

Structural asymmetry as a fraud signature: Detecting collusion with Heron's information coefficient

Allana Tavares Bastos ^a,* , Tiago Alves Schieber ^b, Renato Hadad ^c, Laura C. Carpi ^d,
Martín Gómez Ravetti ^a

^a Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

^b Departamento de Ciências Administrativas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

^c Departamento de Engenharia de Produção, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, MG, Brazil

^d FutureTech Ltda., Belo Horizonte, MG, Brazil

ARTICLE INFO

Keywords:

Complex networks
Information theory
Community structure
Bidding and tendering
Network science
Covert networks

ABSTRACT

Fraud in public procurement remains a persistent challenge, especially in large, decentralized systems like Brazil's Unified Health System. We introduce Heron's Information Coefficient (HIC), a geometric measure that quantifies how subgraphs deviate from the global structure of a network. Applied to over eight years of Brazilian bidding data for medical supplies, this measure highlights possible collusive patterns that standard indicators may overlook. Unlike conventional robustness metrics, the Heron coefficient focuses on the interaction between active and inactive subgraphs, revealing structural shifts that may signal coordinated behavior, such as cartel formation. Synthetic experiments support these findings, demonstrating strong detection performance across varying corruption intensities and network sizes. While our results do not replace legal or economic analyses, they offer an effective complementary tool for auditors and policymakers to monitor procurement integrity more effectively. This study demonstrates that simple geometric insight can reveal hidden dynamics in real-world networks better than other Information Theoretic metrics.

1. Introduction

The outbreak of the COVID-19 pandemic in 2020 brought forth an unprecedented demand for medical supplies worldwide, causing potential risks to the availability of essential healthcare items [1–3]. This surge in demand resulted in fierce competition among suppliers, particularly for individual protection materials such as hospital aprons, rubber gloves, and masks. In Brazil, where health is constitutionally guaranteed as a universal right, the government is responsible for the procurement and distribution medical supplies for over 210 million citizens through the Brazilian Unified Health System (SUS) [4–10].

Public procurement plays a crucial role in ensuring fair competition among companies to meet the nation's healthcare demands. However, unethical practices, such as collusion, have been observed among co-bidding companies, raising concerns about the integrity of the procurement process. Collusion in bidding networks is challenging to detect due to the complexity of such networks [11].

To address these issues, we propose a measure called Heron's Information Coefficient (HIC) to gain insights into bidding processes for health-related items in Brazil. By understanding network structures and dynamics, we aim to identify potential collusion and enhance the transparency and integrity of public procurement. As such, we consider both the topological connectivity and the degree

* Corresponding author.

E-mail address: allanatavares@dcc.ufmg.br (A.T. Bastos).

<https://doi.org/10.1016/j.chaos.2026.118119>

Received 15 November 2025; Received in revised form 13 January 2026; Accepted 19 February 2026

Available online 4 March 2026

0960-0779/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

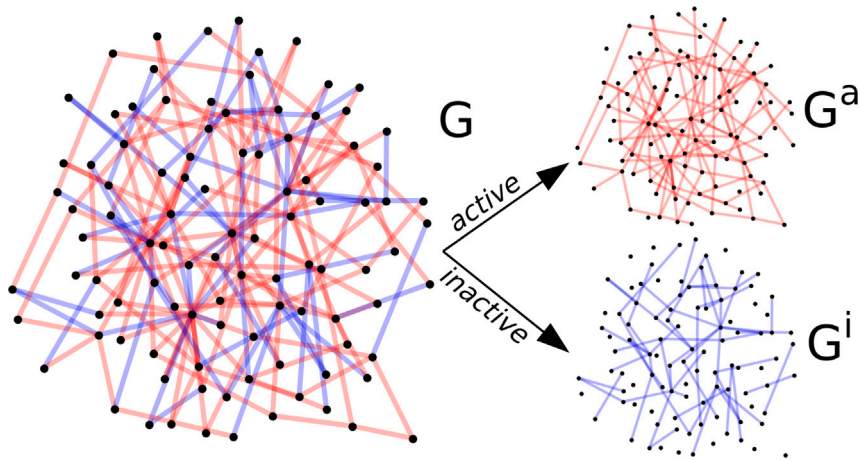


Fig. 1. A small network G with two complementary subgraphs G^a (active) and G^i (inactive). The D-distance between each pair is $D(G, G^i) = 0.4582513$, $D(G, G^a) = 0.3032006$ and $D(G^i, G^a) = 0.2715595$. The largest triangle with perimeter $D(G, G^i) + D(G, G^a) + D(G^i, G^a)$ has area 0.05134147 but the triangle with sides $D(G, G^i)$, $D(G, G^a)$ and $D(G^i, G^a)$ has 0.03964952 and, thus, the fraction between the equilateral triangle and this second triangle $0.03964952/0.05134147 = 0.77$ is the Heron’s coefficient.

of protection provided to the system’s elements. Complex network theory provides effective tools for analyzing various dynamic systems, enabling a better understanding of transportation, diffusion, synchronization, and social networks, among others [12–20].

Our approach focuses on network robustness, quantifying a network’s resilience under failures or changes. By applying Heron’s coefficient, we assess the system’s protection level and its efficiency, considering the activity patterns of individual nodes or links. This approach can be extended to devise intelligent attack strategies and evaluate the impact of intentional failures on interconnected systems.

The proposed methodology was tested on both real and artificial networks, revealing insightful behavior in artificial networks and highlighting unusual patterns in the interactions among co-bidding companies for medical supplies in Brazil. The results of this study can provide valuable information to control agencies, aiding their supervision of public procurement processes and combating corruption effectively.

2. Methods

A network is a set of vertices, V , connected by links, E . Generally, the network is just the structure over which a time-dependent dynamical process occurs, and its elements must not be available to operate during all periods. In this way, at a given instant $t \geq 0$, the original network $G = (V, E)$ can be decomposed into two components that represent the active and inactive subgraphs at that instant: $G^a(t) = (V^a(t), E^a(t))$ and $G^i(t) = (V^i(t), E^i(t))$ (see Fig. 1).

If $G^a = G$, the entire network is active but very vulnerable to failures as any activity can fail. Conversely, the condition $G^i = G$ implies a system in topological stasis, fully protected yet dynamically inert. All other intermediate possibilities result in networks that may be more or less similar to the original one. To compare the active and inactive parts and the original network, we employ the D measure [21], a pseudo-metric between networks. This highly precise measure of network dissimilarity relies on the Jensen–Shannon divergence [22] to assess topological differences in connectivity patterns, effectively capturing the divergence in information flow along shortest paths.

Formally, let $P_i = \{p_i(j)\}$ denote the distance distribution of node i , where $p_i(j)$ represents the fraction of nodes located at a shortest-path distance j . We quantify the topological heterogeneity of the system via the Network Node Dispersion (NND), defined as $NND(G) = \mathcal{J}(P_1, \dots, P_N) / \log(d + 1)$, where \mathcal{J} is the Jensen–Shannon divergence and d is the network diameter. The D -measure synthesizes three distinct topological descriptors: (i) the global difference between the averaged node-distance distributions, μ_G and $\mu_{G'}$; (ii) the aforementioned NND, which captures local variance in distance profiles; and (iii) a spectral comparison of the α -centrality values of the graphs and their complements. This final term, also computed via Jensen–Shannon divergence, is critical for resolving isomorphisms in distance-regular structures and distinguishing between complete graphs of varying sizes or systems containing disconnected components. As such, the distance $D(G, G_0)$ between two networks enables a comparison based on global distance structure rather than local connectivity alone:

$$D(G, G') = w_1 \sqrt{\frac{\mathcal{J}(\mu_G, \mu_{G'})}{\log 2}} + w_2 |NND(G) - NND(G')| + w_3 \left(\sqrt{\frac{\mathcal{J}(P_G^\alpha, P_{G'}^\alpha)}{\log 2}} + \sqrt{\frac{\mathcal{J}(P_{G^c}^\alpha, P_{G'^c}^\alpha)}{\log 2}} \right), \tag{1}$$

where N and M are the sizes of G and G' , respectively, and G^c indicates the complement of G . The weights w_1 , w_2 and w_3 satisfy $w_1 + w_2 + w_3 = 1$, and the authors of the D-measure [21] show that the choice of the weights does not change the metric character

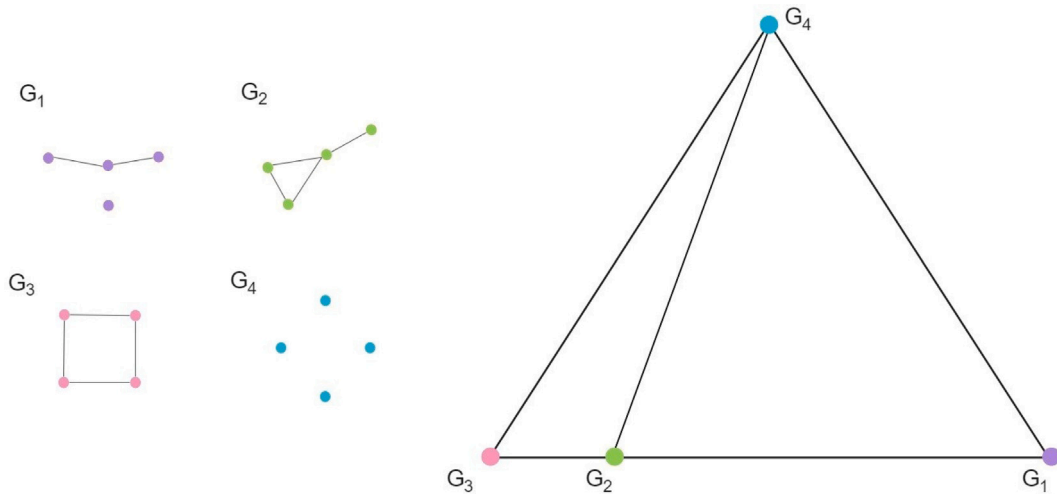


Fig. 2. Multi-dimensional scaling map considering the D -distance and four small networks, G_1 , G_2 , G_3 and G_4 . Here we visualize how non-isomorphic networks with 4 nodes form triangles with different areas based on the D -measure.

and present a detailed discussion on the weight selection. Typically, w_3 is set to 0.1, as a high value for this term would distort the D values, minimizing the power of the NND in detecting these important topological changes. Due to the computational complexity added with this term, w_3 can be avoided without losing precision—unlike w_1 and w_2 , which are considered equally important and thus set to $w_1 = w_2 = 0.5$. By setting $w_3 = 0$, we rely entirely on terms w_1 and w_2 , which are much better at detecting massive structural breaks, such as phase transitions or disconnections.

In addition to its precision, the D -measure also satisfies the triangular inequality, so that, for any given three networks G_1 , G_2 and G_3 , $D(G_1, G_3) \leq D(G_1, G_2) + D(G_2, G_3)$. By satisfying positive definiteness, symmetry, and the triangular inequality, the D -measure effectively projects the complex, non-Euclidean topology of graph structures into a searchable metric space [23]. This projection ensures that the familiar laws of geometry and algebra remain valid, allowing us to reason about the distance between complex systems as rigorously as we reason about distance in the physical world.

