Gao *et al. Applied Network Science* (2025) 10:62
https://doi.org/10.1007/s41109-025-00740-9

Applied Network Science

RESEARCH Open Access

Heterogeneity- and homophily-induced vulnerability of a P2P network formation model: the Mana based auto-peering protocol

Yu Gao[1*], Carlo Campajola[2,3,5†], Nicolò Vallarano[1,2†], Andreia Sofia Teixeira[4,6†] and Claudio J. Tessone[1,2†]

†Carlo Campajola, Nicolò Vallarano, Andreia Sofia Teixeira, and Claudio J. Tessone have contributed equally to this work.

*Correspondence:
Yu Gao
yu.gao2@uzh.ch

Full list of author information is available at the end of the article

Abstract

IOTA is a distributed ledger technology that relies on a peer-to-peer (P2P) network for communications. Recently an auto-peering algorithm was proposed to build connections among IOTA peers according to their "Mana" endowment, which is an IOTA internal reputation system. This paper's goal is to detect potential vulnerabilities and evaluate the resilience of the P2P network generated using IOTA auto-peering algorithm against eclipse attacks. In order to do so, we interpret IOTA's auto-peering algorithm as a random network formation model and employ different network metrics to identify cost-efficient partitions of the network. As a result, we present a potential strategy that an attacker can use to eclipse a significant part of the network, providing estimates of costs and potential damage caused by the attack. On the side, we provide an analysis of the properties of IOTA auto-peering network ensemble, as an interesting class of homophile random networks in between 1D lattices and regular Poisson graphs.

Keywords IOTA, Peer-to-peer, Mana, Eclipse attack, Networks

Introduction

Over the last few decades peer-to-peer networks (P2P) have been widely employed in a variety of settings as a base layer for distributed systems, including the worldwide web(Lawrence and Giles 2000), file sharing (Oram 2001), instant messaging in social networks (Cancho et al. 2001; Heidemann et al. 2012; Masinde and Graffi 2020), and distributed computing (Tanenbaum 2007). In contrast to the client–server system, P2P networks stand out as more egalitarian and decentralised, where peers share data resources with each other without resorting to a central authoritative node (Androutsellis-Theotokis and Spinellis 2004). Particularly, the emergence of Bitcoin (Nakamoto 2008) has brought revolutionary technology in socio-technical and socio-economic systems, and the decentralized P2P network layer is the backbone of its Distributed Ledger Technology (DLT).

A DLT is a consensus-driven system where digital data is replicated, shared, and synchronised across multiple sites, countries, or institutions in the absence of a central

authority, ensuring geographic decentralisation. The process that enables the reaching of distributed agreement is denoted as the consensus mechanism. A DLT is said to be “in consensus" when the participants of the network agree on the replicated information contained in the distributed ledger. A blockchain is a particular example of DLT, with Bitcoin (Nakamoto 2008), the first type of DLT to reach worldwide diffusion, using a blockchain for transaction storing. Bitcoin also introduced the consensus mechanism commonly known as Proof-of-Work, where users solve complex problems in order to mint new tokens and, eventually, ensure the consensus of the system.

We can think of the P2P network in DLTs as the infrastructure that enables the distribution of data among peers, and the establishment of consensus is significantly influenced by the topology of connections within the P2P network (El Ioini and Pahl 2018; Kraner et al. 2023). The nodes in a generic P2P network are computers, and the information they contain and exchange can be files, contracts, and transaction records, to name a few (Wood 2014). To give an example, if we think of DLTs storing specifically economic transactions like Bitcoin, the “miners" i.e. the peers driving consensus play the role of clerks in a bank: they clear the systems transactions in order for the everyday Bitcoin user to be able to execute their transactions. Their incentive in doing so is the coinbase reward associated with each block attached to the blockchain, which also serves as a Bitcoin minting mechanism.

P2P networks

Due to the growing adoption of cryptocurrencies, numerous researchers have extensively explored their P2P networks. One of the main challenges in studying real-world P2P networks follows from their decentralised nature: they rarely can be fully monitored by a single agent, as each node only knows about its peers and has no reliable information beyond them, nor are they incentivised to share this information as it may lead to malicious attacks. To address this problem, several scholars have proposed methodologies for mapping these networks and conducting subsequent analyses. To cite a few, Miller introduced “AddressProbe" to uncover the topology of the Bitcoin P2P network (Miller et al. 2015); Deshpande et al. devised a framework to map the Bitcoin P2P network, presenting an insightful snapshot of its topology (Deshpande et al. 2018); Kim et al. developed NodeFinder, a tool to measure the Ethereum P2P network’s characteristics and dissect the intricate DEVp2p ecosystem (Kim et al. 2018). From a different viewpoint, Imtiaz et al. revealed the intermittent network connectivity displayed by Bitcoin nodes (Imtiaz et al. 2019).

Notably, there exists research concerning attacks on P2P networks: Heilman et al. (2015) proposed employing extremely low-rate TCP connections for network attacks, while Nayak et al. (2016) devised a combination of mining attacks and network-level attacks to undermine the consensus in DLT protocols. Additionally, other papers outline a variety of technical approaches to network attacks (Apostolaki et al. 2017; Tran et al. 2020).

In P2P networks, nodes actively communicate with nearby nodes, facilitating the exchange of information. This dynamic is crucial for maintaining consensus among nodes and deterring potential efforts by individuals or attackers to gain complete control of the network. Despite the original intention of most DLT systems’ P2P networks to build random, uncontrollable networks, vulnerabilities can arise when certain network information is exploited (Müller et al. 2021; Moubarak et al. 2018). To prevent such

attacks, it is essential to have a comprehensive understanding of the inherent topology of the empirical P2P network within a static network model, achieved by substituting technical cryptographic elements with graph-based parameters (Dobson et al. 2007; Arianos et al. 2009; Epstein 2012; Hines et al. 2010). This approach can uncover potential weaknesses in the design of P2P network protocols.

Network Science offers a comprehensive tool-set to effectively model these types of complex systems (Newman 2018; Watts and Strogatz 1998). Socio-technical systems have been largely studied in the context of complex networks (Vespignani 2012; Watts et al. 2005): translating such systems into network representations allows us to exploit their topological properties in order to identify vulnerabilities and nonlinear properties. P2P protocols have a natural network representation, where peers can be represented as nodes, while edges embody peer connections. An obstacle to a straightforward application of network analysis to these systems is that by design P2P protocols make the connection construction unknown globally, as peers only know their local connectivity. On the contrary, peers generally have a huge security incentive not to reveal their connections to possible malicious adversaries (Heilman et al. 2015; Singh et al. 2006). A possible approach to this challenge is to define theoretical models that describe P2P networks, starting from the P2P protocol rules, in order to gain further insight.

IOTA

The prevailing research naturally focuses on Bitcoin and Ethereum, both operating in a permissionless manner. Generally, these platforms incentivise validators—i.e. the peers tasked with verifying transactions and upholding network integrity—by issuing rewards for their honest contribution and penalties if their actions contradict rules or network interests. However, for platforms lacking such reward and penalty systems for peers like IOTA,[1] the design of the peer connection algorithms can have a profound influence on the consensus integrity.

IOTA has been introduced to address one of the main limitations of mainstream blockchains, namely their scalability. Blockchain addresses the challenge of decentralising digital transaction systems while enhancing data transparency and security. However, it may encounter scalability issues when managing a large volume of transactions (Khan et al. 2021). Each node is required to verify and store a complete copy of the blockchain, leading to limited network capacity and slower transaction processing; in addition to this, blockchains have an intrinsic transaction volume limit, depending on how many transactions can fit inside a block and the time that occurs between two blocks. To tackle this, IOTA introduced the Tangle (Popov 2018), a new ledger data structure based on a directed acyclic graph (DAG) instead of the classic linear chain. By enabling concurrent transaction processing, the Tangle can theoretically handle increased transaction throughput as the network expands. Another peculiarity of IOTA is "*Mana*", a scarce quantity associated with a peer that would tokenize its reputation, which has been proposed to be introduced in an upcoming upgrade (Popov et al. 2020). Mana is intended as a Sybil protection mechanism (Douceur 2002; Cheng and Friedman 2005), i.e. a mechanism that sets a cost (in this case, acquiring reputation) for nodes to participate in the consensus formation. This prevents Sybil attacks, where malicious agents are able to manipulate consensus by the creation of a large number of nodes.

