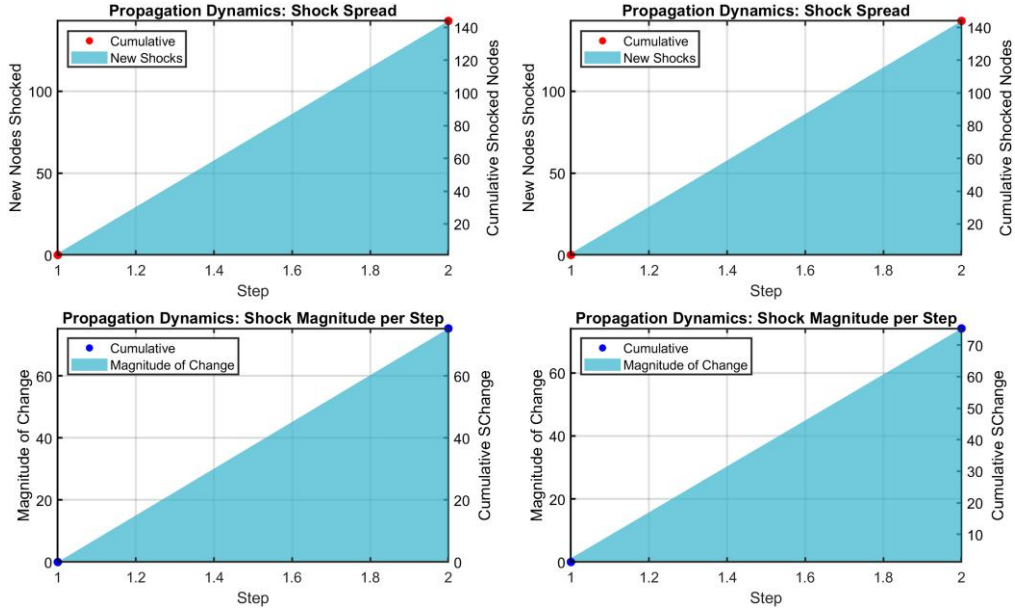

Figure C1 Propagation dynamics of shocks in the cyber attack distribution over the COS network. Top left show propagation when targeting the highest centrality node; right panels represent targeting a low centrality node. Top panels illustrate shock spread dynamics: shade areas indicate newly affected nodes at each step, while dots show cumulative affected nodes. Bottom panels display shock magnitude measured by Wasserstein distance: shade areas represent the magnitude of distribution changes for newly affected nodes at each step, and dots show cumulative impact.