The operational impact of this geometric validity cannot be overstated. The triangular inequality is the cornerstone of efficient high-dimensional search, since it enables both the computational scalability and the semantic definition of structural anomalies. It allows us to bound the distance between two objects without explicitly comparing them: if G_1 is close to G_2 , and G_2 is far from G_3 , the inequality mathematically guarantees that G_1 is also far from G_3 . Without the pruning power granted by this axiom, the index structures required to identify outliers in billion-node graphs would be computationally intractable. Thus, the inequality transforms the concept of an “anomaly” from a subjective observation into a geometric fact: an object that physically cannot be embedded within the dense manifold of the distribution without violating the axioms of the space.

Given the triangular inequality, if at least one of these distances is positive, there are infinite different triangles possessing a perimeter ($2P$) equal to three times the average value between their distances: $2P = D(G_1, G_2) + D(G_2, G_3) + D(G_1, G_3)$. The most symmetrical situation occurs when the distance values are the same, resulting in an equilateral triangle in the metric space. By extension, all other configurations generate more heterogeneous distance patterns and the perimeter of their triangles becomes a measure of the total structural divergence within the triplet. Since the largest triangle for a given perimeter is equilateral and the Heron’s formula gives the area of a triangle with sides $D(G_1, G_2)$, $D(G_1, G_3)$ and $D(G_2, G_3)$ as $\sqrt{P(P - D(G_1, G_2))(P - D(G_1, G_3))(P - D(G_2, G_3))}$, we define Heron’s Information Coefficient as the fraction between the area of that triangle and the area of the equilateral triangle with perimeter $2P$:

$$H_e(G_1, G_2, G_3) = \frac{3\sqrt{3P(P - D(G_1, G_2))(P - D(G_1, G_3))(P - D(G_2, G_3))}}{P^2}. \tag{2}$$

Fig. 2 shows a multidimensional scaling map of the distances G_1 , G_2 , G_3 and G_4 between four small networks. In this representation, we can see that the largest possible triangle formed by triples of these networks has vertices G_1 , G_3 and G_4 with similar sides. When combining G_1 , G_3 and G_4 no triangle is formed because, in fact, $D(G_1, G_3) = D(G_1, G_2) + D(G_2, G_3)$. Thus, among all non-isomorphic networks of size four, $H_e(G_1, G_3, G_4) = 0.996$ represents its highest possible value while $H_e(G_1, G_2, G_3) = 0$ represents its minimum.

When HIC equals 1, the three networks’ topologies in the system are equally different. In contrast, a zero value means a degenerate triangle, that is, either at least two topologies are equal or the topological differences between a pair are as large as the sum of the other two’s topological differences.

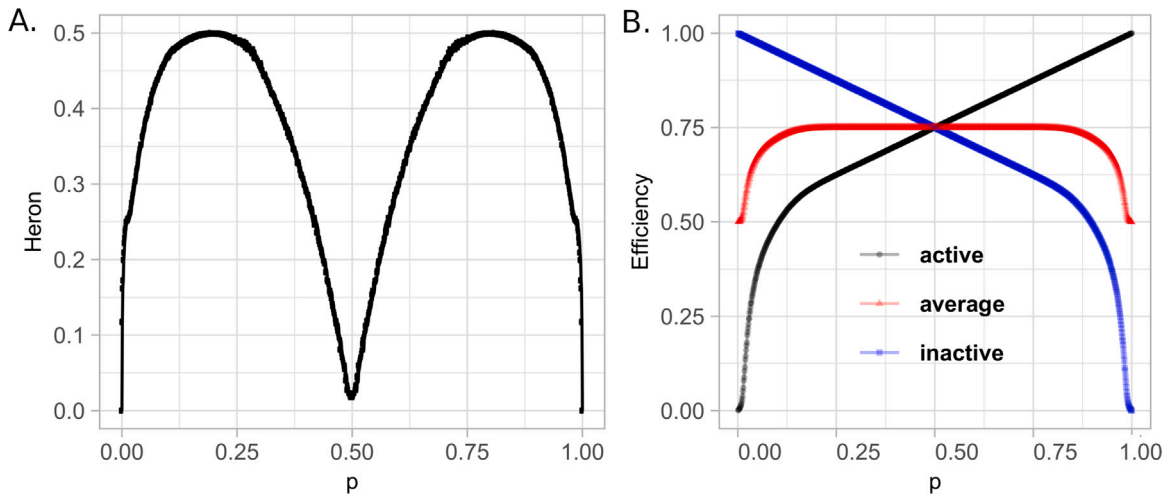


Fig. 3. (A) shows the average values of the Heron's Information Coefficient for 100 experiments. (B) shows the network efficiency evolution for active and inactive subgraphs, as well as the average. The average efficiency of the active and inactive subgraphs tend to plateau close to 0.75 after a spike of the maximum HIC value.

3. Characterizing heron's information coefficient's behavior

To understand the role that Heron's Information Coefficient has as a measure of activity and protection of a networked system, we first consider a random graph model, Erdős-Rényi [24]. In this model, each link i is independently activated according to a uniform distribution $p_i \in [0, 1]$, where $p_i = 0$ represents that link i has no activity in the network, and $p_i = 1$, its complete activity. Fig. 3(A) shows the average values of HIC for 100 experiments with different seeds.

For small p values, the inactive network closely resembles the complete graph, resulting in low HIC values. As p increases, the topologies of the original, active, and inactive networks diverge until a maximum, reached when approximately 20% of the links are active. Beyond this point, the HIC gradually decreases until a local minimum in $p = 0.5$. At this local minimum, the active and inactive networks are topologically similar subgraphs of the original network, as the probability of the existence of a link is 50% in both.

The random graph model is a good example, but it does not tell the whole story. In such a network, each link has an equal probability of being active, a pattern that is not often presented in the real world. In social networks, for example, some ties tend to be stronger than others. To introduce such heterogeneity while maintaining the same expected number of links, we assign link activation probabilities proportional to their edge betweenness centrality (b_e), a measure reflecting the number of shortest paths that traverse each edge. Specifically, for a given parameter γ , the activation probability of an edge is defined as $p_e \propto b_e^\gamma$.

Fig. 4(A) shows an example of an active-inactive configuration in a network with 100 vertices, where the thicker and redder edges denote higher betweenness centrality values. As depicted in Fig. 4(B), we computed the average HIC across 100 random seeds for different γ values. Negative γ values correspond to preferential activation of low-betweenness edges, whereas positive γ values emphasize highly central ones. The results reveal that HIC tends to increase when edges with higher betweenness are more likely to be active, reflecting a more efficient structural differentiation between active and inactive subgraphs.

Overall, a high Heron coefficient is associated with the system's efficiency in transferring information. We observe high values in systems whose active and inactive parts are equidistant from the original network and from each other, which indicates different connectivity patterns without losing information from the original network. In the network represented in Fig. 4, the preferential attachment to high edge betweenness edges ensures that the information transfer during the activity period travels the network more efficiently in different points, whereas inactivity patterns protect the network's periphery. Simultaneously, however, important links in the system – those with high betweenness centrality, for instance – may fail. In this case, a possible strategy would be to interchange the active and inactive networks while retaining Heron's information coefficient so that the information that reaches the system's periphery rapidly through the most central active network can further spread to peripheral nodes.

This analysis highlights that edges connecting distinct network modules, which are typically those with high betweenness, play a crucial role in maintaining global connectivity. Prioritizing their activity preserves inter-community communication and enhances information flow across the network. Notably, betweenness centrality is among the most effective measures for identifying strategic positions within covert or collusive structures [25,26]. Furthermore, this centrality is used to detect collusion through sequential edge removal, in a process analogous to a targeted network attack.

As such, we next evaluate HIC's sensitivity to network perturbations relative to classical robustness metrics: degree centrality, betweenness centrality, clustering coefficient, and modularity. The goal is to assess how effectively each metric captures a network's ability to preserve connectivity and structural integrity under deliberate disruptions.

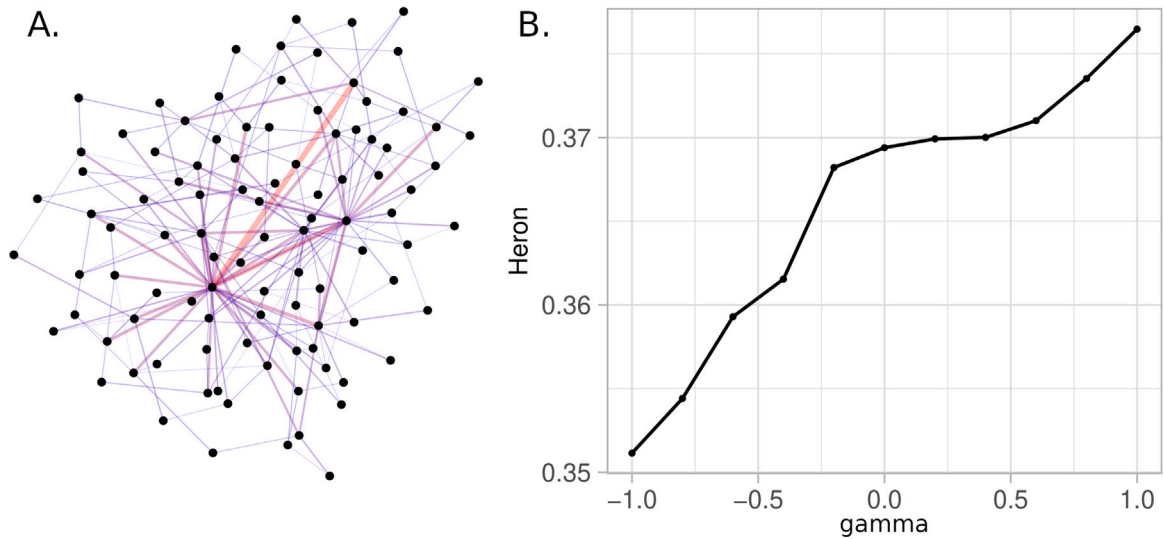


Fig. 4. Network with 10% of active links, where link activation probability depends on edge betweenness centrality (b_e) and a real number γ , such that $p_e \propto b_e^\gamma$. (A) shows an example of active and inactive edges and (B) the average HIC for 100 seeds across different γ values.

To this end, we conducted experiments across multiple network topologies. The scale-free Barabási–Albert model grows through preferential attachment, producing hubs with high degree centrality and a topology resilient to random failures but highly vulnerable to targeted attacks on hubs. The small-world or Watts–Strogatz model combines local clustering with short global path lengths, which confer it balanced robustness across attack strategies. On the other hand, the Erdős–Rényi network, whose robustness depends mainly on edge density, served as a baseline due to its rapid fragmentation under targeted attacks. We also analyzed a community-structured network, characterized by densely connected clusters with sparse inter-community links, to evaluate the effects of bridge removal on global connectivity. In contrast, the hub-and-spoke topology is a centralized structure dominated by a single hub, extremely fragile under targeted removal of that central node, emphasizing dependence on a single point of failure. Finally, the real bidding network exemplifies the problem we aim to tackle, capturing the genuine structural patterns observed in procurement systems.