[1] https://www.iota.org/.

It was proposed to use Mana within a pseudo-random P2P network formation model, the IOTA auto-peering module (Müller et al. 2021), which favors homophilic connections, i.e. among nodes in a close Mana range. Homophily is known in sociology as the tendency of individuals to bind by similarities. In network theory, we observe homophily when nodes tend to connect preferentially with nodes with similar characteristics (Newman 2002; Kossinets and Watts 2009; Shalizi and Thomas 2011; Hasheminezhad and Brandes 2023). Many social networks demonstrate connection mechanisms in which each node possesses a distinct trait, such as wealth, age, or trustworthiness. Connections in these networks arise from similarities, as observed in friendship networks (Moody 2001), collaboration networks (Newman 2001) or actor networks (Watts 1999).

Overview

In the present work, we develop a set of strategies to identify a cost-effective set of nodes that, if controlled by a malicious actor, can functionally disconnect an homophily induced P2P network. In order to provide an empirical analysis, we develop our strategies on a real case example, the proposed IOTA P2P network formation model,[2] and we additionally develop an ulterior attack strategy refined trough the simulations results.

While the proposed results are tailored for the IOTA case, we observe that with minimal additional work, the same strategies apply to any kind of P2P homophilic network.

The rest of the paper is organised as follows: Section II formally describes IOTA network formation model and the attack strategies. Section III presents the strategies simulation's results. Section IV concludes the paper with the result's comments and discussion.

Approach

Mana

The *auto-peering module*, as described by IOTA[2], generates a P2P network composed of N nodes. Each node has a given amount of Mana and, for convenience, we will index them from the largest endowment of Mana (identified as 1) to the lowest (identified as N). These Mana values are publicly known to all nodes and regularly broadcast over the P2P network itself. The distribution of Mana is the result of a dynamical process, where nodes can accrue their endowment through fair participation in the consensus mechanism, and their chance to do so is itself a function of their Mana endowment; in our setting we assume Mana to be an equilibrium value, not depending on time. As is commonly found in similar systems like Proof-of-Work mining (Campajola et al. 2022; Makarov and Schoar 2021), Proof-of-Stake validation (Chohan 2022; Sai et al. 2021) or any system where the growth of a quantity is proportional to its current level, the so-called "Matthew effect" enters the picture and a Zipf's law often emerges after sufficient time. For this reason, we assume the Mana distribution follows a Zipf (2016) with exponent s so that the Mana value of node i is

$$m(i) = Ki^{-s}, \tag{1}$$

where $m(i)$ is the Mana endowment of the i-th ranked node and K is an arbitrary positive constant used for numeric stability, in our case $K = 10^{10}$.

[2] https://github.com/iotaledger/autopeering-sim.

Peer discovery

The protocol prescribes that each node i will add node $j > i$ to its list of eligible neighbours, $\mathcal{N}_i$, if $m(i) < \rho m(j)$, where ρ is a given constant. If this condition is not met for at least R other nodes, i will add the R nodes ranked after itself to the list. Any time a node j is added to the set $\mathcal{N}_i$, i will be added to $\mathcal{N}_j$ to reciprocate. Therefore, the potential neighbours of any node i are formally identified as

$$\mathcal{N}_i = \left\{ j = 1, \ldots, N \,:\, \frac{1}{\rho} m(i) < m(j) < \rho m(i) \;\vee\; |j - i| < R \right\}. \tag{2}$$

From this set, i randomly picks k elements to link to and receives k links from other nodes in the set. The result is a random regular network with coordination number k (or in other words where each node has degree $2k$), and links appear only between nodes with similar Mana endowment. From a modelling perspective, N, k, R, ρ and the Zipf's exponent s are the parameters needed to generate a network.

Network formation

We base our simulation on the publicly available Go code by the IOTA research team.[3] A node will compile its potential neighbours' list and cease searching for neighbours when it has established k connections, while it accepts connection requests from other nodes until it has received another k connections. Communication is then bidirectional along these links, so the resulting network is undirected. As a result, each node ultimately has $2k$ neighbours.

Combining the aforementioned rules, we can deduce that an increase in the Zipf's exponent s leads to the most affluent nodes having relatively short potential neighbour lists. This happens because of the increased inequality in the distribution of Mana among nodes, meaning a larger Mana gap between nodes. The parameters R and ρ, although not as influential, still impact the IOTA P2P network and contribute to its structure by tuning the tolerance that the nodes have in terms of the Mana endowment of their potential neighbours. As can be seen by the definition of Eqs. 1 and 2, when s is small the distribution is relatively homogeneous and peers easily have many potential neighbours through the range condition set by ρ, forming a network that is closely resembling a regular (i.e. where all nodes have the same degree) version of the Erdős-Rényi (ER) random network. However, as s increases, nodes that fail to find enough peers within the Mana ratio condition can acquire sufficient potential neighbours through the alternative condition based on R. In this situation, peers that have fewer potential neighbours can solely establish connections with peers that are in their immediate vicinity in terms of Mana ranking. Consequently, the network will tend towards resembling a k-regular 1D lattice network.

To have a more intuitive understanding of the IOTA auto-peering model, we provide a visualisation of networks generated under various s and ρ values in Fig. 1. As it can be seen, when the Mana (shown in colour in log scale) is relatively homogeneous and when the tolerance parameter ρ is high enough the network does not show any particular asymmetry, but in a low ρ scenario and whenever s is increased there is a clear orientation of the network, with the nodes with the largest Mana endowment being shifted

[3] https://github.com/iotaledger/autopeering-sim.

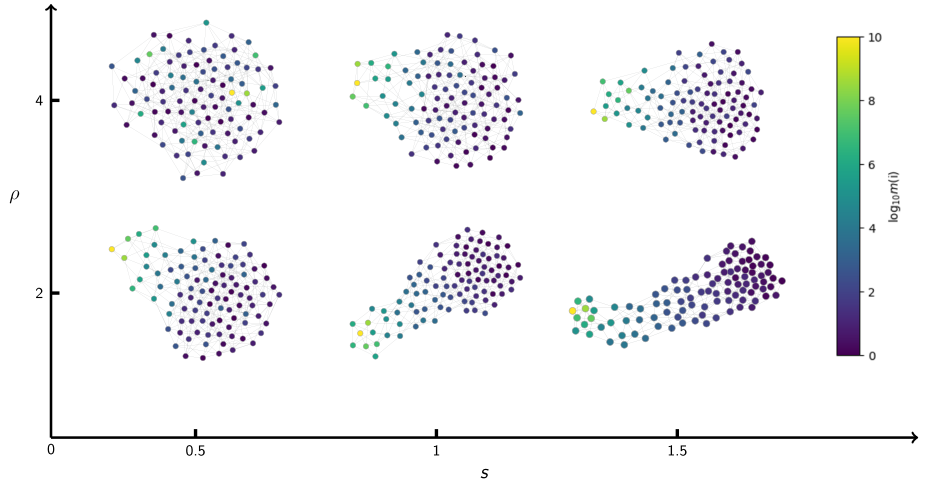

Fig. 1 Intuitive visualisation of the different topologies realised by the IOTA auto-peering formation model as a function of ρ and s. The graphs are generated with fixed $N = 100$, $R = 10$, $k = 4$. Colour represents the Mana endowment of nodes, in a logarithmic scale

towards one end of the network and the shape becomes more linear, following the nodes Mana hierarchy.

Attack damage and cost

Figure 1 neatly highlights the intuition behind a potential attack strategy for this communication network. When s is large and ρ is small, the elongated shape taken by the network results in the creation of a choke point separating the few high Mana nodes from the many low Mana ones.

In an eclipse attack, the attacker's goal is to control all the victim's incoming and outgoing communications, isolating the target from the rest of the peers and destroying the ability of all peers to reach a consensus (Singh et al. 2006). In our work, the goal of the attacker is to achieve control over all links separating two sub-components of the network: this allows the attacker to manipulate the flow of information between the two sub-components, gaining advantage and twisting consensus towards their needs. Since the parameters of the network formation protocol are known to the attacker, they only need to find an allocation of their Mana endowment across multiple Sybil nodes in a way that allows them to control the crucial choke points.

