

Cyclic motifs in the Sardex monetary network

George Iosifidis ^{1,11*}, Yanick Charette^{2,11}, Edoardo M. Airoldi^{3,4}, Giuseppe Littera⁵, Leandros Tassiulas^{6,7} and Nicholas A. Christakis ^{6,8,9,10}

From decentralized banking systems to digital community currencies, the way humans perceive and use money is changing^{1–3}, thus creating novel opportunities for solving important economic and social problems. Here, we study Sardex, a fast-growing community currency in Sardinia (involving 1,477 businesses arrayed in a network with 48,170 transactions) using network analysis to shed light on its operation. Based on our experience with its day-to-day operations, we propose performance metrics tailored for Sardex but also to similar economic systems, introduce criteria for identifying prominent economic actors and investigate the interplay between network structure and economic robustness. Leveraging new methods for quantifying network ‘cyclic density’ and ‘k-cycle centrality,’ we show that geodesic transaction cycles, where money flows in a circle through the network, are prevalent and that certain nodes have a pivotal role in them. We analyse the transactions within cycles and find that the economic turnover of the involved firms is higher, and that excessive currency and debt accumulations are lower. We also measure a similar, but secondary, effect for nodes and edges that serve as intermediaries to many transactions. These metrics are strong indicators of the success of such mutual credit systems at individual and collective levels.

From new transaction technologies, such as blockchain and mobile payment mechanisms^{1,2}, to novel implementations of alternative currencies and decentralized mutual credit systems³, a plethora of new instruments allow people to trade without using legal tender money. Of special interest are ‘complementary currencies’ (also known as community currencies) that have recently resurfaced and that aspire to stimulate depressed local economies by addressing the money liquidity problem. While this concept can be traced back to the nineteenth century, the penetration of the (mobile) Internet and the emergence of sophisticated digital credit management platforms (<http://www.cyclos.org/>) render modern complementary currency systems particularly attractive, as they can be used to support business-to-business trading (<http://www.wir.ch/>), promote sustainable and local consumption (<http://bristolpound.org/>) and even facilitate cooperation in sharing-economy applications⁴.

Despite their importance for the economy and society, we currently lack a clear understanding of the operation of such systems. For instance, what are their salient economic and structural features? How can we quantify the performance of these closed economies as a whole, or characterize the individual performance of their members? Which of the members play a crucial role in the system’s endurance and wealth-creation capacity? How can we assess the

trust that permits use of the currency beyond bilateral-only trading (and that contributes to functional economies)? What are the network properties of such systems and are there any network effects such as those observed in other economic systems? The answers to these questions cannot rely solely on stylized theoretical models but require an in-depth analysis of a real alternative economy. Hence, we investigate these issues using a novel and complete data set from Sardex (<http://www.sardex.net/>), a community currency launched in Sardinia in 2010 as a response to the financial crisis and currently considered one of the most successful in the world^{5,6}.

Sardex uses an electronic-only complementary currency, is based on a decentralized system implementation without a bank, and aims to serve as a means of exchange. It is a ‘closed’ economy in the sense that the currency is not directly exchangeable with the official currency in Italy (that is, the euro), and cannot be used outside of Sardinia. For instance, a company that is located (or, moves subsequently) outside of Sardinia is not allowed to participate in the network. However, Sardex is pegged to the euro with a ratio of 1:1 to avoid the need for price discovery. This is essentially a ‘mutual credit’ (or, zero-sum) system in the sense that every transaction induces a credit surplus for the seller and an equal debt for the buyer, while the aggregate credit accumulation across all members is zero; that is, there is no credit deficit or surplus in the economy at any given time. This creates network externalities among ostensibly independent transactions that involve different (even distant) actors.