Each network underwent controlled node removal experiments, in which 10% of the nodes were sequentially removed in descending order of degree. Across all types of network, our coefficient consistently exhibited a higher sensitivity to structural degradation than the other metrics evaluated. Our coefficient's performance reinforces its suitability for detecting covert or collusive groups, surpassing state-of-the-art methods based on edge betweenness in capturing critical structural vulnerabilities.

Extensive testing across multiple network configurations confirmed the expected behavior and robustness of our coefficient. Together, these findings demonstrate that Heron's Information Coefficient captures both the structural balance between active and inactive subgraphs and the dynamic efficiency of the information flow within the system. Its sensitivity to perturbations across a variety of topologies highlights our coefficient's potential.

3.1. Heron's information coefficient in collusion detection

Within the broader anomaly detection literature, detecting anomalies in data serves as a critical filter for identifying deviations in high-dimensional systems, spanning domains from financial forensics to social network security. While applicable to various data modalities, we prioritize graph-based approaches due to their unique capacity to model long-range correlations and structural interdependencies. In the context of fraud, for instance, anomalies are rarely isolated events; they are inherently relational. For instance, fraudulent activity typically manifests through specific topological patterns: either as opportunistic contagion, where fraud propagates through established social links; or as organized collusion, where dense subgraphs of actors synchronize their behavior [27].

Historically, detection frameworks relied on heuristic feature engineering and domain-specific statistical models. This paradigm is labor-intensive but also brittle, struggling to generalize to previously unseen anomaly types [28]. While traditional machine learning techniques, such as Support Vector Machines (SVMs) and matrix factorization, attempted to automate feature extraction, they frequently succumb to the curse of dimensionality and prohibitive computational costs when scaling to massive adjacency matrices. Consequently, the field has pivoted toward Deep Learning to capture the non-linear complexity of graph data and learn expressive latent representations capable of separating normal from anomalous manifolds. However, this gain in predictive performance introduces a new critical limitation: these models operate as "black boxes", offering superior accuracy but lacking the interpretability required for high-stakes decision-making.

Metric Sensitivity to Structural Perturbations

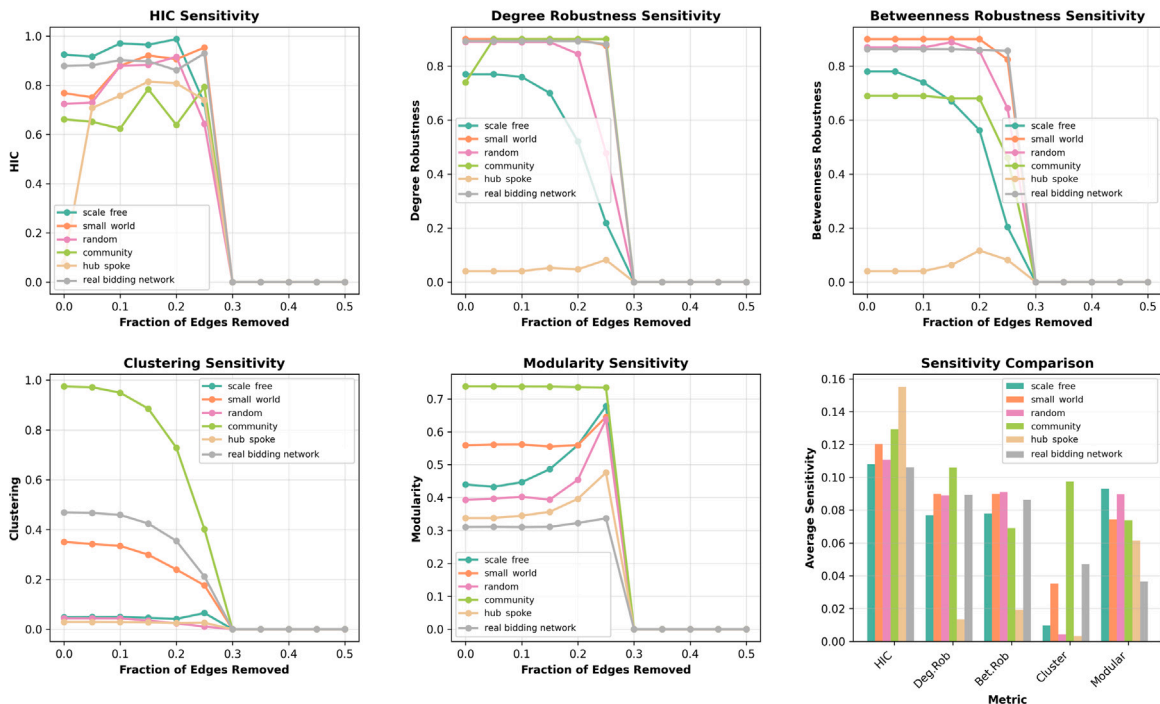


Fig. 5. Sensitivity comparison of the edge removal across relevant metrics (HIC, degree robustness, betweenness robustness, clustering, and modularity) for different network types (scale-free, small-world, Erdős–Rényi, community-structured, hub-and-spoke, and a real bidding network). The real network corresponds to the 2016 Surgical Masks network, where our metric outperformed the others, as it consistently did for Brazilian bidding networks.

More specifically, the study of covert networks is challenging due to a myriad of reasons, one of which is data incompleteness and unreliability. Publicly available and reliable datasets, such as the Sicilian Mafia dataset [25], are rare and heavily contingent on judicial disclosures. Criminal network research is also uniquely sensitive to sampling biases. For example, random sampling can severely distort the understanding of a network clustered around a few focal nodes. In fact, the collection of complete data is a virtually impossible task, failing to properly capture the complete dynamics of the networks, with bias being inevitable [26]. This uncertainty is compounded by the boundary-specification problem: in the absence of explicit membership rosters, defining the limits of a clandestine network requires inferring latent connections within a secretive, non-cooperative environment.

Moreover, the current detection methods used by law enforcement agencies are restricted from the greater public. Understandably so. Knowing the exact criteria used to identify potential criminal agents could ease organized crime’s efforts in evading capture, in an adversarial adaptation. In the criminology literature, a recurrent approach consists of Network Science techniques. Its primary utility lies in its ability to map dynamic covert systems without relying on *a priori* organizational templates. Rather than forcing data into presumed hierarchies, graph-theoretical approaches allow the structural properties of the network to emerge from the empirical data.

However, the interpretation of these emergent topologies requires significant nuance to avoid misleading inferences. For instance, the high degree connectivity of an element in the network does not necessarily equate to high status and, conversely, peripheral actors, with low connectivity, may exercise disproportionate control through strategic compartmentalization. For instance, in human trafficking rings, “madams” may appear peripheral because they operate independently, despite driving the market demand and controlling “managers” who appear more central in terms of degree centrality [29]. Therefore, structural measures must be contextually grounded. Researchers must integrate edge attributes and qualitative metadata, recognizing that the absence of a tie carries as much information-theoretic weight as its presence.

Notably, sequential node removal based on betweenness centrality is among the most effective measures for identifying strategic positions within covert or collusive structures [25,26]. Betweenness produced a greater impact in dismantling a covert network, showing its capability to properly identify relevant elements. This evidence underscores that in covert networks, the capacity to bridge structural holes and control information flow is a more accurate predictor of systemic importance than mere connectivity.

For bidding processes, the collusion between companies is hard to identify and can take multiple forms, using the techniques mentioned in this section [30]. For instance, the recurrence of bids for a particular item [2,31] can contribute to cartel companies’ prior planning and their subsequent organization in rounds and punishment of the players who fail to cooperate in the next bid.

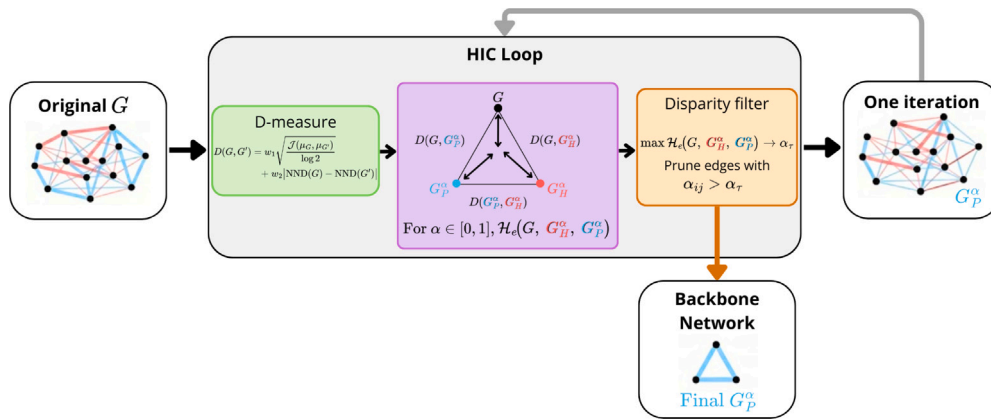


Fig. 6. Diagram of our HIC method for collusion detection with a disparity filter. The HIC loop consists of finding the optimal separation threshold α_T , defined as the point where the structural divergence between the original network (G), the backbone (G_p^α), and its complement (G_H^α) is maximized. Geometrically, this corresponds to maximizing the area of the triangle formed by these three states in the D-measure metric space in each iteration. HIC is maximized for a set number of iterations or until the graph is entirely disconnected, depending on the experiment.

Companies that act in collusion usually participate in a large amount of bidding to share the average earnings or confer an impression of honesty throughout the process [31]. Therefore, we model each bidding item as a network in which each node is a company participating in the bidding process. Each link exists if two companies are competitors in the same contest. To that end, multiple links represent multiple instances where the companies were co-bidding. As such, the bidding process possesses a complex topology structure that is very informative but often difficult to analyze.

The theoretical benchmark of a public bidding network consists of nodes uniformly connected and characterized by the nonexistence of strong and connected communities, leading to a challenging trade-off: extracting connections from this set and retaining those that represent a high-level interaction between companies without overemphasizing what could be simple statistical deviations. In this scenario, Heron’s Information Coefficient is useful due to its quantification of structural imbalance in dynamic networks, revealing signs of the coordinated behavior typical of collusion.

Here, we use the disparity filter algorithm [32] alongside the maximization of HIC to highlight the central players in the BID network. Fig. 6 shows how our method works iteratively.

The idea of the disparity filter is to associate with each network’s link a weight α_{ij} that represents a significant level provided by the distribution of the local uniform randomness null hypothesis [32]. Relevant links are chosen so that $\alpha_{ij} < \alpha$, $\alpha \in [0, 1]$ being a fixed significance level. Instead of arbitrarily choosing the significance levels, we maximize HIC when α varies. Disconnected companies are excluded from the resulting network. The process is repeated until reaching a significance level as small as desired. The selection of the most relevant group is made iteratively, starting with the original network with a significance level equal to 1. A new subgraph is obtained at each step by removing edges that possess a higher significant level. The goal is to find a lower significance level for each step, resulting in a higher HIC. By definition, this happens when the topologies of the active, inactive, and original networks are close to equidistant.