To measure the attack efficiency, we introduce the concept of damage D and cost x. The damage is measured in terms of the fraction of Mana the attacker is able to disconnect from the network: if the total Mana in the system is denoted as $M = \sum_i m(i)$ and the network nodes are split into two components A and B where without loss of generality $\sum_{i\in A} m(i) \leq \sum_{j\in B} m(j)$, then

$$D(A) = \sum_{i\in A} m(i)/M. \tag{3}$$

Clearly, the maximum attainable damage is then $D_{max} = 1/2$.

To control a link, the attacker needs to control one of the endpoint nodes. The cost of controlling a link (i, j) is defined as the minimum Mana of its endpoint nodes $x(i, j) = \min(m(i), m(j))$. Then, for a set of links $C(A, B)$ which

separates partition A from B and the set of least Mana endpoints (the "frontier") $F = \{\arg\min(m(i), m(j))\ \forall (i,j) \in C(A,B)\}$, the cost of attack reads

$$x(C) = \sum_{i \in F} m(i). \tag{4}$$

The goal of the attacker is, for a given Mana cost x_{max} they are able to commit, to find the optimal set C such that $x(C) \leq x_{max}$ and $x(C') > x_{max} \forall C'$ such that $D[x(C')] \geq D[x(C)]$.

In the following, we propose two simplified strategies for the attacker: a "Betweenness" strategy and a "Greedy" strategy, which both require full information about the P2P structure, and a derived "Blind" strategy that only requires the more realistic assumption of knowledge of the parameters used to generate the network.

Attack strategies

To investigate the security of the IOTA P2P network, we develop two attack strategies: the "Betweenness" strategy and the "Greedy" strategy. To quantify the impact of the attack, we introduced in the previous section the concepts of "Damage" and "Cost", whose ratio we employ as a measure of attack efficiency.

Although both strategies rely on complete topological information about the network, it is important to note that in real-world scenarios the specific P2P network topology is not publicly accessible. Therefore, to address this limitation, we also propose a "Blind" strategy. This strategy is inspired by the results of the simulations of the "Betweenness" and "Greedy" strategies, but it is executed assuming the attacker has no knowledge of the actual topology of the P2P network except for the knowledge of the Mana distribution and the IOTA auto-peering network formation model parameters, which are public information available to all IOTA nodes.

Betweenness strategy

The betweenness strategy, as the name suggests, relies on edge betweenness centrality to identify the optimal cut. Similar to node betweenness, edge betweenness serves as a metric that quantifies the fraction of shortest paths passing through a specific edge between any two nodes within a network (Freeman 1977). Inspired by the Girvan-Newman algorithm for community detection (Girvan and Newman 2002), the strategy prescribes removing edges from the network starting from the ones with the highest betweenness centrality. However, unlike the community detection process, in this scenario, the removal of the top betweenness links is halted once the network is divided into two parts. It is crucial to note that after each removal of the top betweenness link, the link betweenness is recalculated for the entire network. The attacker proceeds iteratively, removing the top betweenness links until the original network is fragmented into two distinct sub-networks, denoted as A and B. The removed links constitute the associated cut C, and the attack damage D is the total Mana associated to the nodes in the subnetwork deriving from cut C, which has the smaller total Mana share.

Greedy strategy

In the literature, we can find many different instances of network partitioning algorithms, such as spectral partitioning (Fiedler 1973), clustering coefficient (Saramäki et al. 2007), community detection (Yang et al. 2016) or heuristic based partitioning (Kernighan and Lin 1970). However, none of these algorithms takes advantage of node

attributes, such as the Mana endowment. Here, we introduce a simple method to split the auto-peering network according to the Mana ranking information and the network topological information, which we call the “Greedy" strategy.

The Greedy strategy is much more simplistic than the Betwenness one: the attacker splits the nodes into two sets $A = \{i \leq i^*\}$, and $B = \{i > i^*\}$ for each possible target node $i^* = 1, \dots N-1$. It follows that the cut is the set of links connecting A and B. For each i^* we calculate the cost of the split and choose the i^* that maximises damage per unit cost to run the attack.

Blind strategy

In practical situations, the specific topology of the P2P network is not available, as nodes do not have a full picture of all connections in the network but only know their own peers. A blind strategy, one that does not require full knowledge about the network topology, is therefore needed. We design a blind strategy that is *informed* by the results obtained with the Betweenness and Greedy strategies, namely measuring the frequency with which a specific node, given the network formation parameters, ends up being in the control set identified by the full-information strategy. We then take the most frequently targeted node i_σ, where $\sigma \in \{B, G\}$ (for Betweenness and Greedy, respectively) and define the control set for the strategy $\mathcal{C}_{\text{blind}}$ as all nodes within a range L in the Mana ranking from node i_σ, i.e.

$$\mathcal{C}_{\text{blind}}(\sigma) = \{i \ : \ |i_\sigma - i| < L\} \tag{5}$$

The attacker is assumed to take control of this set of nodes and we verify whether this was successful to split the network and quantify the attack efficiency according to our Damage and Cost measures.

Results

Simulation results for full information attacks

We explore the parameter space and find that the resulting network structure is most sensitive to the value of Zipf’s exponent s, which describes Mana heterogeneity (see Sect. 4.3). We fix all other parameters to reasonable values $N = 100$, $R = 10$, $\rho = 4$, $k = 4$, and focus on the attack vulnerability varying s from 0.5 to 1.5 at intervals of 0.1. For each set of parameters, we generate a sample of 1000 networks and apply our attack strategies, measuring the damage over cost ratio D/x.

In Fig. 2, we present the results for the average damage over cost $\mathbb{E}[D/x]$ over 1000 simulations and across varying values of s for the full information attacks. We observe that both the Betweenness and Greedy strategies, respectively represented by the blue and orange violins, attain maximum efficiency when $s \approx 1$. Additionally, we show the result of applying the Betweenness strategy to a random-regular network where Mana does not have a role in the formation of connections, i.e. the fully randomised Watts-Strogatz (WS) model of small-world networks (Watts and Strogatz 1998) with coordination number k. Compared to the auto-peering network formation model, the attack efficiency in the WS model appears to be much closer to 0 in the vast majority of cases. The result in Fig. 2 shows that in specific Mana distribution conditions, the attacker can achieve relatively large damage to the network with a small cost and that this vulnerability is induced by the presence of Mana in the network formation protocol. Both strategies achieve maximal $D/x \approx 3.5$ for $s = 1$: this is particularly relevant because Zipf laws

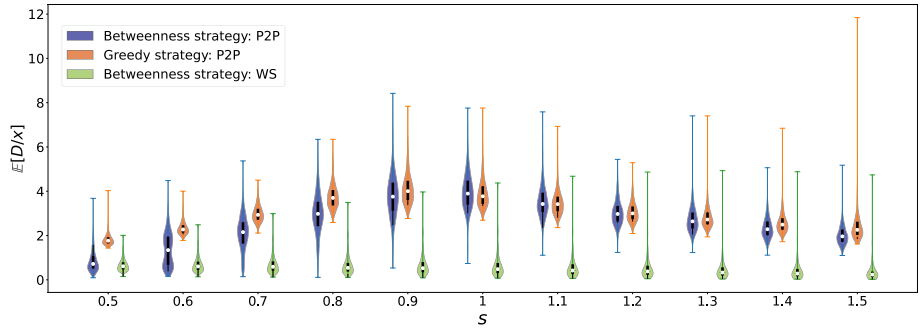

Fig. 2 The average damage over cost ratio by attack strategy and underlying network formation model. Each auto-peering generated network has $N = 100$, $\rho = 4, R = 10$, $k = 4$. WS generated networks have $N = 100, k = 4, p = 1$, the data size is 1000 graphs

with exponent in the close range around 1 have been observed to describe wealth distribution in cryptocurrency token holdings in the literature (Kusmierz and Overko 2022).

Next, we compare the frequency of endpoint removals between the betweenness strategy and the greedy strategy under the conditions of $\rho = 4$, $R = 10$, and $N = 100$, based on a sample of 1000 graphs. Our findings, which for the sake of conciseness we report in Figs. 7 and 8 in the Appendix, reveal that the greedy strategy, which aims to optimise the D/x ratio, typically removes node $i_G = 14$ as the maximum frequency endpoint. On the other hand, for the betweenness strategy, the maximum frequency endpoint that is removed is node $i_B = 12$. We find that in the case of $s = 1$, the endpoints selected with maximum frequency F by both the "betweenness" strategy and the "greedy" strategy are typically in the same range, albeit not precisely the same nodes.

We are going to use the maximum frequency endpoints hereby identified to inform the blind attacks analysed in the next section.