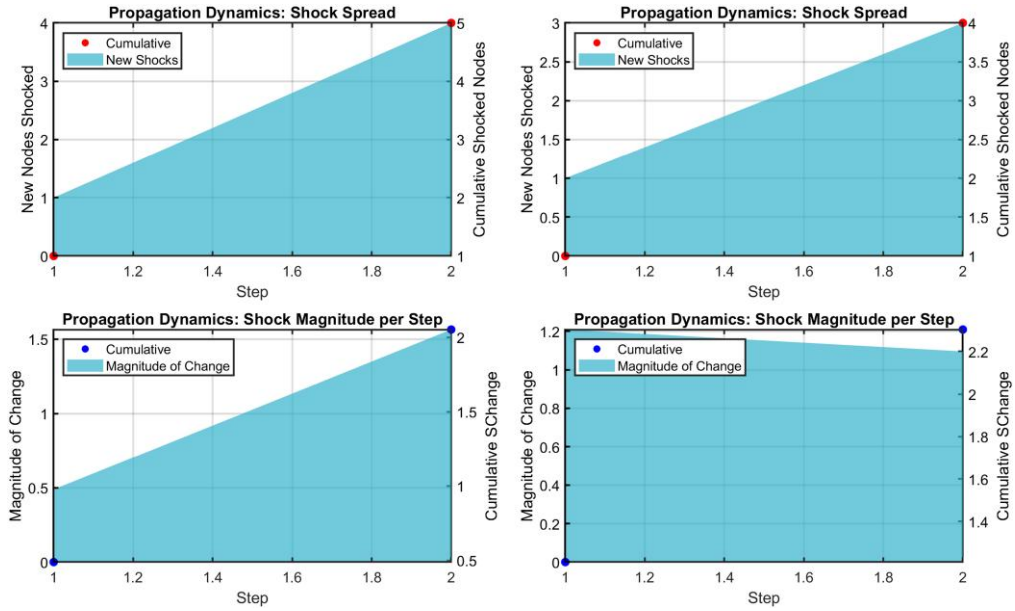

Figure C2 Propagation dynamics of shocks in the cyber attack distribution over the MST network. Top left show propagation when targeting the highest centrality node; right panels represent targeting a low centrality node. Top panels illustrate shock spread dynamics: shaded areas indicate newly affected nodes at each step, while dots show cumulative affected nodes. Bottom panels display shock magnitude measured by Wasserstein distance: shade areas represent the magnitude of distribution changes for newly affected nodes at each step, and dots show cumulative impact.

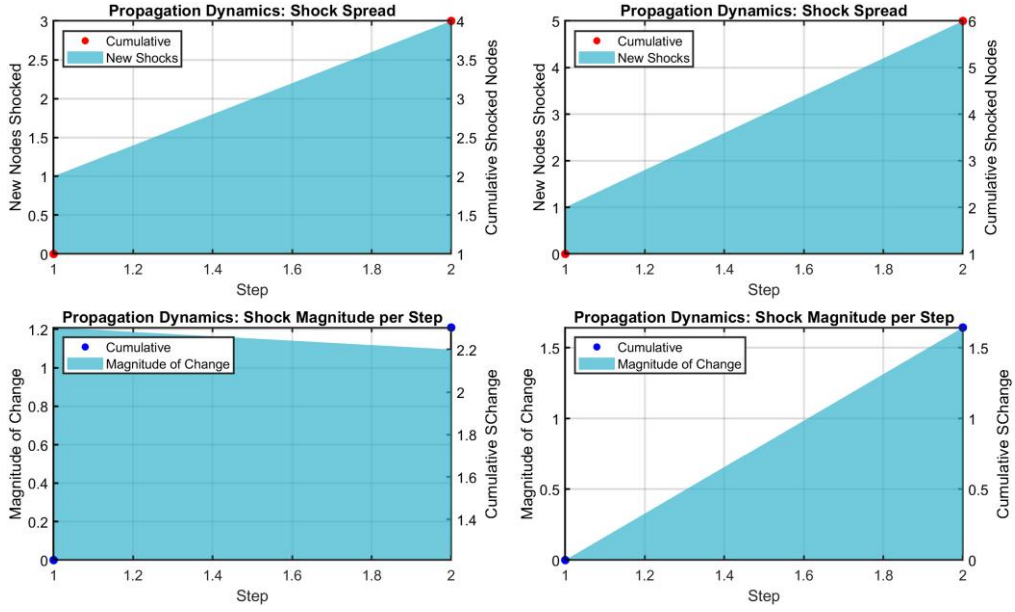

Figure C3 Propagation dynamics of shocks in the cyber attack distribution over the KNN network. Top left show propagation when targeting the highest centrality node; right panels represent targeting a low centrality node. Top panels illustrate shock spread dynamics: shaded areas indicate newly affected nodes at each step, while dots show cumulative affected nodes. Bottom panels display shock magnitude measured by Wasserstein distance: shade areas represent the magnitude of distribution changes for newly affected nodes at each step, and dots show cumulative impact.

Appendix C. Cascade dynamics with different networks

We examine the cascade dynamics across different types of networks, highlighting significant differences in spreading behaviour compared to the SAL network. A common observation across all networks is that the propagation process concludes in a few iterations, regardless of the targeted node. However, the scale and pattern of propagation vary significantly depending on the network topology.

In the COS network (Figure C1), the cascade is completed in just two steps, with all nodes affected. This rapid and comprehensive spread is a direct consequence of the fully connected structure of the network, which provides numerous pathways for shock transmission. Such a topology ensures that shocks quickly reach all nodes, leaving little room for gradual or localized propagation.

In contrast, the MST (Figure C2) and KNN (Figure C3) networks exhibit a more constrained propagation process. While the number of propagation steps remains similarly limited to a few iterations, the total number of affected nodes at the end of the cascade is significantly lower. This outcome reflects the sparse connectivity of these networks, which limits the availability of pathways for shock transmission. Specifically, the MST, being a tree structure, restricts propagation to a single unique path between nodes, while the KNN network confines connections to local neighbourhoods, further curbing the cascade's reach.