Sardex Spa is the legal entity that monitors the market and ensures the secure operation of the electronic ledger. The system includes businesses representing almost all sectors of the Sardinian economy, spanning the entire island. When a business joins Sardex it obtains a credit line based on the committed capacity of resources it brings into the system. This commitment is expressed in euros and is noted in the contract signed by the business. The goal of the commitment is twofold. It ensures the businesses will not have unwanted Sardex credits, and that there will be enough commodities and services to render Sardex an attractive market place. While the businesses can trade also in euros (outside the network) and are free to decide to what extent they will be involved in Sardex transactions, they must abide by certain trading rules. For instance, the prices they charge in Sardex should be equal to the respective charged prices in euros. Each Sardex member can leave the network, as there are no exit barriers, under the condition that it brings its balance to zero, namely by selling (if it is negative) or buying (if positive) commodities. Finally, it is important to emphasize that there is no interest rate, positive or negative, a design decision aiming to increase the

¹School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland. ²School of Social Work and Criminology, Laval University, Quebec City, Quebec, Canada. ³Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, USA. ⁴Institute for Quantitative Social Sciences, Harvard University, Cambridge, MA, USA. ⁵Sardex Spa, Sardinia, Italy. ⁶Yale Institute for Network Science, Yale University, New Haven, CT, USA. ⁷Department of Electrical Engineering, Yale University, New Haven, CT, USA. ⁸Department of Sociology, Yale University, New Haven, CT, USA. ⁹Department of Biomedical Engineering, Yale University, New Haven, CT, USA. ¹⁰School of Organization and Management, Yale University, New Haven, CT, USA. ¹¹These authors contributed equally: G. Iosifidis and Y. Charette. *e-mail: iosifidg@tcd.ie

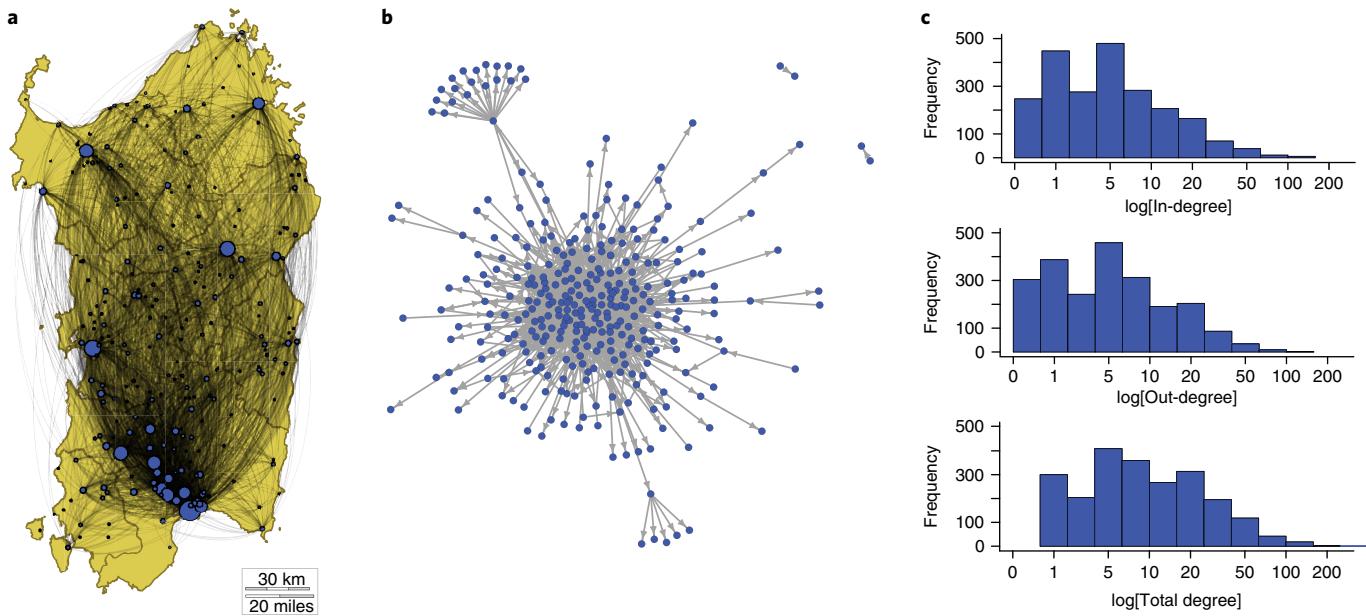


Fig. 1 | A representation of the Sardex transaction network from 1 January 2013 to 31 December 2014. **a**, Each super-node represents the collection of Sardex nodes located within a city or town. The node size is proportional to the number of traders in that city, and edges capture transactions between different locations; $N = 276$ cities; $E = 48,170$ transactions. **b**, The Sardex transaction network within the capital city of Cagliari ($N = 137$ businesses; $E = 498$ partnerships). **c**, In-, out- and total degree distribution of the Sardex network on a log-log scale ($N = 1,477$ businesses; $E = 13,753$ partnerships). Map in **a** adapted from <http://d-maps.com>.

circulation of Sardex credit. Additional information is provided in Supplementary Note 1.