Simulation results for blind attacks

A pitfall of the two strategies above is that they both require the attacker to have perfect knowledge of the network structure, which is unrealistic. A more realistic relaxation of this assumption is that the attacker is only aware of the network formation model parameters. Given such information, the attacker can reproduce the above results and inform a "Blind" strategy, aiming to control the nodes that are most often needed to perform the split. In particular, for the specific case of $N = 100$, we take the "target" nodes $i_B = 12$ and $i_G = 14$ (for the Betweenness and Greedy strategy respectively) and define the parameter L such that the blind attacker controls nodes $\mathcal{C}_{\text{blind}}(\sigma)$ for $\sigma \in \{B, G\}$. We then measure the D/x ratio over 1000 simulated networks along with the *success ratio* p, i.e. the fraction of simulations where the network is successfully split by the blind attacker, and show the results in Fig. 3 as a function of L.

The red and yellow lines show $\mathbb{E}[D/x]$, the average damage over cost attack efficiency associated with a Blind strategy informed either by the Greedy (red, "BG") or Betweenness (yellow, "BB") results, averaged over a sample of 1000 networks. The blue and green lines track the attack success rate p using the target node from the "Greedy" strategy (blue line) and from the "Betweenness" strategy (green line).

As should be expected, for low choices of L the strategy is unsuccessful in splitting the network. The Blind-Betweenness (BB) strategy reaches a peak average efficiency of $\mathbb{E}[D/x] = 1.75$ for $L_b^* = 7$, corresponding to 100% attack success. On the other hand,

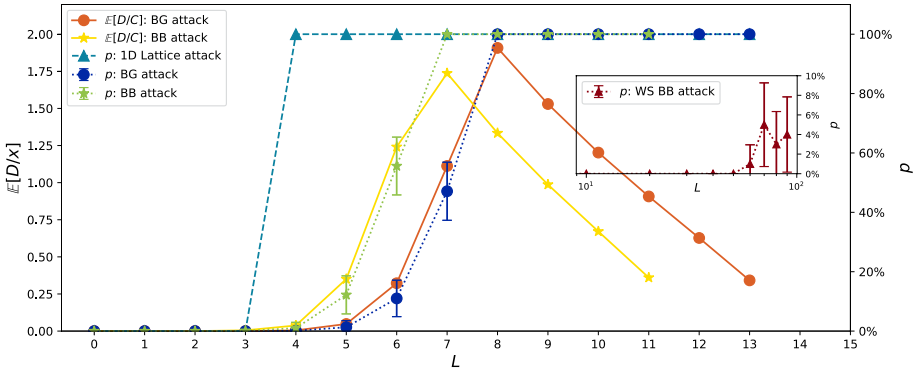

Fig. 3 Success ratio p (on the right y-axis) and average damage $\mathbb{E}[D/x]$ (on the left y-axis) per unit cost of a Blind strategy informed by the Greedy ("BG") and Betweenness ("BB") strategies. The results are averaged over a sample of 1000 simulated networks with $N = 100$, $\rho = 4$, $R = 10$ and $s = 1$ while the error bars are 95% confidence intervals. As comparisons, we also show the success ratio p on an equivalent 1D lattice and a fully randomised regular WS-type network in the inset

the Blind-Greedy (BG) strategy reaches peak efficiency for $L_g^* = 8$: while requiring more nodes to be controlled, the efficiency is larger on average ($\mathbb{E}[D/x] = 2$). Summarising, both strategies allow the attacker to successfully split the network from a low information context. Based on our simulations the BG strategy attains a better efficiency, but we observe that we did not assign an additional cost to the number of controlled nodes, we only considered the Mana required to control the node: the number of nodes to control may affect the cost of an attack in a real-world situation. Interestingly, both strategies at their peak efficiency L command the same cost of around 24% of all the Mana in the system, as shown in Fig. 5. This means that an attacker controlling 24% of the system's Mana would be able to identify a distribution strategy to control the choke points of the P2P network 100% of the times.

To compare the severity of the attacks with some baseline scenarios, we also run the same attacks on a 1D lattice with the same coordination number k as the auto-peering network and on a fully randomised Watts-Strogatz (WS) network model, i.e. a random-regular network of degree $2k$. For the one-dimensional lattice, for which we show the success ratio with the cyan triangles in Fig. 3, it is straightforward to see that the network is always split whenever $L \geq k$, where in this case $k = 4$. Compared with a 1D k-regular lattice, the auto-peering networks are more resilient. This makes perfect sense: 1D lattice are completely predictable networks, where the attacker does not need to acquire information to predict the topology of the network. On the other extreme the WS networks, being fully randomised, are much more resistant to attacks and the networks are almost never successfully split unless L is chosen to be of the same order of magnitude as N. We show this result in the inset of Fig. 3.

This result shows clearly that the IOTA auto-peering protocol produces networks that are in between the fully deterministic 1D lattice and a fully random regular network. Depending on the parameter values, the Mana-based connection rules can severely restrict the options for nodes to connect between themselves, resulting in the loss of the security provided by randomness and consequently in predictable network structures that can be exploited.

Strategy robustness

In order to verify the robustness of our results, we test the proposed strategies by simulating the model on a realistic range in the parameters' space. We choose $\rho \in [1.5, 4]$ and $s \in [0.5, 1.5]$ as they should provide a comprehensive overview of realistic parameters choices.

We find that, as long as N is large enough compared to k, varying the number of nodes has no particular effect on the results. This is mostly due to the fact that the Mana is distributed according to Zipf's law, and so most of the Mana is concentrated in the first few nodes in the ranking regardless. Similarly, we do not find effects of particular interest when changing R within reasonable ranges. For the sake of space, we report additional results testing N and R dependence in Figs. 10 and 11 (see Appendix). For each combination of parameters, the results were averaged over a sample of 1000 graphs generated by the auto-peering network formation model. On the other hand, we find an interesting interplay between the ρ and s parameters, which we discuss in the present section.

In Fig. 4 we show the expected damage over cost ratio across different parameter combinations for the two informed strategies. While the performances of the two strategies are slightly different, we do observe a similar trend: both attain maximum success (in terms of damage over cost) for low ρ and for s between 0.5 and 1, i.e. when the Mana distribution is sufficiently heterogeneous compared to nodes' "tolerance" ρ, but not so

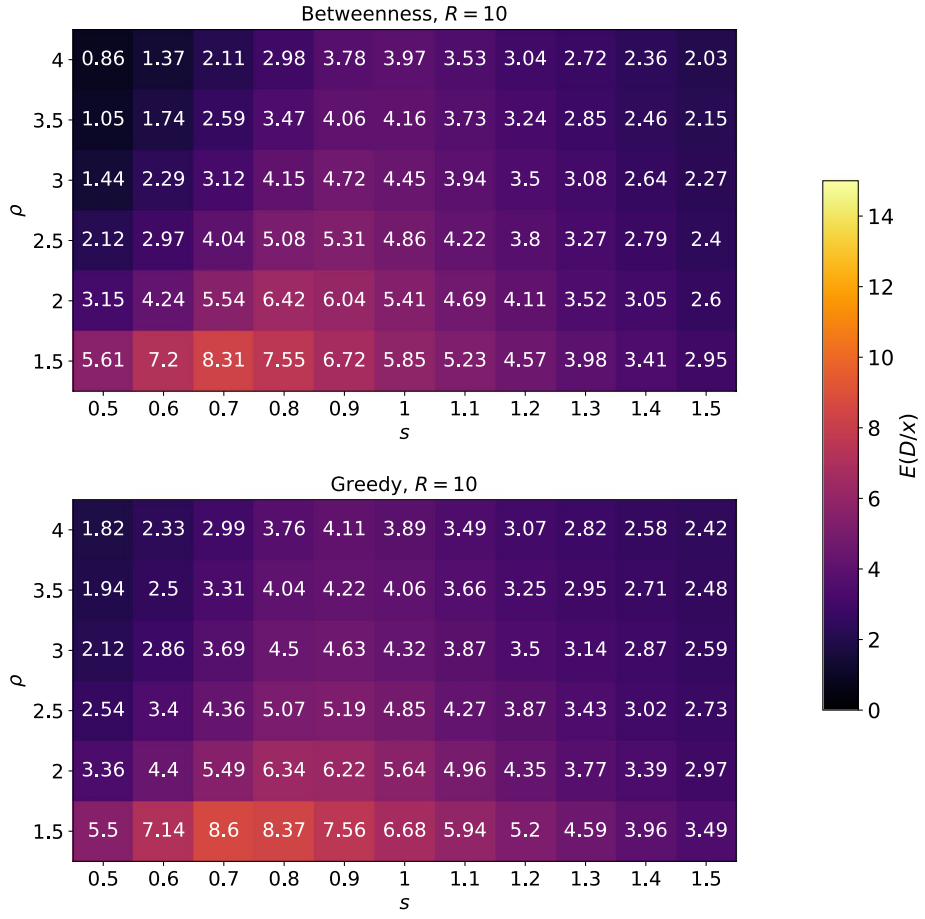

Fig. 4 Heatmaps of the expected damage over cost, $\mathbb{E}[D/x]$, plotted over ρ and s, for the betweenness (up plot) and greedy (down plot) strategy. $R = 10$ is kept constant. The maximum for each fixed s is when ρ is minimal, meaning that whenever the P2P model is more similar to a chain, the network is maximally vulnerable