These results underscore the critical role of network topology in shaping shock propagation dynamics. Fully connected networks, such as those based on COS similarity, enable rapid and widespread cascades, potentially amplifying systemic risks. Conversely, sparsely connected networks like the MST and KNN act as natural buffers, containing the spread of shocks due to their limited connectivity. This distinction has important implications for network design and risk management. For example, in scenarios where rapid information or influence dissemination is desired, highly connected networks may be preferable. However, when the goal is to contain systemic risks, sparsely connected structures offer more controlled propagation pathways, reducing the likelihood of large-scale cascades.

Appendix D. Simulation analysis of risk propagation patterns

Figure 6 shows that many countries without direct cyber attack data are inferred to have distributions concentrated in the critical risk category. Since this outcome may appear unintuitive or overly skewed, a simulation study was conducted to clarify whether it is driven by (i) properties of the Wasserstein propagation algorithm, (ii) the structure of the network, or (iii) the empirical distribution of the observed data.

This experiment analyses the impact of data characteristics and network structure on inference accuracy. Specifically, we contrast the proposed optimal transport approach with a baseline method based on local Euclidean averaging.

We simulate a network composed of two distinct groups of nodes, V_1 and V_2, with each node associated with a univariate probability distribution. These distributions are generated by sampling from log-normal models of the form: $X \sim \text{LogNormal}(\mu, \sigma)$, where μ and σ control the central tendency and dispersion. Nodes in group V_1 are assigned a fixed parameter set ($\mu_1 = 1$, $\sigma_1 = 0.2$), while nodes in V_2 are assigned parameters from a grid of values: $\mu_2 \in \{4, 3, 2, 1\}$ and $\sigma_2 \in \{2, 1, 0.5, 0.2\}$. This setup creates varying degrees of shape divergence between the two groups. Each distribution is estimated by drawing 1000 samples and applying kernel density smoothing over 25 bins to produce empirical distributions β_v.

To introduce realistic network complexity, we generate binary undirected graphs with controlled modularity using a variant of the stochastic block model. This simplification helps isolate and analyse the key dynamics of the propagation process, avoiding the additional complexity of weight heterogeneity, which could obscure the system's fundamental behaviour. Each node is randomly assigned to one of two communities. Edges are established based on intra- and intercommunity connection probabilities:

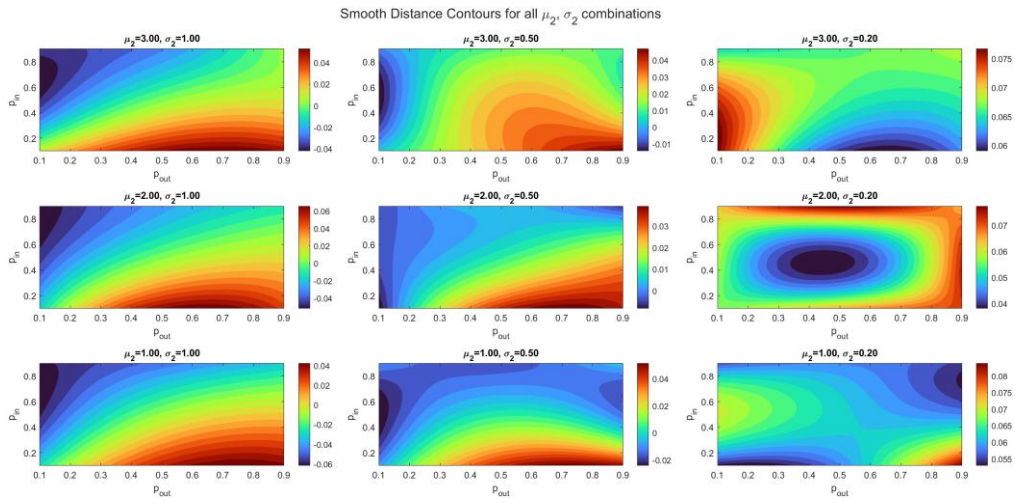

Figure D1 Relative inference performance of Wasserstein propagation versus a Euclidean baseline across synthetic scenarios. Each subplot represents a distinct combination of log-normal parameters (μ_2, σ_2) defining the shape of distributions in the second community, and network modularity settings (p_{in}, p_{out}). The values shown reflect the average difference in Wasserstein distances between the true and estimated distributions obtained via our method and the baseline. Negative values indicate improved performance of the Wasserstein propagation approach.