Each company that participated in the bidding process is a node connected by links weighted by the number of times they have competed in the same bidding process. The way nodes interact and connect provides essential information about the system [33]. In particular, information about their connectivity patterns represents a valuable tool on how different dynamics occur in the system. For example, the speed at which information travels within a cohesive community is much greater than in a more sparse one. As such, let G be the original weighted network after applying the disparity filter, for a fixed $\alpha \in [0, 1]$, and G_a^α be the network obtained from G by removing links whose weight is greater than or equal to α . As such, $G_i^\alpha = G - G_a^\alpha$ is the complement of G_a^α to G . We define α_T as the smallest value of α such that $H_e(G, G_a^\alpha, G_i^\alpha)$ is as large as possible. This optimal threshold α_T thus identifies the most informative backbone of the network, balancing cohesion and separation between active and inactive subgraphs. In doing so, it reveals the most representative organization of competitive interactions among companies.

To characterize HIC’s behavior in real procurement data, we use a Brazilian dataset of infrastructure projects procurement [30], which documents the procurement activities of the state-controlled oil company Petrobrás between 2002 and 2013. This era encompasses the systemic market manipulation uncovered by the “Operation Car Wash” investigation, a landmark inquiry that exposed the “Club of 16” – a cartel of Brazil’s largest construction firms that orchestrated bid-rigging and price – fixing across major infrastructure projects. Despite lacking the temporal aspect, this dataset provides valuable insights from known collusive agents in Brazilian biddings. The original authors established, between machine learning methods, the superior efficacy of ensemble methods, specifically identifying Extra Trees, Random Forest, and AdaBoost as the top-performing algorithms, with detection rates between 81% and 95%. We reproduced their experiments and tested disparity filters with HIC and degree robustness, given its strong performance in the experiments shown in Fig. 5.

The network comprises 272 corporate nodes, of which 47 (17.28%) are identified as corrupt; notably, only 80 companies successfully won at least one of the 101 analyzed auctions. Within the total volume of 683 bids, 128 were flagged as collusive.

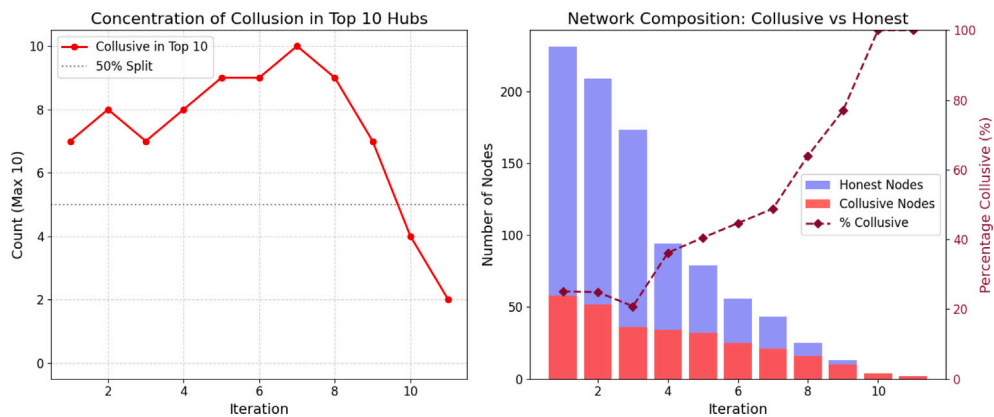


Fig. 7. Evolution of network composition during iterative backbone extraction for Petrobrás procurement from 2002 to 2013. (Left) Hub Dominance: The count of collusive agents within the top 10 most central nodes identified with HIC, across iterations. The algorithm consistently retains collusive “ringleaders” in the top ranking, peaking at 100% dominance by Iteration 7. (Right) Network Composition: The stacked bars show the absolute number of nodes remaining at each step, colored by ground truth (Blue = Honest, Red = Collusive). The dashed line tracks the percentage of collusive nodes in the remaining subgraph. As the algorithm iterates, it disproportionately sheds honest nodes (noise), resulting in a monotonic increase in cartel density, reaching 100% purity at Iteration 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A defining structural characteristic of this network is its low collusion assortativity ($r = 0.1480$). This suggests a deliberate “cover bidding” strategy, where collusive agents dilute their presence among honest participants to mask topological clustering. Despite this challenging scenario, HIC was able to isolate only collusive agents after 10 iterations, as shown in Fig. 7. The only moment our method failed to identify at least 5 collusive agents in its top 10 was in the tenth iteration, when the network only had 4 companies—all of which were corrupt.

To validate these findings, we conducted a rigorous sensitivity analysis using Monte Carlo simulations with bootstrap sampling for stability and null models for significance testing. The results reveal a sharp divergence in metric utility. Standard topological measures, specifically Degree Robustness and Modularity, failed to reject the null hypothesis, suggesting they are blind to the cartel’s non-modular structure. In contrast, HIC identified a significant structural anomaly ($Z=4.32$, $p < 0.001$). The observed HIC value (0.856) lies far outside the null distribution mean ($\mu \approx 0.69$), confirming that the collusion manifests not as simple clustering, but as a subtle “information backbone”. This provides compelling statistical evidence that HIC is a superior metric for detecting the subtle coordination characteristic of collusion, which standard topological measures overlook.

4. Experimental results

All the results on this section are based on the extraction of the network’s backbone with a disparity filter based on HIC maximization. A simple complexity analysis can be found in the Supplementary Material S1. The source code for the characterization of our coefficient’s behavior and other experiments can be found on GitHub (https://github.com/FutureLab-DCC/Heron_coefficient).

4.1. Artificial bidding network

We tested our method in artificial procurement networks generated using the Bernoulli Weighted Random Network (BWRN) model [34], which combines the Stochastic Block Model (SBM) for group formation with hierarchical structures for management layers. This generator creates synthetic networks that are statistically equivalent substitutes for real networks, allowing for research on covert networks. Its validity has been established through comparison with the publicly available Sicilian Mafia network [25], which is a covert network that covers phone calls and in-person meetings extracted from court proceedings. Thus, these synthetic networks replicate the behaviors characteristic of real colluding agents while addressing a critical challenge in covert network research: data availability [29]. We enhanced realism by incorporating successful bids to better simulate procurement networks with known corrupt agents. Then, we iterated the HIC filtering process until network sparsity prevented further analysis.

Fig. 8 demonstrates how the Heron coefficient progressively identifies corrupt agents. As the algorithm iterates, we can see that important corrupt nodes are maintained. These results corroborate our method’s potential to detect colluding agents in real-world procurement data.

To validate robustness, we tested the method in 100 random seeds and varying parameters, iterating until all HIC values reached zero, that is, the network dissolved. The results were consistent, as seen in 9. With a lower percentage of corrupt agents, the method stopped faster. Networks with lower corruption levels (5%, 10%) maintained higher accuracy throughout, generally staying above 0.8. Furthermore, the method showed great final accuracy for networks containing up to 30% corrupt agents, with the final

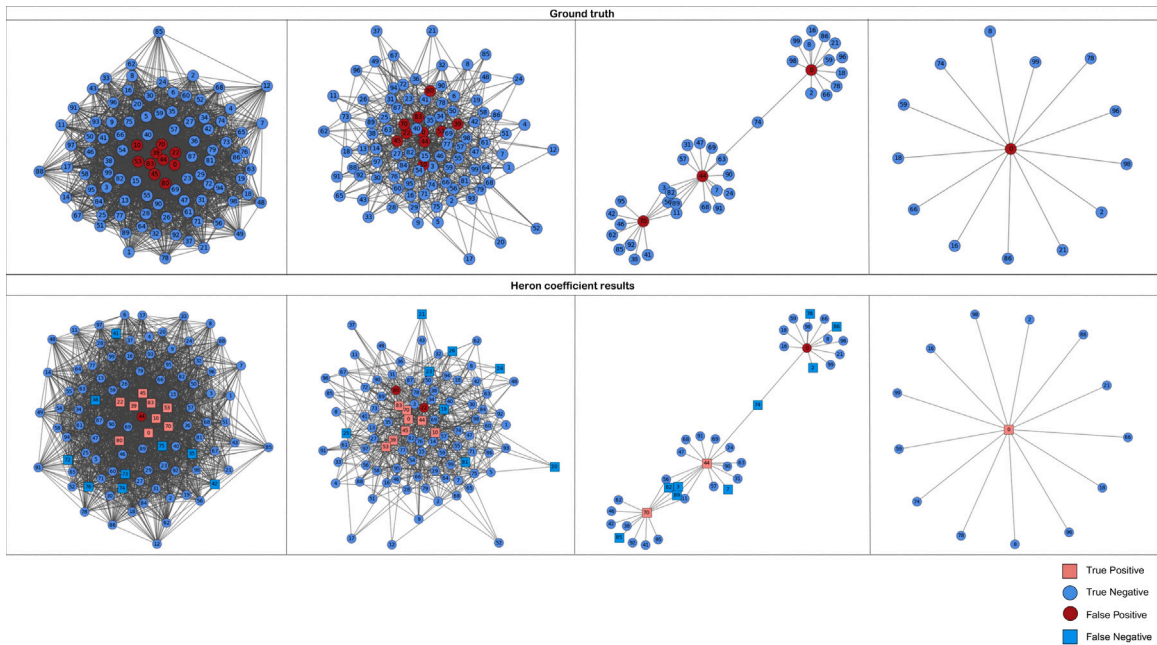


Fig. 8. Disparity filter with HIC to detect corrupt agents in a synthetic network. The image shows iterations 1, 3, 5 and 6 (the last one) for 10 corrupt agents in a network of 100 agents. Above are the original networks with the correct labels, with the analyzed network of the iteration below.

iteration successfully identifying at least one corrupt agent in most cases. Performance remains stable across network scales, although optimization would be necessary for graphs exceeding 10,000 nodes.

Scaling experiments on networks with up to 5000 nodes (all generated with 10% corrupt nodes across 100 seeds with the BWRN generator) revealed a fundamental structural principle. As shown in Fig. 10, larger organizations exhibit reduced structural robustness during backbone extraction, precisely because our method successfully identifies their sparse, efficiency-optimized architectures. This inverse relationship demonstrates that covert networks create fragile organizational skeletons that rapidly fragment when key nodes are removed. The HIC's sensitivity to these optimized structures confirms its effectiveness in exposing the critical trade-off between scale and vulnerability in covert network design.

Our method also performs effectively with partial network information, as shown in Fig. 11. We tested on networks growing to have 10 corrupt nodes and 100 overall nodes in 10% increments until their final size. We found that the average HIC behavior is consistent with that seen on Fig. 3. For 10 to 30% of the network, we have a higher HIC and an early detection. On the other hand, we have a steep decline for 70 to 90% of the network available. This trajectory suggests these networks possess a well-defined backbone where a subset of critical bidding relationships captures the system's essential dynamics.