heterogeneous that attacks can only be successful by controlling the top Mana nodes themselves. At the same time, we observe that ρ has a similar effect on both strategies for all values of s, with a stronger effect when s is low (so when Mana distribution is homogeneous). This is explained by the role of ρ, which is increasing (by a multiplicative factor) the tolerance for each node in terms of Mana difference to its potential neighbours; this means that when the network is homogeneous, a large ρ implies longer range connections, while when s is very large, the randomness effect of ρ is diminished.

In Fig. 5 we present the Mana cost (as a percentage of the total Mana) that is necessary to successfully split the network with 100% success, following a blind attack with the BB or BG strategy. We notice that the regions of the parameters space which deliver the worst output in terms of attack efficiency are also the ones with the largest cost of attack, where graphs generated form the auto-peering model resemble random regular graphs. The cost is monotonically decreasing in s and increasing in ρ for both the blind strategies.

Conclusion

In conclusion, our results point out that a malicious agent controlling just over 24% of the total Mana can divide their Mana into a set of nodes that will most likely act as choke points of the P2P network from the auto-peering network formation model, leading to

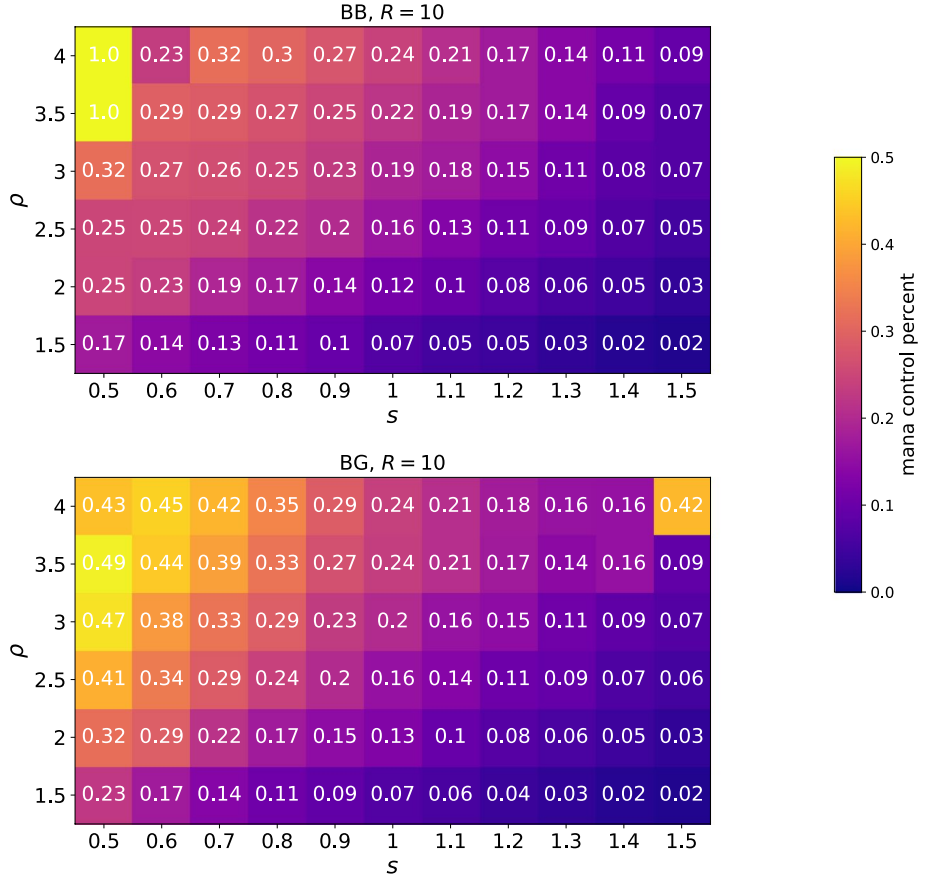

Fig. 5 Cost in Mana (as a percentage of the total Mana in the system) of a blind attack strategy inspired by betweenness strategy results (bottom plot) and greedy strategy results (upper plot) necessary to obtain a successful network split 100% of the times, the results are averaged over 1000 simulations for each parameters combination. s and ρ vary, while $R = 10$ and $N = 100$ are kept constant

the attacker successfully splitting the P2P network and achieving damage that is close to the maximum possible. Such an attack would potentially end up compromising the operations of the IOTA network, as the attacker can control the flow of information between the high-Mana portion of the network and the vast majority of nodes on the other side. For comparison, we also apply the strategies on 1D lattices and fully rewired Watts-Strogatz (WS) networks, finding that the IOTA auto-peering protocol is far less resistant than the latter while slightly more resistant than the 1D lattice.

To conclude, we want to provide a practical comment to interpret our results: the auto-peering network formation model is not currently implemented on the IOTA system; as such there is no immediate risk to the IOTA network.

The results of the present work are to be interpreted as a policy suggestion, presenting a potential attack vector: our results suggest that homophilic P2P network formation models may increase the attack surface of a DLT, increasing the risk of Consensus failure.

At the same time, from a network science perspective, we believe that the IOTA auto-peering model provides an interesting interpretation of random assortative formation models, and provides an alternative bridge between k-regular Poisson graphs and 1D lattices, complementing the popular Watts-Strogatz model.

Appendix

In this appendix, for a better understanding of the strategies employed in the paper, we present additional details regarding the simulation results of the two full information strategies ("Betweenness strategy" and "Greedy strategy"), which in turn were used to develop the "Blind strategy."

In Fig. 6, we show the simulation results of $E[D/x]$ for 1000 graphs, where $N = 100$, $R = 10$, $\rho = 1$, and $N = 100$ when applying the Greedy strategy. This figure explains the target selection mechanism of the Greedy strategy, which selects the target node i^* based on the highest $E[D/x]$ value among the $N - 1$ potential splits.

Betweenness strategy VS Greedy strategy

We then compare the consistency of the Betweenness strategy and Greedy strategy in terms of the nodes that are typically included in the control set. Under the same parameters conditions used to produce Fig. 6, our findings in Fig. 7 reveal that the greedy strategy most often targets node $i_G = 14$, whereas the betweenness strategy's most frequently targeted node is $i_B = 12$.

To compare the attack efficiency in these two strategies, we also reframe the result shown in Fig. 4 of the main text by showing the ratio between the two $E[D/x]$. The result is displayed in Fig. 8, which shows that while for the large majority of parameters the two strategies are similarly effective, in the top-left corner (when ρ is large and s is small) the greedy strategy significantly outperforms the betweenness strategy. This is the same region of the parameters space where both strategies have the worst efficiency, as reported in Fig. 4 of the main text. It is also the range of parameters where the networks produced by the auto-peering model are maximally close to a random-regular topology. In Erdos-Renyi graphs the betweenness information is homogeneous and the attacker gains no advantage from the distribution of link betweenness.