$$\mathbb{P}(S_{v,w}=1)=\begin{cases} p_{\text{in}}, & \text{if } c_v=c_w,\\ p_{\text{out}}, & \text{if } c_v\neq c_w,\end{cases} \quad \text{with } S_{w,v}=S_{v,w},$$

where S is the binary adjacency matrix of the network. We vary p_{in} and p_{out} over {0.9, 0.7, 0.5, 0.3, 0.1}, enabling systematic exploration of network modularity from highly clustered to nearly random structures. To ensure network sparsity, which better reflects empirical conditions, we impose a cap on the number of edges by selecting only a proportion 20% of the possible $E_{\max}=\frac{N(N-1)}{2}$ edges, uniformly removing any excess after generation.

This controlled environment allows us to test how well the Wasserstein propagation method reconstructs unobserved distributions when informed by only a partial subset of the network. In each simulation, 80% of nodes in group V_2 are treated as known (fixed) while the remaining nodes (including 20% of V_1 and V_2) are set as targets for inference. For comparison, we include a local Euclidean baseline in which the inferred distribution at an unobserved node is computed as a convex combination of its observed neighbours:

$$\hat{\beta}_v^{\text{Euclid}}=\sum_{w\in V(v)\cap V_{\text{fixed}}}\tilde{S}_{v,w}\beta_w, \quad \tilde{S}_{v,w}=\frac{S_{v,w}}{\sum_{w\in V_{\text{fixed}}}S_{v,w}},$$

where $V(v)$ denotes the neighbourhood of node v and the weights $\tilde{S}_{v,w}$ are normalized.

We assess performance by computing the regularized Wasserstein distance between the inferred and true distributions for each method. Each unique configuration of distribution parameters and network structure is replicated $N_{\text{rep}}=100$ times. For every repetition, the average Wasserstein error is recorded across all unobserved nodes, providing a robust basis for performance comparison.

Figure D1 presents a comparative evaluation of the Wasserstein propagation method relative to a Euclidean benchmark under a broad set of synthetic conditions. Each subplot corresponds to a unique scenario defined by the distributional parameters (μ_2, σ_2) for the second community and the network's structural properties (p_{in}, p_{out}), which govern intra- and intercommunity connectivity. For each configuration, we compute the average Wasserstein distance between the inferred and true distributions over all unobserved nodes and simulation repetitions. The plotted values reflect the difference in inference error between our proposed method and the Euclidean approach. Negative values

indicate cases where Wasserstein propagation achieves lower reconstruction error. The figure shows that Wasserstein propagation delivers consistently superior results in settings characterized by pronounced distributional heterogeneity (larger μ_2, σ_2) and strong community structure (high p_{in}, low p_{out}). Conversely, its advantage narrows in scenarios with more homogeneous distributions or weaker network modularity. These results emphasize the algorithm's effectiveness in capturing both geometric and structural nuances in the data, particularly when the underlying system displays complex interdependencies and non-Euclidean patterns.

The results of the simulation studies confirm that the Wasserstein propagation algorithm is not inherently biased towards high-risk outcomes. In scenarios where the true distributions are concentrated in low or medium severity, the method accurately reconstructs those profiles. The skew seen in Figure 6 instead arises from the empirical structure of the input data: most observed countries in the 2023 dataset already exhibit high or critical attack severity, and the unobserved countries are closely connected to them in the similarity network. In such settings, the propagation process faithfully reflects the concentration of risk in the observed nodes. The network's modularity further reinforces this behaviour by amplifying signals within connected clusters. In conclusion, the inference results in Figure 6 are not artefacts of the methodology but data-driven consequences of propagating distributions over a high-risk, tightly connected global structure. The simulations demonstrate that the method remains robust and sensitive to genuine variation when present.