In summary, our method is computationally cheap, converges fast, and shows promising results in detecting corrupt agents in bidding processes. The Heron Information Coefficient effectively identifies corrupted agents in early iterations, particularly at lower corruption levels. While accuracy becomes less stable at corruption levels exceeding 50% – an expected limitation since anomaly detection inherently struggles when anomalous nodes constitute the majority – the method remains valuable. Even after six iterations in high-corruption scenarios, it substantially reduces network size while retaining corrupt nodes. As such, our results suggest that the Heron coefficient can serve as a practical aid for corruption inspection, even in resource-constrained environments with limited detection capabilities.

4.2. The BID network

Brazil is a country of continental dimensions, the fifth largest in the world and the sixth in terms of population [35] with considerable regional and social inequalities [4,9,10]. As determined by the country's Constitution, the provision of services in healthcare and education is a citizen's right, met with resources collected through taxation at the federal, state, and municipal levels [4,6,7,10]. Supplied by public procurement, the national unified healthcare system (SUS) has promoted access to primary and emergency care and reached universal coverage of vaccination and prenatal care. As this system has become increasingly weak over the decades, the current infrastructure of primary care clinics and emergency units is mainly public. In contrast, hospitals, outpatient clinics, and diagnostic and therapeutic services are mostly private [5,10]. Consequently, even before a strenuous situation on the health system, such as a pandemic, the decreasing public health expenditures are insufficient and underfunded [4,5]. Furthermore, the inadequate funds allocated to the system are severely undermined by corruption in public procurement. The diversion of funds

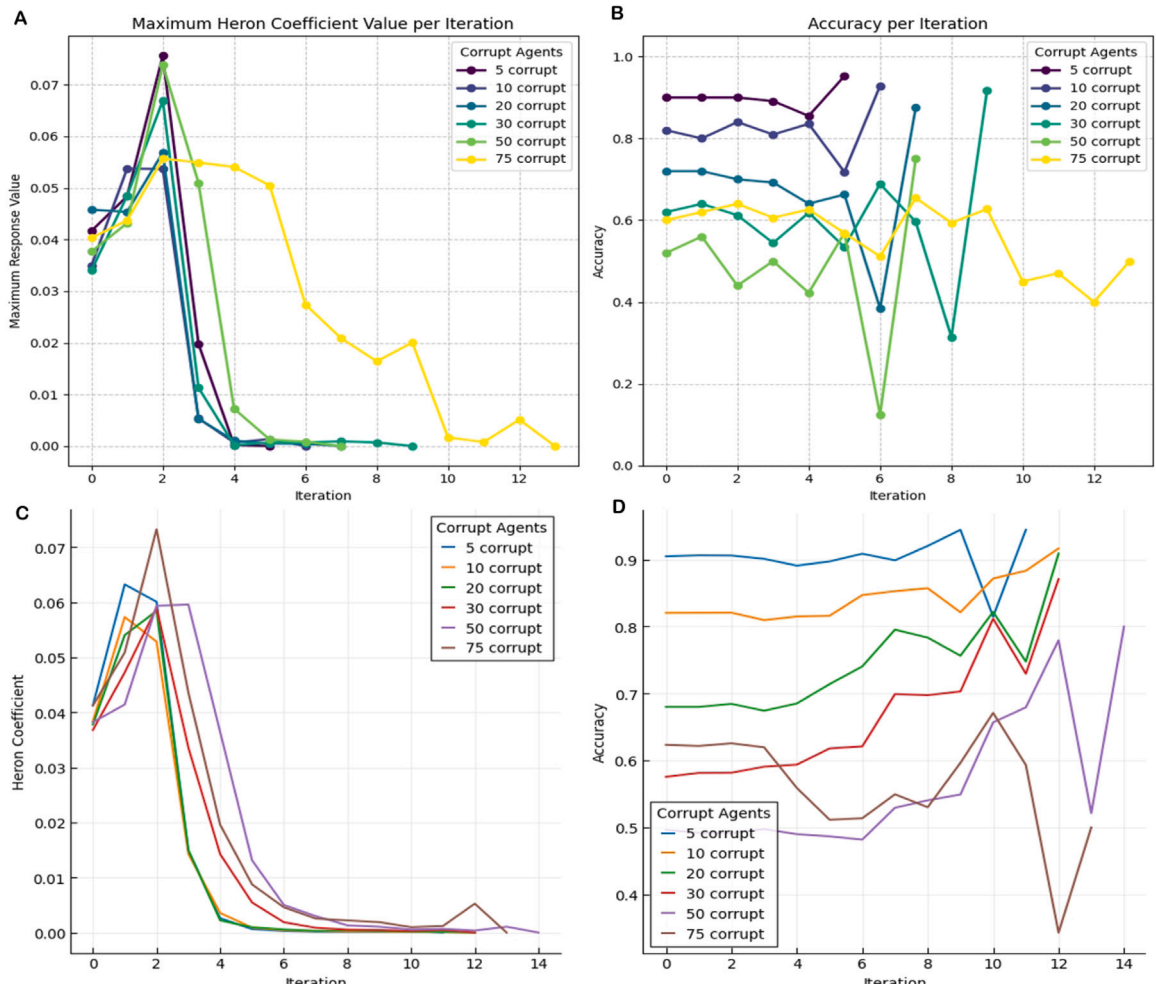


Fig. 9. Comparison of different numbers of corrupt agents for networks with 100 agents, for 1 seed (A and B) and the average on 100 seeds (C and D) showcasing the method’s stability. In later rounds, our method performs well even in challenging scenarios such as those where the deviating behavior is not isolated in a minority of agents, but in its majority.

meant to address social needs contributes to the increase in poverty and inequality [1,31] in an already unequal country[9]. Even with these setbacks, SUS has been able to attend to the population [4–6], albeit with execution flaws often engendered by the ultimately political challenges the system faces [7,36].

In 2020, the world faced a pandemic that urged the acquisition of essential and high-demand goods [1,3], such as personal protection equipment. Consequently, the COVID-19 outbreak highlighted significant inconsistencies in national plans to contain a pandemic and the lack of good governance and corruption prevention [1,2,37]. The response of the Brazilian system to the pandemic also faced political challenges [5] such as significant investments in the purchase of more than 5 million units of an ineffective treatment of the disease, chloroquine [38–40]. This context only exacerbates a historical problem in public procurement around the world: an estimated loss of 10% to 25% of global spending is due to corruption [31,41].

Public procurement in healthcare is a significant area of risk of corruption [31], which can occur through bid-rigging, funds embezzlement, opacity in governance, misuse of power, nepotism and favoritism in management, petty corruption in levels of service, fraud and theft of medical resources [1]. One such corrupt practice is cartel formation, which exposes the government to conspiring agents who violate the objective of bidding, incurring a loss of efficiency in public biddings on price and welfare in varying degrees [1,2,31,42]. Although descriptive analyzes can provide information on how susceptible to fraud the Brazilian procurement system is, as shown by the high price variability for some bidding items [43], we obtain little information on how the companies participating in a bidding process act together to bypass the inspection system.

We consider data from the Brazilian Federal Government’s public bids from 2013 to 2021 for three personal protective equipment: a surgical mask, rubber gloves and endotracheal tubes. The data is freely available through the Transparency Portal [44]. This channel monthly publicizes information gathered from several Federal Government systems. Considering all the bids of a given item, the network is composed of nodes weighted by the number of participating bids, linked by edges weighted by the number of

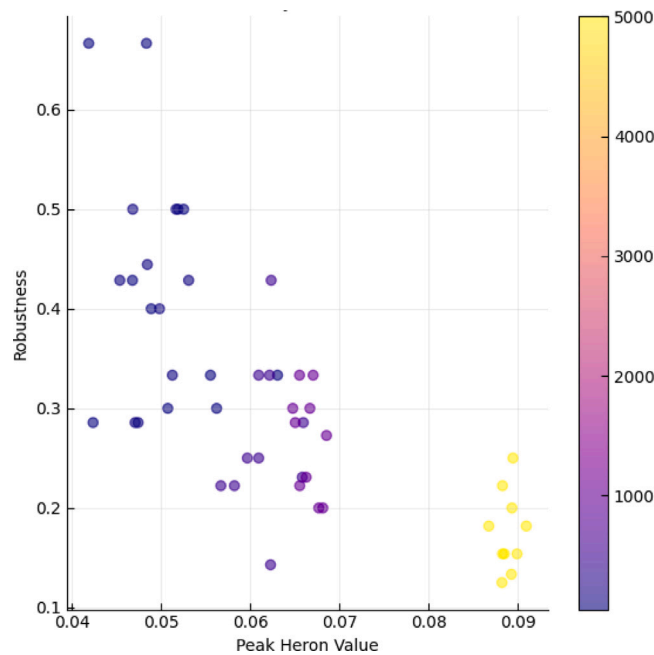


Fig. 10. Comparison of robustness and peak HIC value for BWRN networks with 10% of corrupt nodes, from 10 to 5000 corrupt agents. This scaling test demonstrates the effectiveness of using HIC to expose the critical backbone of covert networks.

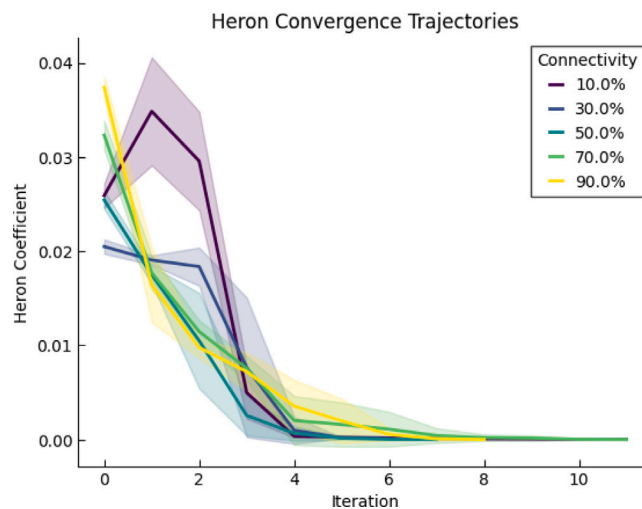


Fig. 11. Average HIC convergence in incrementally growing networks from 10% to 100% of their final size, over 100 seeds. The networks were generated with the BWRN generator to be ultimately composed of 10 corrupt nodes among 100 total agents.

times two companies took part in a bidding process. We tested the networks of these three products over six steps. Fig. 12 shows the networks for each one of the items, and a more detailed look at the companies year by year can be found in the Supplementary Material S3.

The market for each item has some peculiarities. Masks were in great demand by institutions and individuals during the pandemic period, with a massive increase in demand and offer. Conversely, gloves are items with a relatively predictable annual order, used exclusively by healthcare workers. Despite the pandemic, canceling elective surgeries can correlate to maintaining this market’s stable behavior.