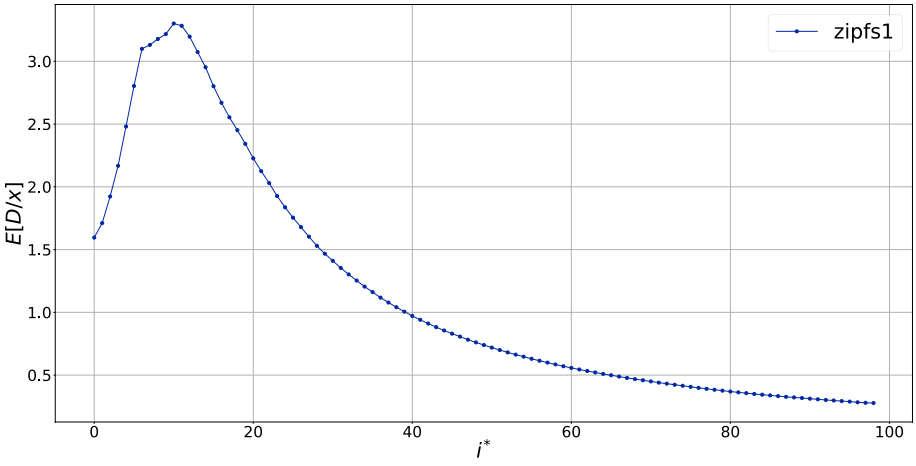

Fig. 6 The average Greedy damage over cost by targeted node $i^* = 1, \ldots N-1$, where N = 100, $\rho = 4$, $R = 10$, and sample size is 1000 graphs

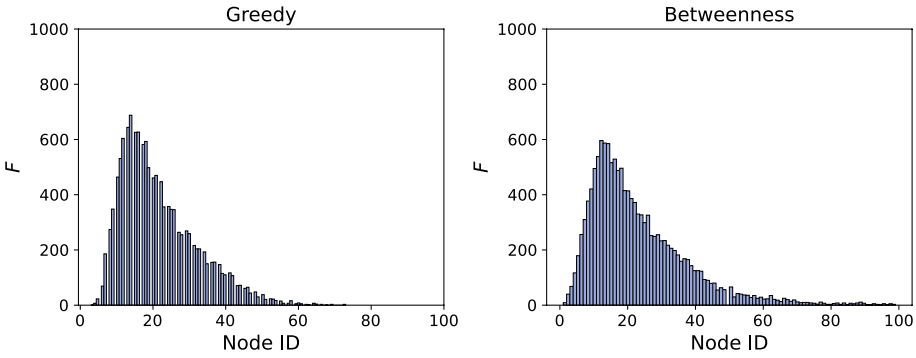

Fig. 7 Histogram of frequencies *F*(*i*) with which node *i* is in the control set according to the Greedy and Betweenness strategies on a sample of 1000 graphs, where $\rho = 4$, $R = 10$, $N = 100$, $k = 4$, $s = 1$

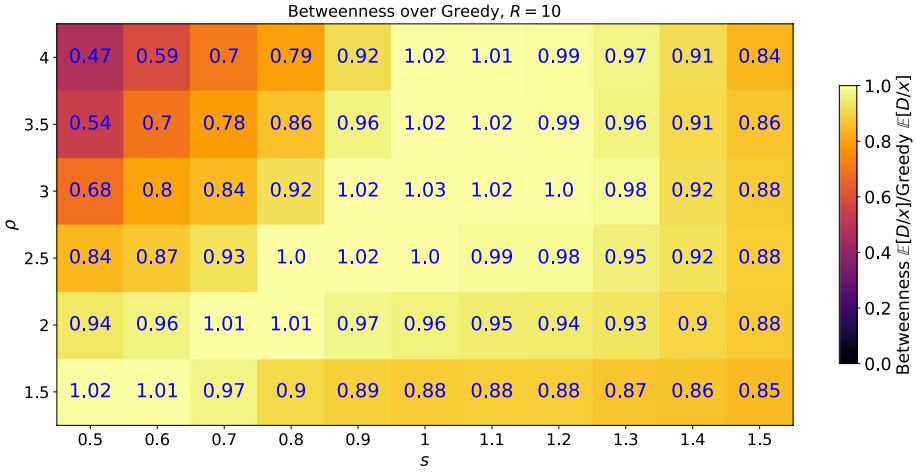

Fig. 8 The heatmap shows the ratio of $\mathbb{E}[D/x]$ for the Betweenness strategy over the $\mathbb{E}(D/x)$ for the Greedy strategy. *s* and ρ varies, while $R = 10$ is kept constant. The reader can notice a minimum placed at $s = 0.5$ and $\rho = 3.5$. That is the region of the parameters' space where the auto-peering model networks most resemble regular Erdos-Renyi random graphs, and the link betweenness becomes homogeneous and highly uninformative

Figure 9 provides an overview of the effect of BB (Blind-Betweenness) and BG (Blind-Greedy) attacks. Here the heatmap indicates the minimum L, that is the count of nodes removed before and after the target nodes i_σ identified by the full-information strategies, such that the blind attack is 100% successful in splitting the simulated graphs. This measure can be intended as a measure for the robustness of the generated networks. We see that in BB attacks the value of L is relatively stable until it explodes in the region of the parameters' space where the network is more akin to an unstructured random regular network, i.e. low s and high ρ. The L values of BG attacks are more gradually increasing when moving towards the same region of the parameters' space. The overall take-home message about the robustness of the different methodologies is that BG attacks are less sensitive to parameter variation.

Finally, we provide a sensitivity analysis for our results with respect to the parameters R and N in Figs. 10 and 11. The analysis demonstrates that both network size N and parameter R present minimal quantitative influence on the results, and do not affect the overall qualitative behaviour. In Fig. 10, varying N yields consistently marginal differences in $E[D/x]$ across the subplots. Similarly, Fig. 11 indicates that altering R produces negligible variations in $E[D/x]$ within each subplot grid. These observations confirm the robustness of the outcomes to changes in both network size N and the parameter R in the additional peer discovery condition.

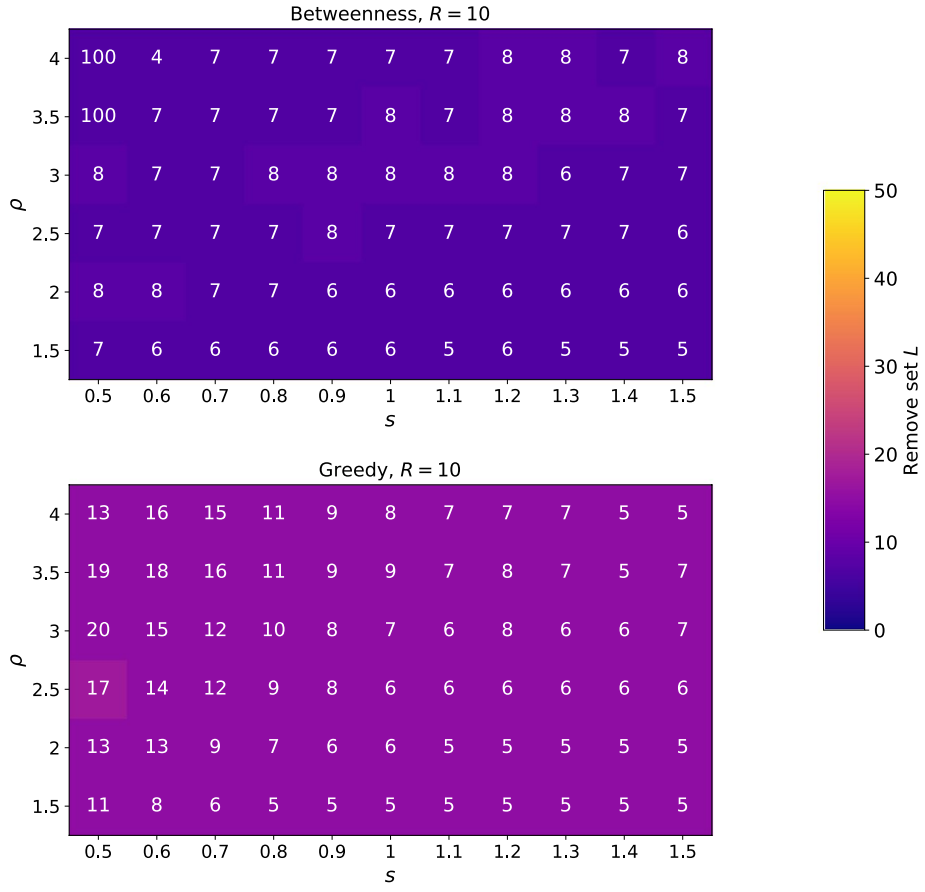

Fig. 9 Each cell's number indicates the minimum value of L necessary to obtain a successful network split 100% of the times in the betweenness and greedy strategies respectively. The results are averaged over 1000 simulations for each parameters combination. s and ρ vary, while $R = 10$ is kept constant