Appendix E. Closeness centrality measure

In Section 4.4, we used eigenvector centrality to identify influential nodes because of its recursive property, which highlights globally connected hubs. Nevertheless, eigenvector centrality may overemphasize nodes embedded in dense clusters. To assess robustness, we complement our analysis with closeness centrality, which captures a different dimension of structural importance: proximity and reachability.

While eigenvector centrality identifies globally influential nodes, i.e. those connected to other well-connected nodes, closeness centrality captures a different aspect of structural importance: proximity. Specifically, closeness measures how close a node is, on average, to all other nodes in the network, making it a natural indicator of reachability and information accessibility. The map in Figure E1 shows that countries with high closeness centrality (depicted in yellow) are primarily located in Europe, North America, and parts of Central Asia. These regions are structurally well-positioned to quickly interact with the rest of the network. In contrast, many African and some South American countries appear in cooler colours, indicating longer average distances to other countries and therefore lower centrality under this metric.

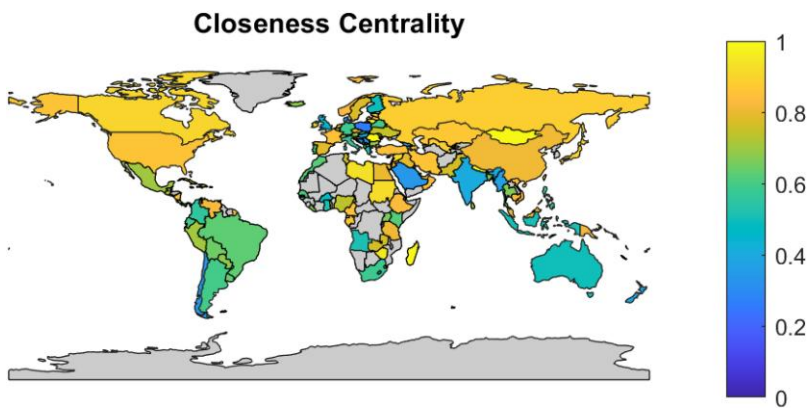

Figure E1 Normalized closeness centrality. The figure reports on the map the normalized eigenvector centrality of each country. Lighter shades indicate higher centrality, darker shades indicate lower centrality, and grey areas denote countries with unavailable data.

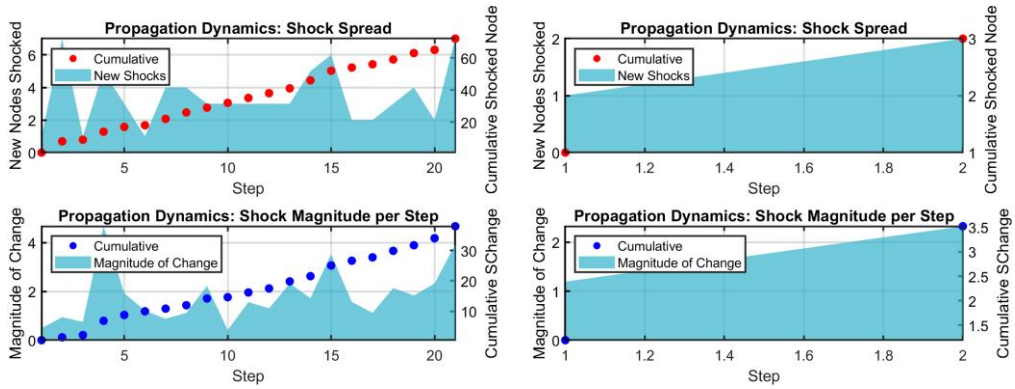

Figure E2 Propagation dynamics of shocks in the cyber attack distribution. Top left show propagation when targeting the highest centrality node; right panels represent targeting a low centrality node. Top panels illustrate shock spread dynamics: shaded areas indicate newly affected nodes at each step, while dots show cumulative affected nodes. Bottom panels display shock magnitude measured by Wasserstein distance: shaded areas represent the magnitude of distribution changes for newly affected nodes at each step, and dots show cumulative impact.