It is important to note that a few companies hugely dominate the public bidding market for these items. Of the 733 companies participating in public bids for endotracheal tubes, 55% won at least one process. However, a surprising 3.5% won half of all submissions. Of the public proposals for masks and gloves, these numbers are 4% and 2%, respectively. Since the internal procedures

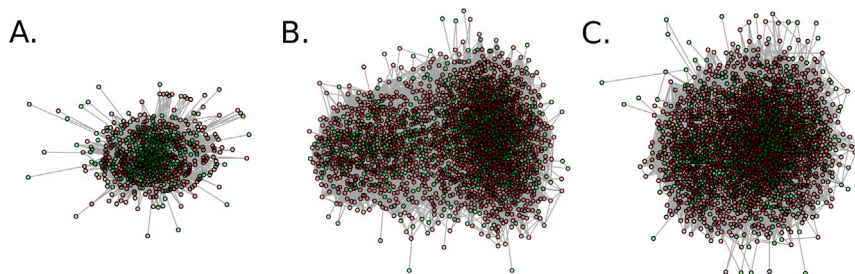


Fig. 12. (A) Tube (B) Mask and (C) Glove between 2013 and 2019. In each graph, the green nodes are the companies that won at least one bidding process, whereas the red ones never won. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

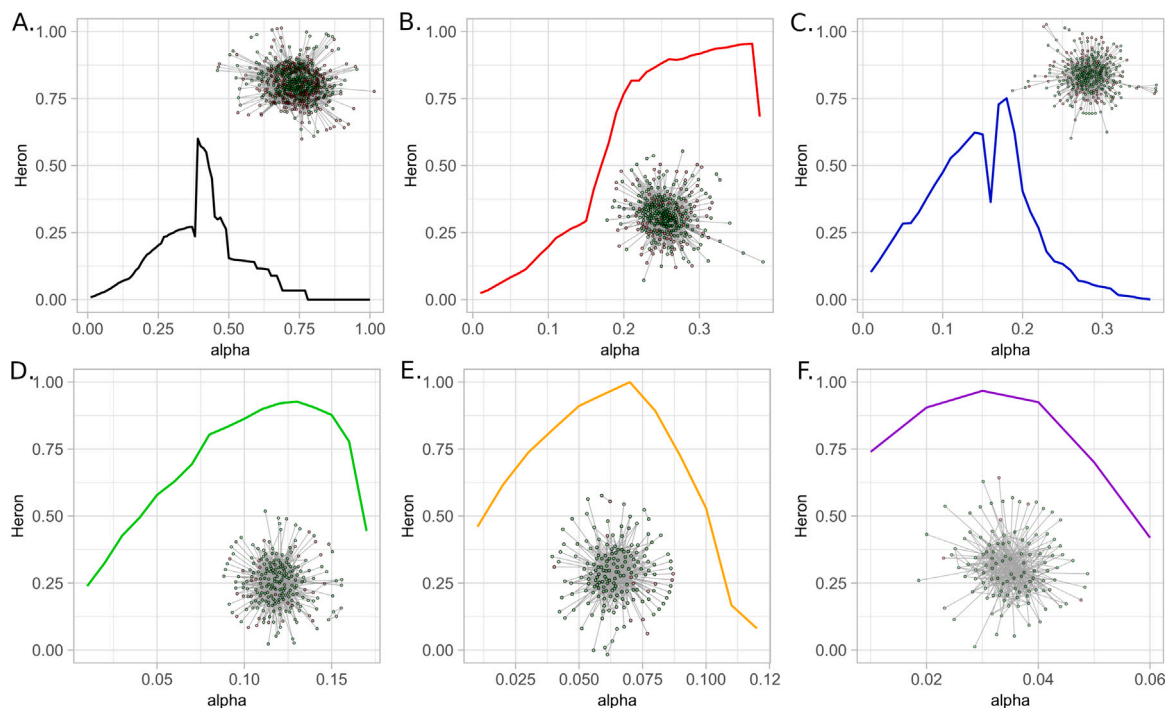


Fig. 13. Significance level α over six iterations for the tube network in 2021, from A to F. The graph also shows the shrinking networks, as edges are removed every step with our method, where green nodes are companies that won at least one bidding and red nodes never won. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of Brazilian bidding supervision agencies are, understandably, not publicly available, our approach relies on metrics and methods from covert network analysis, focusing on the sequential removal of edges [25,26] via the disparity filter.

The most informative subgraph is chosen by maximizing the Heron Coefficient in the network, as illustrated in Fig. 13, which depicts the aggregate bidding processes for tubes in 2021. This optimization is performed iteratively, adjusting the significance level α at each step to maximize the network's Heron Coefficient. Interestingly, the reduction of the significance level implies in an increase of the fraction of winners and, at the sixth step the resulting network usually contains only winners companies.

Fig. 14 shows the evolution of the highest significance level of links in the networks, the percentage of winning companies in the network and the common elements in 3 different periods: pre-pandemic (2013–2019), the pandemic's first year (2020) and its second year (2021). Further details of the significance level's experiments and results can be checked on the Supplementary Material S2.

The evolution of α with the removal of redundant links is similar for gloves (panels A to C) and endotracheal tubes (panels D to F), both before and after the pandemic. The six steps quickly find a subgraph with relevant links, with $\alpha \leq 10\%$. Concurrently, the proportion of winning companies within these subgraphs increases at each step, indicating that the most significant connections persist among successful bidders. In the tubes network, all companies in the sixth step had won at least one bid. The most significant difference between these two networks emerges in their temporal dynamics. The gloves network exhibits a stable composition, with

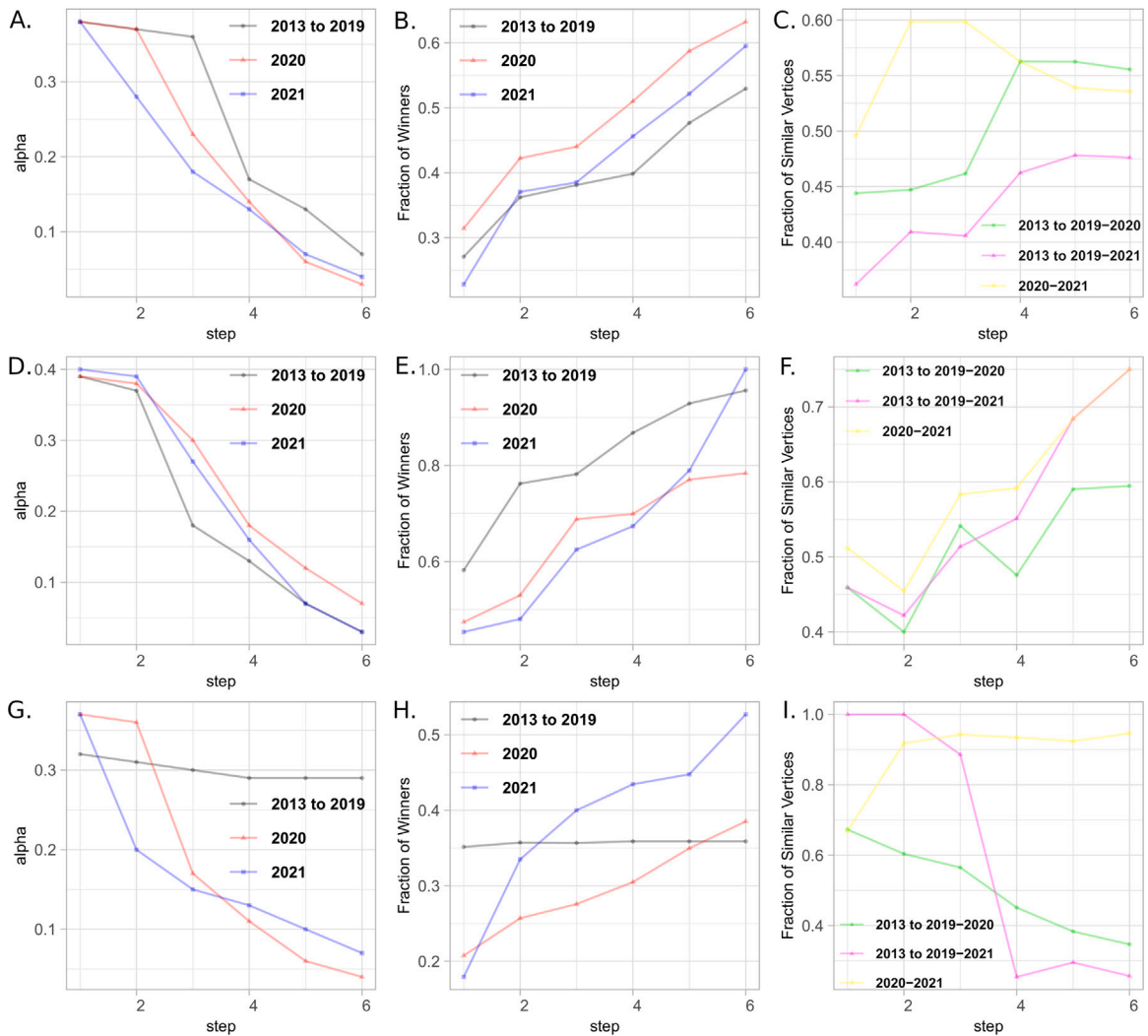


Fig. 14. Evolution of the significance level (α), fraction of winners and fraction of similar nodes for gloves (panels A to C), tubes (panels D to F) and masks (panels G to I) for each one of the algorithm’s steps.

2021 closely resembling 2020 and earlier years (2013–2019), reflecting market continuity. In contrast, the tubes network underwent a marked disruption in 2020, followed by a reversion in 2021 toward its pre-2020 structure. Although many firms exited the market between 2019 and 2020, several long-standing participants reappeared in 2021. Thus, the data for 2013–2019 can predict 2021 participants, but not those from the anomalous year of 2020. This evolution illustrates the sensitivity of the Heron metric to shifts in market structure: its earlier peaks captured potential collusive patterns, while the 2020 surge signals renewed concentration that warrants closer monitoring.

For a more in-depth analysis, we tested the Heron Information Coefficient shifts with sliding windows every trimester. By examining the HIC value at a point in time and contrasting it with the expected value given previous behavior, we further stress the temporal aspects of the evolving networks. Here, we can better differentiate the behavior of the gloves and tubes networks.

In the surgical gloves network, shown in Fig. 15, HIC values remain consistently high, but moderate anomalies emerge in 2019 and 2020, marking localized reorganizations rather than systemic disruptions during the pandemic. The smooth upward trend and moderate variability in density and clustering both suggest a gradual structural change, with increasing efficiency achieved through selective, rather than expansive, connectivity. The persistence of multiple components in earlier intervals, followed by consolidation, reinforces the interpretation of slow but steady and selective market concentration.

In the network of endotracheal tubes, seen in Fig. 16, we find a distinct pattern of structural contraction. HIC remains stable at high values for most of the period, but decreases sharply toward the end, coinciding with a reduction in nodes and edges. The presence of strong positive and negative anomaly scores signals alternating phases of expansion and collapse, suggesting a system characterized by transient structural fragility rather than gradual stabilization. Therefore, even though the general trend is the same as the gloves network, its finer details have significant differences.