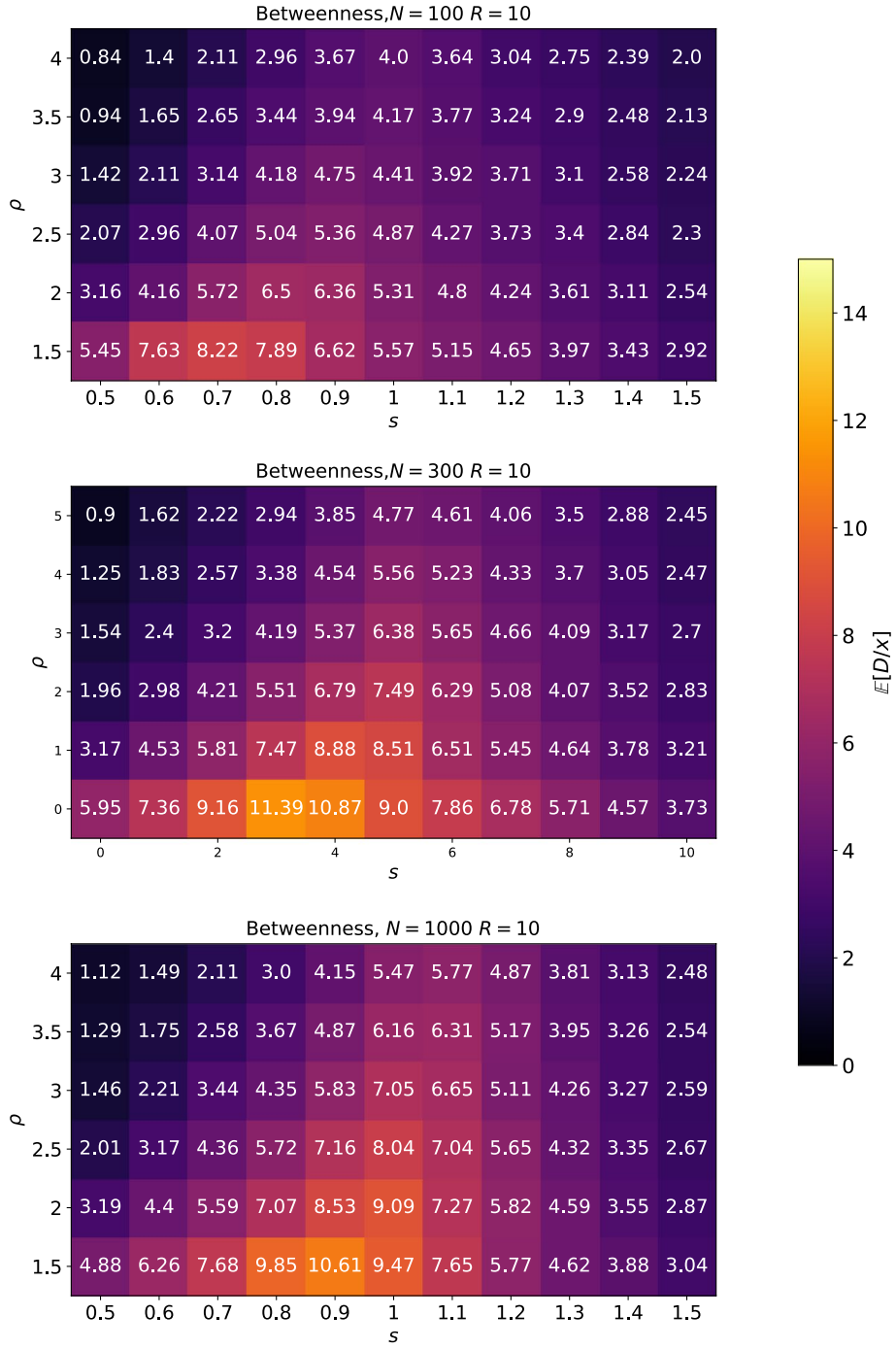

Fig. 10 The heatmap shows the ratio of $\mathbb{E}[D/x]$ for the Betweenness strategy $\mathbb{E}(D/x)$, where *s*, *N* and ρ varies, while $R = 10$ is kept constant. From the three heat map plots, we can see that there is no big gap for the damage over cost in different network sizes

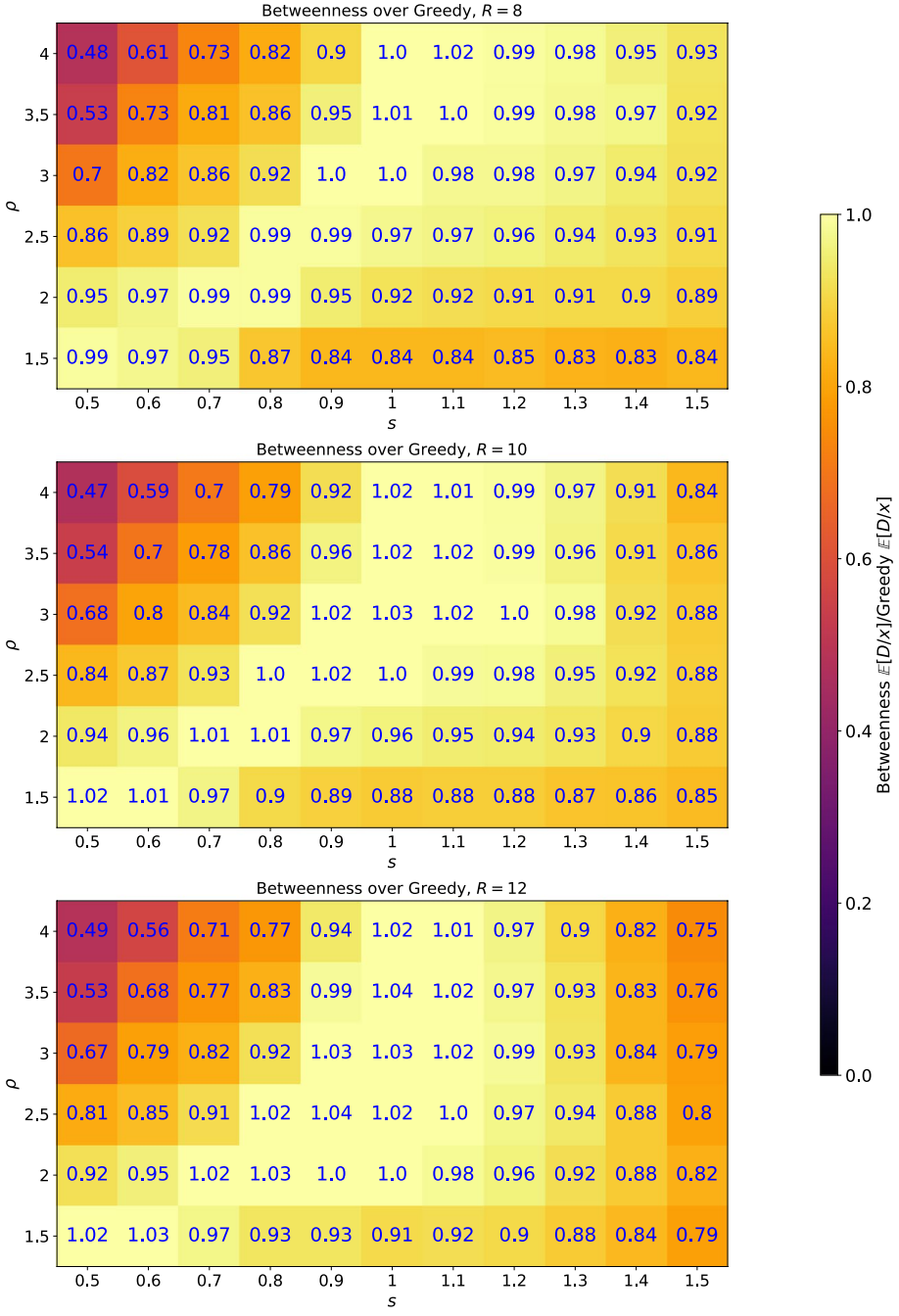

Fig. 11 The heatmap shows the ratio of $\mathbb{E}[D/x]$ for the Betweenness strategy over the $\mathbb{E}(D/x)$ for the Greedy strategy varying s, R and ρ, where $N = 100$

Abbreviations

DLT	Distributed ledger technology
P2P	Peer-to-peer networks
DAG	Directed acyclic graph
BG	Blind-greedy
BB	Blind-betweenness

Acknowledgements
Not applicable.