Comparing this to the eigenvector centrality results in the article, we observe both overlap and divergence. High eigenvector centrality countries tend to be globally connected hubs, which often coincides with high closeness (e.g. the U.S., Germany, and France). However, eigenvector centrality can overemphasize nodes in dense clusters, inflating their perceived importance due to mutual reinforcement, whereas closeness provides a more balanced view of global accessibility, unaffected by localized concentration of influence. While eigenvector centrality is suitable for modelling influence diffusion through recursive importance, closeness centrality highlights efficiency in information flow, making it particularly relevant for modelling the speed of propagation.

The shock propagation results obtained by using closeness centrality as the targeting criterion closely mirror the dynamics observed when shocks originate from nodes with high eigenvector centrality (see Figure E2 left panel). Indeed, we observe a steady increase in the cumulative number of shocked nodes and shock magnitude, alongside intermittent waves of new shocks per step. The progression is nonlinear, with some periods of plateau followed by bursts, indicating that even under closeness-based targeting, the network retains threshold and bottleneck dynamics typical of influence diffusion processes. By contrast, initiating the shock from a low-centrality node (right panels) results in a delayed and minimal cascade. This limited diffusion highlights the confined influence of peripheral nodes, which, due to their sparse connectivity, are unable to trigger widespread transmission through the network. These patterns validate that closeness centrality, despite capturing a different aspect of network position than eigenvector centrality, also identifies structurally influential nodes, capable of initiating broad cascades. The similarity in results stems from the fact that both high-closeness and high-eigenvector nodes are typically located in core regions of the network, where they are either directly connected to or centrally positioned among many other nodes. In summary, these findings reinforce the robustness of our main result: targeting structurally central nodes, whether defined by global (eigenvector) or semiglobal (closeness) metrics, leads to effective and widespread propagation, while peripheral nodes play a more limited role. The similarity in cascade dynamics under both centrality definitions underscores their practical alignment in this network and supports the general validity of our centrality-based vulnerability analysis.

References

Andreolini M., Apruzzese A., Bechelli L., Butti G., Castelluccio S., Cesarone G., Chiantore L., Cicognini M., Dragoni G., Faggioli G., Gabrielli I., Gatti C., Girdinio P., Giudice P., Giustozzi C., Greco A., Ivaldi L.,

Livelli F. M. R., Livrieri L. N., & Zapparoli Manzoni A. (2004). *Rapporto 2024 sulla Sicurezza ICT in Italia*. In Clusit.

Arjovsky M., Chintala S., & Bottou L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 214–223). PMLR.

Belkin M., Niyogi P., & Sindhwani V. (2004). *Regularization and semi-supervised learning on large graphs* (pp. 624–638). PMLR.

Bengio Y., Delalleau O., & Le Roux N. (2006). *Label propagation and quadratic criterion*. In *Semi-supervised learning* (pp. 193–216). MIT press.

Brantly A. F. (2021). Risk and uncertainty can be analyzed in cyberspace. *Journal of Cybersecurity*, *7*(1), 1–12. https://doi.org/10.1093/cybsec/tyab001

Cardillo G., Giordani P., Levantesi S., Nigri A., & Spelta A. (2023). A multi-way analysis of similarity patterns in longevity improvements. *Statistical Methods & Applications*, *32*(5), 1805–1828. https://doi.org/10.1007/s10260-023-00714-0

Chizat L., Peyré G., Schmitzer B., & Vialard F.-X. (2018). Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, *274*(11), 3090–3123. https://doi.org/10.1016/j.jfa.2018.03.008

Cuturi M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, *26*, 22922300.

Facchinetti S., Osmetti S. A., & Tarantola C. (2023). Network models for cyber attacks evaluation. *Socio-economic Planning Sciences*, *87*(19), 101584. https://doi.org/10.1016/j.seps.2023.101584

Facchinetti S., Osmetti S. A., & Tarantola C. (2024). A statistical approach for assessing cyber risk via ordered response models. *Risk Analysis*, *44*(2), 425–438. https://doi.org/10.1111/risa.14186

Gallouët T., Lavenant H., & Santambrogio F. (2021). Regularized unbalanced optimal transport and Cournot-Nash equilibria. *Calculus of Variations and Partial Differential Equations*, *60*(5), 1–27.