We model the Sardex economy as a network where the nodes are businesses and the edges the currency flow among them. We perform a thorough network analysis that sheds light on the operation of this large-scale digital economic system. We also introduce analytic performance and robustness metrics for the economy, and centrality metrics for identifying prominent nodes. We particularly focus on cyclical geodesic transaction motifs, namely ‘cycles’—where the beginning and end of a series of transactions is the same entity. Such cycles are necessary to sustain the continuing flow of money and also to suppress excessive debt or credit accumulation (which is one of the major causes of failures of these systems). We also perform a secondary analysis, focusing on ‘betweenness’ of nodes and edges, which measures the extent to which they act as transaction intermediaries in Sardex. The insights we obtain are not only relevant to complementary currencies, but also to a range of collaborative platforms with similar operational principles. Our analysis thus sheds light on possible network-related mechanisms pertaining to how humans adopt and use novel monetary instruments.

The spatial characteristics of the Sardex network and the trading relations across different cities can be seen in Fig. 1a, while a network instance for the capital city of Cagliari is shown in Fig. 1b. In our analysis, each individual trader is modelled as a node, where the node degree represents the number of its trading partners, while the weighted directed edges capture the aggregate currency flow from each buyer to each seller in the time interval of interest. The Sardex network has a skewed degree distribution, as depicted in Fig. 1c, with an average degree equal to 18.6 partners (median = 10; s.d. = 26.9) and a maximum of 259. After the fast-growing phase of Sardex in the first three months of this two-year period, several graph properties remained relatively unchanged. For example, the average directed path length (that is, the average length of the sequence of pair-wise directed ties between any two nodes) was approximately 3.5 (median = 3; s.d. = 0.9), and the diameter

of the network stabilized at 10. Similarly, the clustering coefficient was 0.14 and the transitivity was approximately constant and equal to 0.10. Sardex has a single network component and low average path length, similar to small-world⁷ and scale-free networks⁸, but has a high clustering coefficient (12 times higher than Erdos–Renyi networks⁹ but 5 times lower than small-world graphs). A detailed descriptive analysis can be found in Supplementary Note 1.

Our network analysis here focuses on cycles and cyclic transactions; that is, sets of transactions where a group of traders buy and sell from each other in a cyclic fashion. The length k of a cycle is defined as the number of edges that it comprises (see Fig. 2a). A reciprocal (bilateral-only) transaction is a cycle of length $k=2$, while lengthier cycles involve more participants. Obviously, a lack of cycles in dynamic flow (here, credit) networks with no sink or source nodes results in high accumulations of surpluses and debts, disrupting in practice the credit balance (see Fig. 2b). Moreover, the existence of cycles is crucial to distinguish economically healthy mutual credit systems from speculative pyramid or Ponzi schemes that result in isolated and exploited nodes. Besides, there is an intuitive relationship between the system’s overall performance on the one hand, and the existence of transaction cycles, especially the lengthier ones, on the other. That is, lengthier cycles need more time to be completed (counting the time the currency needs to return to, and be redeemed by, the initial seller), meaning, in practice, that they involve a higher trust of the participants towards the currency; this enables the solution of a larger set of ‘double coincidence of wants’ problems and helps to yield a functional economy (Supplementary Fig. 5).