Funding
YG acknowledges financial support from the China Scholarship Council. AST acknowledges support by FCT through the LASIGE Research Unit, ref. UIDB/00408/2020 (Persistent Web Link) and ref. UIDP/00408/2020 (Persistent Web Link).

Data availability
The scripts to generate the datasets analysed during the current study are available in the "SimulationsAutoPeeringNetworkWithMana" repository, Persistent Web Link.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no conflict of interest.

Author details
[1]Department of Informatics, University of Zürich, Zürich, Switzerland
[2]UZH Blockchain Center, University of Zürich, Zürich, Switzerland
[3]Institute of Finance and Technology, University College London, London, UK
[4]BRAN Lab, Network Science Institute, Northeastern University London, London, United Kingdom
[5]DLT Science Foundation, London, UK
[6]LASIGE, Departamento de Informática, Faculdade de Ciencias, Universidade de Lisboa, Lisbon, Portugal

Received: 31 May 2024 / Accepted: 12 September 2025
Published online: 27 November 2025

References

Lawrence S, Giles CL (2000) Accessibility of information on the web. Intell 11(1):32–39
Oram A (2001) Peer-to-Peer: Harnessing the Power of Disruptive Technologies. O'Reilly Media, Inc.
Cancho RF, Janssen C, Solé RV (2001) Topology of technology graphs: small world patterns in electronic circuits. Phys Rev E 64(4):046119
Heidemann J, Klier M, Probst F (2012) Online social networks: a survey of a global phenomenon. Comput Netw 56(18):3866–3878
Masinde N, Graffi K (2020) Peer-to-peer-based social networks: a comprehensive survey. SN Comput Sci 1(5):299
Tanenbaum AS (2007) Distributed Systems Principles and Paradigms
Androutsellis-Theotokis S, Spinellis D (2004) A survey of peer-to-peer content distribution technologies. ACM Comput Surv (CSUR) 36(4):335–371
Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. Decentralized business review, 21260
El Ioini N, Pahl C (2018) A review of distributed ledger technologies. In: On the Move to Meaningful Internet Systems. OTM 2018 Conferences: Confederated International Conferences: CoopIS, C &TC, and ODBASE 2018, Valletta, Malta, October 22-26, 2018, Proceedings, Part II, pp. 277–288 Springer
Kraner B, Vallarano N, Schwarz-Schilling C, Tessone CJ (2023) Agent-based modelling of ethereum consensus. In: 2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 1–8. IEEE
Wood G (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper 151(2014):1–32
Miller A, Litton J, Pachulski A, Gupta N, Levin D, Spring N, Bhattacharjee B et al (2015) Discovering bitcoin's public topology and influential nodes
Deshpande V, Badis H, George L (2018) Btcmap: Mapping bitcoin peer-to-peer network topology. In: 2018 IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN), pp. 1–6. IEEE
Kim SK, Ma Z, Murali S, Mason J, Miller A, Bailey M (2018) Measuring ethereum network peers. Proc Internet Meas Conf 2018:91–104
Imtiaz MA, Starobinski D, Trachtenberg A, Younis N (2019) Churn in the bitcoin network: Characterization and impact. In: 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 431–439. IEEE
Heilman E, Kendler A, Zohar A, Goldberg S (2015) Eclipse attacks on bitcoin's peer-to-peer network. In: 24th {USENIX} Security Symposium ({USENIX} Security 15), pp. 129–144
Nayak K, Kumar S, Miller A, Shi E (2016) Stubborn mining: Generalizing selfish mining and combining with an eclipse attack. In: 2016 IEEE European Symposium on Security and Privacy (EuroS &P), pp. 305–320. IEEE
Apostolaki M, Zohar A, Vanbever L (2017) Hijacking bitcoin: Routing attacks on cryptocurrencies. In: 2017 IEEE Symposium on Security and Privacy (SP), pp. 375–392. IEEE
Tran M, Choi I, Moon GJ, Vu AV, Kang MS (2020) A stealthier partitioning attack against bitcoin peer-to-peer network. In: 2020 IEEE Symposium on Security and Privacy (SP), pp. 894–909. IEEE
Müller S, Capossele A, Kuśmierz B, Lin V, Moog H, Penzkofer A, Saa O, Sanders W, Welz W (2021) Salt-based autopeering for dlt-networks. In: 2021 3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS), pp. 165–169. IEEE
Moubarak J, Filiol E,Chamoun M (2018) On blockchain security and relevant attacks. In: 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), pp. 1–6. IEEE
Dobson I, Carreras BA, Lynch VE, Newman DE (2007) Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. Chaos Interdiscipl J Nonlinear Sci **17**(2)
Arianos S, Bompard E, Carbone A, Xue F (2009) Power grid vulnerability: a complex network approach. Chaos Interdiscip J Nonlinear Sci **19**(1)
Epstein JM (2012) Generative Social Science: Studies in Agent-based Computational Modeling. Princeton University Press

Hines P, Cotilla-Sanchez E, Blumsack S (2010) Do topological models provide good information about electricity infrastructure vulnerability? Chaos Interdiscipl J Nonlinear Sci **20**(3)
Newman M (2018) Networks. Oxford university press
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world'networks. Nature 393(6684):440–442
Vespignani A (2012) Modelling dynamical processes in complex socio-technical systems. Nat Phys 8(1):32–39
Watts DJ, Muhamad R, Medina DC, Dodds PS (2005) Multiscale, resurgent epidemics in a hierarchical metapopulation model. Proc Natl Acad Sci 102(32):11157–11162
Singh A, Ngan T-W, Druschel P, Wallach DS et al (2006) Eclipse attacks on overlay networks: Threats and defenses
Khan D, Jung LT, Hashmani MA (2021) Systematic literature review of challenges in blockchain scalability. Appl Sci 11(20):9372
Popov S (2018) The tangle. White paper **1**(3)
Popov S, Moog H, Camargo D, Capossele A, Dimitrov V, Gal A, Greve A, Kusmierz B, Mueller S, Penzkofer A et al (2020) The coordicide. Accessed Jan, 1–30
Douceur JR (2002) The sybil attack. In: International Workshop on Peer-to-peer Systems, pp. 251–260 Springer
Cheng A, Friedman E (2005) Sybilproof reputation mechanisms. In: Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-peer Systems, pp. 128–132
Newman ME (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208701
Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network. Am J Sociol 115(2):405–450
Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. Sociol Methods Res 40(2):211–239
Hasheminezhad R, Brandes U (2023) Robustness of preferential-attachment graphs. Appl Netw Sci 8(1):1–19
Moody J (2001) Race, school integration, and friendship segregation in America. Am J Sociol 107(3):679–716
Newman ME (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci 98(2):404–409
Watts DJ (1999) Networks, dynamics, and the small-world phenomenon. Am J Sociol 105(2):493–527
Campajola C, Cristodaro R, De Collibus FM, Yan T, Vallarano N, Tessone CJ (2022) The evolution of centralisation on cryptocurrency platforms. arXiv preprint arXiv:2206.05081
Makarov I, Schoar A (2021) Blockchain analysis of the bitcoin market. Technical report, National Bureau of Economic Research
Chohan UW (2022) Cryptocurrencies and inequality. In: Cryptofinance: A New Currency for a New Economy, pp. 49–62. World Scientific
Sai AR, Buckley J, Gear A (2021) Characterizing wealth inequality in cryptocurrencies. Front Blockchain 4:38
Zipf GK (2016) Human Behavior and the principle of least effort: an Introduction to Human Ecology. Ravenio Books
Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry, 35–41
Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826
Fiedler M (1973) Algebraic connectivity of graphs. Czechoslov Math J 23(2):298–305
Saramäki J, Kivelä M, Onnela J-P, Kaski K, Kertesz J (2007) Generalizations of the clustering coefficient to weighted complex networks. Phys Rev E 75(2):027105
Yang Z, Algesheimer R, Tessone CJ (2016) A comparative analysis of community detection algorithms on artificial networks. Sci Rep 6(1):1–18
Kernighan BW, Lin S (1970) An efficient heuristic procedure for partitioning graphs. Bell System Tech J 49(2):291–307
Kusmierz B, Overko R (2022)How centralized is decentralized? comparison of wealth distribution in coins and tokens. In: 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), pp. 1–6. IEEE

Publisher's Note