GhasemiGol M., Ghaemi-Bafghi A., & Takabi H. (2016). A comprehensive approach for network attack forecasting. *Computers & Security*, *58*(4), 83–105. https://doi.org/10.1016/j.cose.2015.11.005

Grady D., Thiemann C., & Brockmann D. (2012). Robust classification of salient links in complex networks. *Nature Communications*, *3*(1), 864. https://doi.org/10.1038/ncomms1847

Kaufmann D., Kraay A., & Mastruzzi M. (2010a). The worldwide governance indicators: A summary of methodology, data and analytical issues. *World Bank Policy Research Working Paper*, 5430.

Kaufmann D., Kraay A., & Mastruzzi M. (2010b). *The worldwide governance indicators: Methodology and analytical issues*. Technical report, World Bank Policy Research Working Paper No. 5430.

Kondor R. I., & Lafferty J. (2002). Diffusion kernels on graphs and other discrete structures. In *Proceedings of the 19th International Conference on Machine Learning* (pp. 315–322). ICML.

Kuhn H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1-2), 83–97. https://doi.org/10.1002/nav.v2:1/2

Li J., Shao C., Xu W., & Dong C. (2015). Traffic density estimation from surveillance video using deep learning methods. *IET Intelligent Transport Systems*, *9*(10), 878–887.

Panaretos V. M., & Zemel Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Application*, *6*(1), 405–431. https://doi.org/10.1146/statistics.2019.6.issue-1

Pele O., & Werman M. (2009). Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 460–467). IEEE.

Peyré G., & Cuturi M. (2019). *Computational optimal transport*. Now Publishers.

Rabin J., Peyré G., Delon J., & Bernot M. (2011). *Wasserstein barycenter and its application to texture mixing*. In *Scale space and variational methods in computer vision* (pp. 435–446). Springer.

Rubner Y., Guibas L. J., & Tomasi C. (1997). The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop* (Vol. 661, p. 668).

Schulzke M. (2018). The politics of attributing blame for cyberattacks and the costs of uncertainty. *Perspectives on Politics*, *16*(4), 954–968. https://doi.org/10.1017/S153759271800110X

Séjourné T., Flamary R., Courty N., Gramfort A., & Peyré G. (2021). Sinkhorn divergences for unbalanced optimal transport. *Journal of Machine Learning Research*, *22*, 1–46. 10.48550/arXiv.1910.12958

Solomon J., De Goes F., Peyré G., Cuturi M., Butscher A., Nguyen A., Du T., & Guibas L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, *34*(4), 1–11. https://doi.org/10.1145/2766963

Solomon J., Rustamov R., Guibas L., & Butscher A. (2014). Wasserstein propagation for semi-supervised learning. In *Proceedings of the 31st International Conference on Machine Learning* (pp. 306–314). PMLR.

Spelta A. (2026). Density-based machine learning model averaging for inflation forecasting. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *75*(2), 364–406. https://doi.org/10.1093/jrsssc/qlaf048

Spelta A., Flori A., & Pammolli F. (2018). Investment communities: Behavioral attitudes and economic dynamics. *Social Networks*, *55*(1), 170–188. https://doi.org/10.1016/j.socnet.2018.07.004

Spelta A., Pagnottoni P., & Pecora N. (2025). Modelling emergency disaster mortality around the world: A network-based distributional inference approach. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnaf140. https://doi.org/10.1093/jrsssa/qnaf140

Spelta A., & Pecora N. (2024). Wasserstein barycenter for link prediction in temporal networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *187*(1), 180–208. https://doi.org/10.1093/jrsssa/qnad088

van den Berg B. (2024). Dealing with uncertainty in cyberspace. *Computers & Security*, *144*, 103939. https://doi.org/10.1016/j.cose.2024.103939

Villani C. (2008). *Optimal transport: Old and new*. Springer.

Zhou D., Bousquet O., Lal T. N., Weston J., & Schölkopf B. (2003). Learning with Local and Global Consistency. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in Neural Information Processing Systems* (Vol. 16). MIT Press. https://proceedings.neurips.cc/paper_files/paper/2003/file/87682805257e619d49b8e0dfdc14affa-Paper.pdf

Zhu X., & Ghahramani Z. (2002). *Learning from labeled and unlabeled data with label propagation*. Technical Report CMU-CALD-02-107.

Zhu X., Ghahramani Z., & Lafferty J. D. (2003). Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning* (pp. 912–919). ICML.