The idea of cyclic network motifs, and procedures for identifying cycles, have been previously described^{10,11}, but here we propose and implement a set of analytical metrics for assessing the presence of cycles at the network level, and for quantifying the contribution of each node to cyclic transactions. We first introduce the cyclic density, which captures the extent to which the network overall contains cycles. This metric is defined as the ratio of the total number of cycles in the trading graph over the expected number of cycles in a

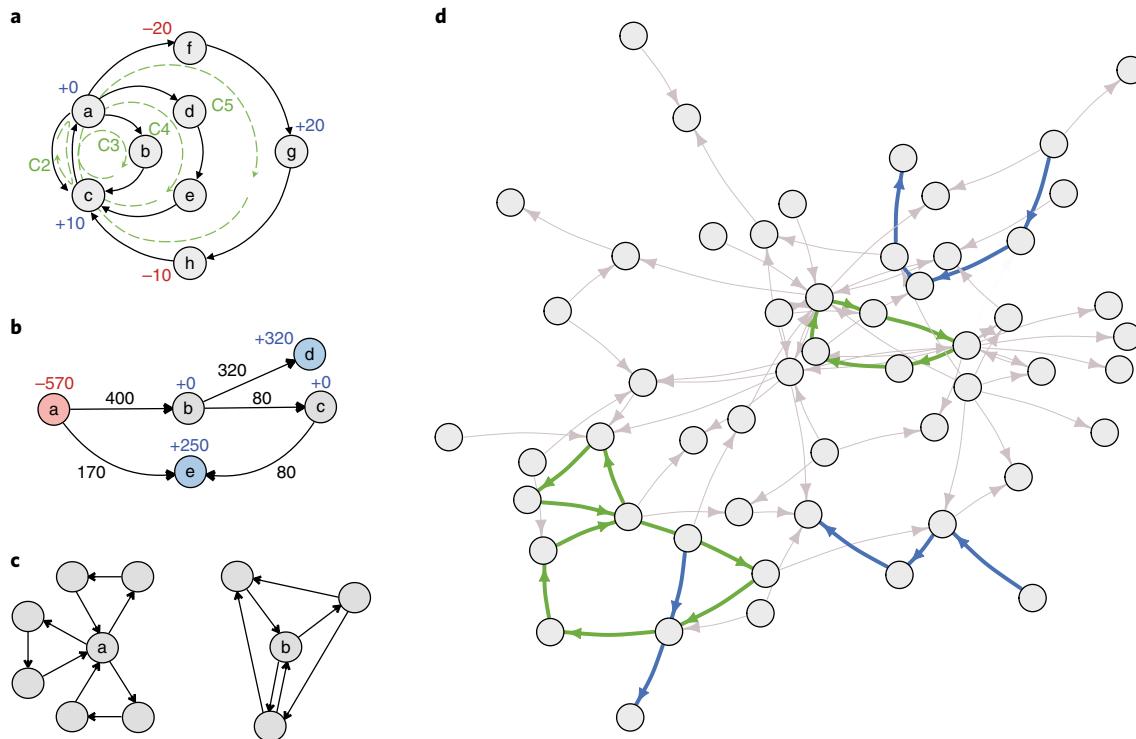


Fig. 2 | The concept and importance of cycle transactions. Edges indicate money flow from buyers to sellers. **a**, Graph with 4 transaction cycles of lengths 2, 3, 4 and 5. Illustrative positive and negative balances shown in red and blue. For example, node d has 4-cycle centrality equal to 1, but 3-cycle centrality equal to 0. **b**, An economic network with no cycles; node d lies at the end of the transaction path (a, b, d), and node e at the end of the path (a, e) and (a, b, c, e). Nodes d and e accumulate credit that they will not be able to spend, while node a builds up a debt that it cannot service. **c**, An example of k -cycle centrality and k -cycle coverage. Nodes a and b both have 3-cycle centrality equal to 1, but node a has 3-cycle coverage of 1 (6/6), while for b this value is 0.5 (3/6). **d**, Cycles of length 3 and 5 (green) and paths of length 2, 3 and 4 (blue) in the Sardex network of the city of Sassari (2013, $N=54$ businesses; $E=91$ partnerships).

properly defined null model. The higher the value of the cyclic density, the higher the potential performance of the trading network in all likelihood. Building on earlier work¹⁰, we also introduce the notion of k -cycle node centrality, which quantifies the portion of the network's trading cycles of length $k \geq 2$ in which a node participates. The larger the value of the k -cycle centrality, the more important that node probably is for the performance of the system. For example, removing a node with a 5-cycle centrality value of 0.5 will break down 50% of the cycles of length 5 in the whole network, and this will increase the likelihood of having isolated nodes with high debts or surpluses. Note that it is known that severed trading relationships are often difficult and time consuming to replace¹², even more so in closed economic systems such as Sardex. Finally, we introduce the k -cycle node coverage metric, which quantifies, for each node, the overlap of its cycles, and hence how many different nodes appear in its cycles. Two nodes may have the same k -cycle centrality but different k -cycle coverage, as shown in Fig. 2c.