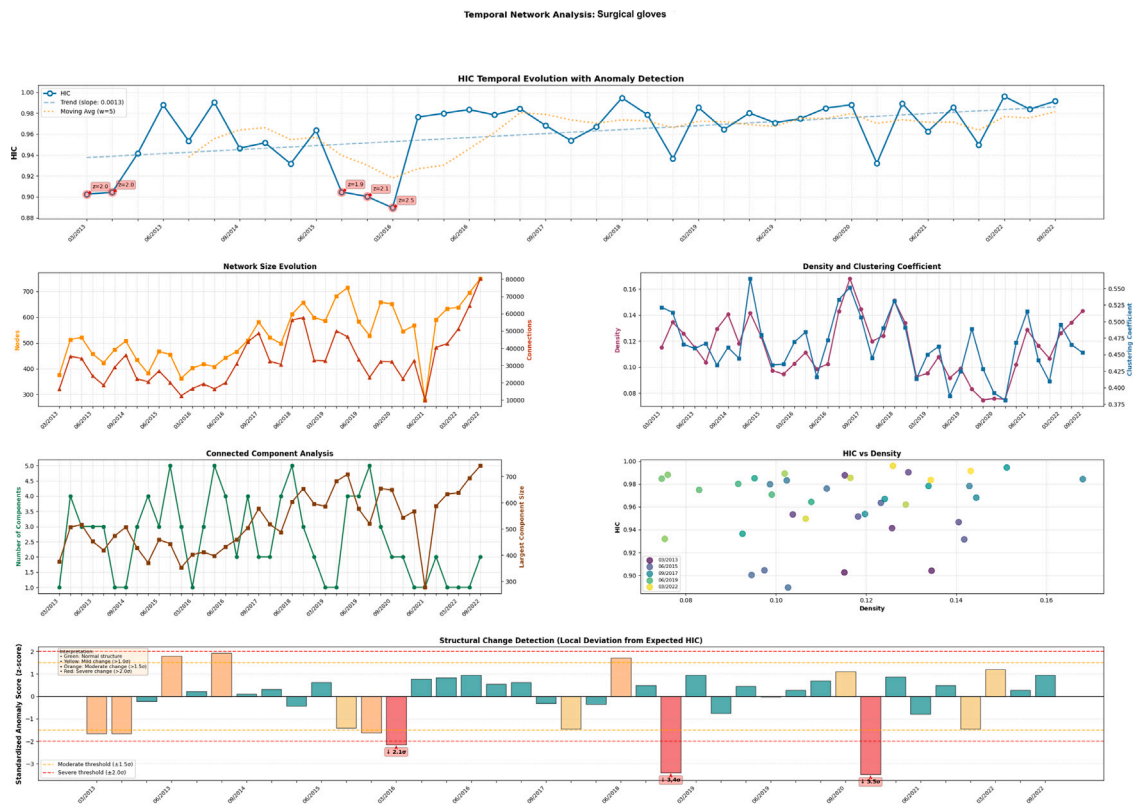


Fig. 15. Summary of sliding windows experiment of the Surgical Glove network from 2013 to 2022, per trimesters. The pandemic had an expressive effect on the network behavior.

The network of masks was the most affected by the pandemic. In 2019, 928 companies were competing. In 2020, this number increased to 2558; in 2021, it decreased to 1673—which is still an expressive increase compared to pre-pandemic levels. Unlike the other networks, the mask network underwent a clear phase transition as seen in Fig. 14. Before the pandemic, the α failed to isolate a subgraph of statistically significant links. In 2020 and 2021, however, the α is reduced, indicating that the remaining links were more structurally relevant. During these years, the proportion of winning companies increased at each filtering step, contrasting with the pre-pandemic pattern where this percentage remained nearly constant. After 2020, significant connections formed preferentially among winning firms, signaling a reorganization of the competitive structure. This drastic shift in the network’s behavior and composition persisted into 2021, involving largely the same dominant players.

More specifically, as shown in Fig. 17, the masks network exhibits both large-scale expansion and pronounced anomalies, in contrast to the other two. A severe deviation in early 2020 coincides with a major increase in network size and connections, reflecting abrupt topological shifts as expected with the pandemic. The subsequent HIC recovery and stabilization indicate rapid adaptation and reorganization toward a more heavily dominated market configuration.

These experimental results show the effects on different networks before and after an unparalleled disturbance. For gloves, there are higher link importance levels in the subgraphs, with increasingly more winners and a gradual differentiation of the participating companies over the analyzed periods. There is a higher link’s importance level for tubes but a less accurate selection of companies in the first pandemic year. Even then, 2021 has a higher percentage of winning companies than before the pandemic and similar elements to previous years after the six steps. Finally, there is a dramatic shift in significance level for masks, winning remaining companies and composition when comparing the pre-pandemic scenario with the pandemic. The pandemic has higher significance levels and percentages of winning companies, as well as similar arrangements in 2020 and 2021 that are very dissimilar to previous years. The sliding-window evolution of HIC effectively mirrors changes in network centrality distributions and topological robustness—high and stable HIC values correspond to configurations with strong central nodes and cohesive connectivity cores, whereas low or volatile HIC signals fragmentation and vulnerability to structural perturbations.

Overall, our method captures the significance of the remaining links and elements in the system. Allied with the composition comparison with previous years, this approach composes a detailed characterization of the system and its dynamics. Across all three cases, the inverse relationship between HIC and density remains consistent. Moreover, the method’s anomaly detection accurately pinpoints critical transition points, distinguishing the gradual evolution seen with surgical gloves and endotracheal tubes from more abrupt reconfigurations, as seen with masks. These results confirm the robustness and interpretability of our HIC-based temporal



Fig. 16. Summary of sliding windows experiment of the Endotracheal Tube network from 2013 to 2022, per trimesters. Its pre-pandemic behavior is quite different from the other networks analyzed.

framework. Our method effectively captures both smooth and disruptive structural dynamics and identifies critical change intervals, providing a useful tool for covert networks detection.

This work establishes Heron’s Information Coefficient (HIC) as a robust metric for quantifying structural imbalances within dynamic networks. Applied to Brazilian health procurement data during the COVID-19 pandemic, HIC successfully isolated topological anomalies consistent with collusive behavior and market concentration—findings further corroborated by stress-testing on synthetic graphs with varying corruption densities. While the method’s geometric intuition effectively exposes deviations often smoothed over by conventional statistics, we emphasize that it functions best as a complementary diagnostic rather than a standalone proof of illicit activity, particularly given the inherent scarcity of ground-truth criminal data. Future research should focus on benchmarking HIC against established information-theoretic metrics and developing real-time monitoring pipelines, thereby broadening its applicability to procurement integrity, fraud detection, and the resilience of complex systems.

5. Conclusion

This work introduces Heron’s Information Coefficient as a robust measure to quantify structural imbalances in dynamic networks. Applied to Brazilian health procurement data spanning the COVID-19 pandemic, HIC successfully isolated topological anomalies consistent with collusive behavior and market concentration—findings further corroborated by stress-testing on synthetic graphs with varying corruption densities and an aggregate procurement dataset.

While promising, this approach should be treated as complementary to existing detection frameworks and not as a stand-alone diagnostic. The method’s geometric intuition helps interpret deviations that conventional measures might smooth over, but it does not capture all institutional or legal nuances. As in many studies of covert criminal networks, our work also faces limited access to reliable real-world data for broader validation. We emphasize that our method functions best as a complementary diagnostic rather than a standalone proof of illicit activity, particularly given the inherent scarcity of ground-truth criminal data and the ever-evolving dynamics of criminal activity.

Future studies should systematically compare the Heron coefficient with other dissimilarity or entropy-based metrics and explore its use in other anomaly detection scenarios. We believe that this measure can add value to research on procurement integrity, fraud prevention, the resilience of complex systems, as well as, more broadly, community detection.

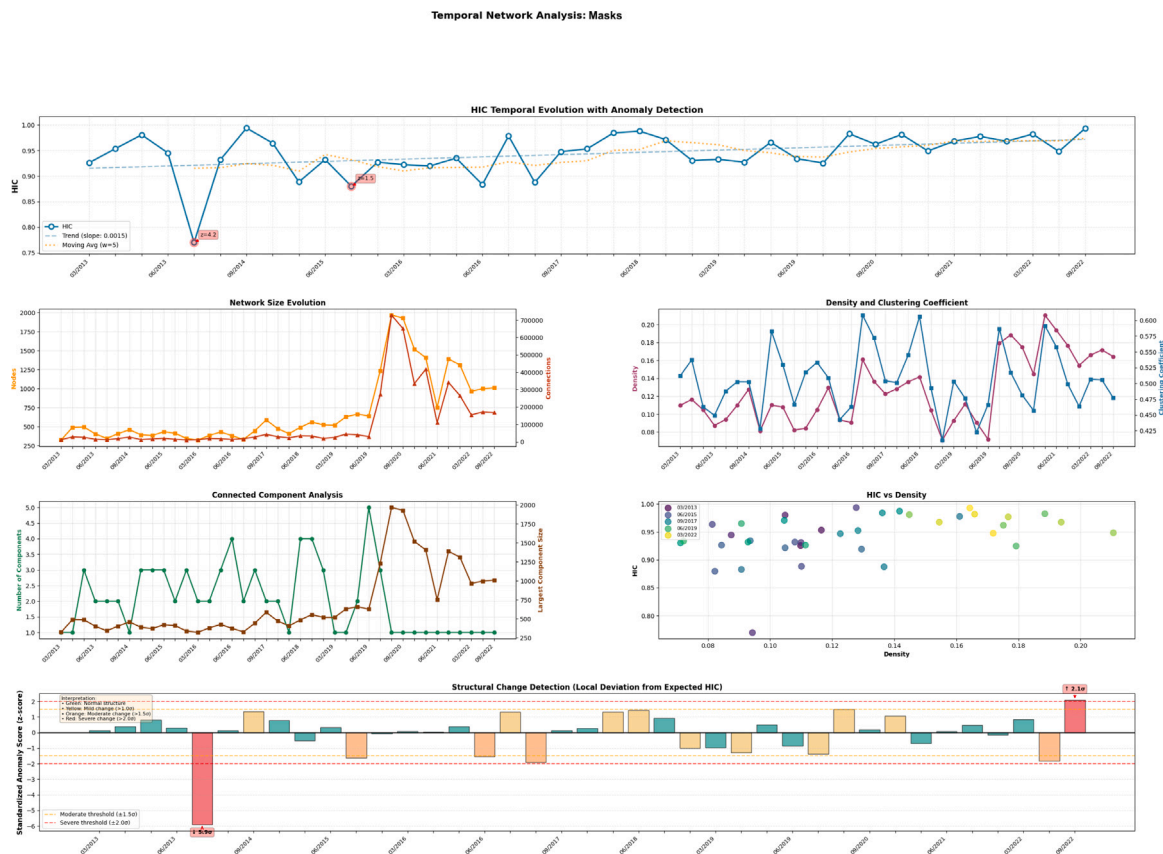


Fig. 17. Summary of sliding windows experiment of the Surgical Mask network from 2013 to 2022, per trimesters. This network was particularly affected by the pandemic.