To assess the performance of this mutual credit system, we propose and quantify two metrics: the aggregate volume of transactions (or turnover) of the system and the credit healthiness for each of its members. The first metric has been used by practitioners to describe the system-level performance of various community currencies. The second, node-level metric captures excessive and prolonged credit (positive) or debt (negative) accumulations by traders. In particular, nodes that have had a high credit or debt for many days during the time interval of interest are those with undesirable credit healthiness (examples are provided in Supplementary Note 2). Such accumulations indicate inactive members who pose a threat to the robustness of the system, either because they immobilize credit surpluses or because they are incapable of reducing their debts.

Leveraging our approaches to identifying and quantifying trading cycles, we then explore two fundamental questions. At the network level: does Sardex have many cycles, and are there nodes with a pivotal role in them? At the cycle level: are cycles associated with the above performance metrics, and what is the economic activity within the cycles? Moreover, we investigate these questions for cycles of different length: is it preferable in such local economies to have small trading relations, involving two or three nodes, or lengthier cycles, involving transactions among four or five nodes in cycles involving 'unseen' others?

First, we examine cyclic density and credit healthiness at the overall network level, across time. Based on our experience from the day-to-day management of the system, we consider as unhealthy the state of having debt or credit higher than 90% of a node's capacity (its credit line). Figure 3a depicts the network's cyclic density for different sizes of the Sardex graph, which has a higher number of cycles compared to different null models. For the latter, we have used an analytical formula that quantifies the expected number of cycles in a graph built by randomizing the edges of the observed Sardex network while preserving its in/out-degree distribution (henceforth referred to as reshuffled-degree graphs)^{13–15}. We have also used an Erdos–Renyi⁹ graph and a small-world⁷ graph as null models, and further verified the results by a Z-test using a population of 100 reshuffled-degree graphs, as shown in Fig. 3b. Regarding the performance metrics, the aggregate volume of new transactions conducted in every month increases with time, as shown in Fig. 3c. Moreover, Fig. 3d depicts the portion of unhealthy nodes at the end of each month; interestingly, we observe a 50% reduction of this quantity during the two-year operation of the market. In other words, as the network grows, Sardex traders not only manage to