CRedit authorship contribution statement

Allana Tavares Bastos: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tiago Alves Schieber:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Renato Hadad:** Writing – original draft, Data curation, Conceptualization. **Laura C. Carpi:** Writing – review & editing, Writing – original draft, Validation. **Martín Gómez Ravetti:** Writing – review & editing, Validation, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.chaos.2026.118119>.

Data availability

The bidding data used in this study is publicly available from the Brazilian government. We structured the dataset and released it, along with all relevant code, in our GitHub repository.

References

- [1] Teremetskiy V, Duliba Y, Kroitor V, Korchak N, Makarenko O. Corruption and strengthening anti-corruption efforts in healthcare during the pandemic of Covid-19. *Med Leg J* 2021;89(1):25–8. <http://dx.doi.org/10.1177/0025817220971925>, Publisher: SAGE Publications.
- [2] Steingruber S, Kirya M, Jackson D, Mullard S. *Corruption in the time of COVID-19: A double-threat for low income countries*. 2020, p. 18, Bergen: U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute (U4 Brief 2020:6).
- [3] OECD. Exploitative pricing in the time of COVID. 2020, URL <https://www.oecd.org/daf/competition/Exploitative-pricing-in-the-time-of-COVID-19.pdf>.
- [4] Paim J, Travassos C, Almeida C, Bahia L, Macinko J. The Brazilian health system: history, advances, and challenges. *Lancet* 2011;377(9779):1778–97. [http://dx.doi.org/10.1016/S0140-6736\(11\)60054-8](http://dx.doi.org/10.1016/S0140-6736(11)60054-8).
- [5] Puppim de Oliveira JA. Brazilian public administration: Shaping and being shaped by governance and development. *Chin Political Sci Rev* 2017;2(1):7–21. <http://dx.doi.org/10.1007/s41111-017-0052-4>.
- [6] Gragnolati M, Lindelow M, Couttolenc B. Twenty years of health system reform in Brazil: an assessment of the sistema único de saúde. *The World Bank*; 2013, <http://dx.doi.org/10.1596/978-0-8213-9843-2>.
- [7] Viana ALd, da Silva HP, Yi I. Universalizing Health Care in Brazil: Opportunities and challenges. In: *Social policy in a development context*. Palgrave Macmillan UK; 2017, p. 181–211. http://dx.doi.org/10.1057/978-1-137-53377-7_7.
- [8] Collins C, Araujo J, Barbosa J. Decentralising the health sector: issues in Brazil. *Health Policy* 2000;52(2):113–27. [http://dx.doi.org/10.1016/S0168-8510\(00\)00069-5](http://dx.doi.org/10.1016/S0168-8510(00)00069-5).
- [9] Almeida C, Travassos C, Porto S, Labra ME. Health sector reform in Brazil: A case study of inequity. *Int J Health Serv* 2000;30(1):129–62. <http://dx.doi.org/10.2190/NDGW-C2DP-GNF8-HEW8>, Publisher: SAGE Publications Inc.
- [10] de Almeida Botega L, Andrade MV, Guedes GR. Brazilian hospitals' performance: an assessment of the unified health system (SUS). *Health Care Manag Sci* 2020;23(3):443–52. <http://dx.doi.org/10.1007/s10729-020-09505-5>.
- [11] Wachs J, Kertész J. A network approach to cartel detection in public auction markets. *Sci Rep* 2019;9(1):10818. <http://dx.doi.org/10.1038/s41598-019-47198-1>, Number: 1, Publisher: Nature Publishing Group.
- [12] Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. *Nature* 2000;406(6794):378–82. <http://dx.doi.org/10.1038/35019019>.
- [13] Aytac A, Berberler ZNO. Network robustness and residual closeness. *RAIRO - Oper Res* 2018;52(3):839–47. <http://dx.doi.org/10.1051/ro/2016071>.
- [14] Bachmann I, Bustos-Jiménez J, Bustos B. A survey on frameworks used for robustness analysis on interdependent networks. *Complex, Orgname=Hindawi* 2020. <http://dx.doi.org/10.1155/2020/2363514>.
- [15] Otsuka M, Tsugawa S. Robustness of network attack strategies against node sampling and link errors. *PLOS ONE* 2019;14(9):e0221885. <http://dx.doi.org/10.1371/journal.pone.0221885>.
- [16] Smith AM, Pósfai M, Rohden M, González AD, Dueñas-Osorio L, D'Souza RM. Competitive percolation strategies for network recovery. *Sci Rep* 2019;9(1):1–12. <http://dx.doi.org/10.1038/s41598-019-48036-0>.
- [17] Tejedor A, Longjas A, Zaliapin I, Ambroj S, Foufoula-Georgiou E. Network robustness assessed within a dual connectivity framework: joint dynamics of the active and idle networks. *Sci Rep* 2017;7(1):8567. <http://dx.doi.org/10.1038/s41598-017-08714-3>.
- [18] Wang B, Tang H, Guo C, Xiu Z. Entropy optimization of scale-free networks' robustness to random failures. *Phys A* 2006;363(2):591–6. <http://dx.doi.org/10.1016/j.physa.2005.08.025>.
- [19] Wu J, Tan S-Y, Liu Z, Tan Y-J, Lu X. Enhancing structural robustness of scale-free networks by information disturbance. *Sci Rep* 2017;7(1):7559. <http://dx.doi.org/10.1038/s41598-017-07878-2>.
- [20] Zhou F, He X, Yuan Y, Zhang M. Influence of Interlink Topology on Multilayer Network Robustness. *Sustainability* 2020;12(3):1202. <http://dx.doi.org/10.3390/su12031202>.
- [21] Schieber TA, Carpi L, Díaz-Guilera A, Pardalos PM, Masoller C, Ravetti MG. Quantification of network structural dissimilarities. *Nat Commun* 2017;8(1):1–10. <http://dx.doi.org/10.1038/ncomms13928>.
- [22] Lin J. Divergence measures based on the Shannon entropy. *IEEE Trans Inform Theory* 1991;37:145–51. <http://dx.doi.org/10.1109/18.61115>.
- [23] Klette R, Rosenfeld A. *Digital geometry*. Elsevier; 2004, p. 77–116. <http://dx.doi.org/10.1016/b978-1-55860-861-0.x5000-7>.
- [24] Erdős P, Renyi A. On the evolution of random graphs. *Bull Inst Intern Stat* 1960;38.
- [25] Cavallaro L, Ficara A, Meo PD, Fiumara G, Catanese S, Bagdasar O, Song W, Liotta A. Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia. *PLOS ONE* 2020;15(8):e0236476. <http://dx.doi.org/10.1371/journal.pone.0236476>, 5 de ago. de.
- [26] Ficara A, Currier F, Fiumara G, De Meo P, Liotta A. Covert Network Construction, Disruption, and Resilience: A Survey. *Mathematics* 2022;10(16):2929. <http://dx.doi.org/10.3390/math10162929>.
- [27] Akoglu L, Tong H, Koutra D. Graph based anomaly detection and description: A survey. *Data Min Knowl Discov* 2015;29(3):626–88. <http://dx.doi.org/10.1007/s10618-014-0365-y>.
- [28] Ma X, Wu J, Xue S, Yang J, Zhou C, Sheng QZ, Xiong H, Akoglu L. A Comprehensive Survey on Graph Anomaly Detection With Deep Learning. *IEEE Trans Knowl Data Eng* 2023;35(12):12012–38. <http://dx.doi.org/10.1109/TKDE.2021.3118815>.
- [29] Campana P. Explaining criminal networks: Strategies and potential pitfalls. *Methodol Innov* 2016;9:2059799115622748. <http://dx.doi.org/10.1177/2059799115622748>.
- [30] García Rodríguez MJ, Rodríguez-Montequín V, Ballesteros-Pérez P, Love PED, Signor R. Collusion detection in public procurement auctions with machine learning algorithms. *Autom Constr* 2022;133:104047. <http://dx.doi.org/10.1016/j.autcon.2021.104047>.
- [31] Vian T. Review of corruption in the health sector: theory, methods and interventions. *Health Policy Plan* 2008;23(2):83–94. <http://dx.doi.org/10.1093/heapol/czm048>.
- [32] Serrano MÁ, Boguñá M, Vespignani A. Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci* 2009;106(16):6483–8. <http://dx.doi.org/10.1073/pnas.0808904106>.
- [33] Akbarpour M, Jackson MO. Diffusion in networks and the virtue of burstiness. *Proc Natl Acad Sci* 2018-07-24;115(30):E6996–7004. <http://dx.doi.org/10.1073/pnas.1722089115>, Publisher: National Academy of Sciences Section: PNAS Plus.
- [34] Elsisy A, Mandviwalla A, Szymanski BK, Sharkey T. A network generator for covert network structures. *Inform Sci* 2022;584:387–98. <http://dx.doi.org/10.1016/j.ins.2021.10.066>.
- [35] IBGE. *Projeção da população*. 2021, <https://www.ibge.gov.br/apps/populacao/projecao/index.html>.
- [36] Mattos C. Procurement procedures and bid-rigging in Brazil. In: *Competition law enforcement in the BRICS and in developing countries: legal and economic aspects*. International law and economics, Springer International Publishing; 2016, p. 169–85. http://dx.doi.org/10.1007/978-3-319-30948-4_6.
- [37] Kirya M. *Anti-corruption in Covid-19 preparedness and response*. Bergen: U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute; 2020, (U4 Brief 2020:8).
- [38] Wessel L. 'It's a nightmare.' How Brazilian scientists became ensnared in chloroquine politics. *Sci Mag* 2020. <http://dx.doi.org/10.1126/science.abd4620>.
- [39] Adler M. *Exército e Ministério da Saúde gastaram milhões para distribuir cloroquina*. 2021, Estado de Minas.
- [40] Darlington JB. How Brazil gambled on unproven drugs to fight Covid-19. 2021-03-14, CNN..
- [41] Kohler JC, Mitsakakis N, Saadat F, Byng D, Martinez MG. Does pharmaceutical pricing transparency matter? Examining Brazil's public procurement system. *Glob Health* 2015;11(1):34. <http://dx.doi.org/10.1186/s12992-015-0118-8>.
- [42] Harrington J. Detecting cartels. *The Johns Hopkins University, Department of Economics*; 2005, URL <https://econpapers.repec.org/paper/jhupapers/526.htm>.
- [43] Harrington J, Chen J. Cartel pricing dynamics with cost variability and endogenous buyer detection. *Int J Ind Organ* 2006;1185–212. <http://dx.doi.org/10.1016/j.ijindorg.2006.04.012>.
- [44] CGU. *Licitações realizadas* - Portal da transparência. 2021, <https://www.transparencia.gov.br/download-de-dados/licitacoes>.