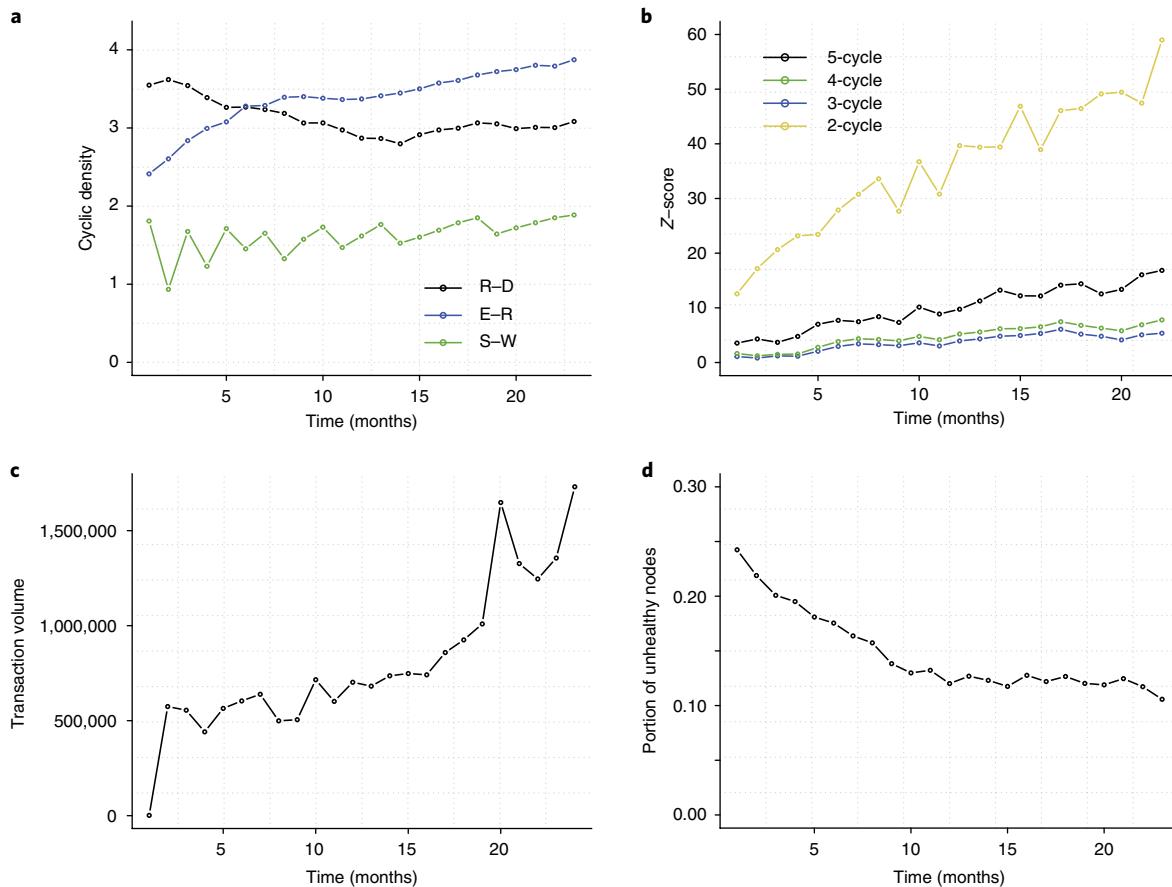


Fig. 3 | Overall performance of the Sardex exchange. **a**, Sardex cyclic density calculated with respect to the expected number of cycles in a population of 200 reshuffled-degree graphs (using the formula in ref. ¹³; denoted R-D), a population of 200 Erdos-Renyi graphs with the same number of nodes and edges (denoted E-R) and a population of 200 small-world graphs (denoted S-W). **b**, Z-test scores for Sardex comparing the number of observed cycles with the cycles in a population of 100 random graphs preserving the Sardex in-/out-degree distribution (R-D graphs). **c**, Volume of new transactions per month. **d**, Credit ‘unhealthiness’ in Sardex; that is, portion of nodes with average absolute balance greater than 90% of their capacity, at the end of each month. Figure 3a has been created with Supplementary Codes 4, 5 and 6.

increase the volume of their transactions, but they do so in a fashion that avoids prolonged, excessive debt or credit accumulations.

Regarding the role of nodes in Sardex cycles, we see that the distribution of the k -cycle centrality has a heavy tail, with a few nodes participating in a large percentage of cycles, while many nodes do not participate in any cycle. These results depart substantially from the cycle centrality distribution in the null models and reveal that a small subset of nodes plays a crucial role in the economy as, for example, their removal would increase idle currency accumulations (prolonged surpluses or debts). We also correlated the cycle centrality with other node centrality metrics—that is, the degree and betweenness centrality¹⁶—to assess the information load of this new metric. We find that the correlation of cycle centrality with betweenness centrality ($r=0.70$, $P<0.001$) is comparable to the correlation of betweenness with degree centrality ($r=0.77$, $P<0.001$). Additionally, we find that the correlation of the cycle centrality with degree centrality can be high ($r=0.92$, $P<0.001$) for certain ranges of values of these centralities. Interestingly, however, a closer look at the scatter plots (Supplementary Fig. 9) reveals that nodes with small-to-medium values of cycle centrality have a large range of betweenness and degree centrality. This suggests that cycle centrality does carry additional information. Finally, we calculated the k -cycle coverage distribution in Sardex and found substantial variation and non-redundancy (see Supplementary Fig. 10).

Next, we turn our focus to the transactions within cycles to assess the association of a firm’s being in cycles with its economic performance. First, we explore the role of cycles at the edge level using a dyadic generalized linear regression model¹⁷. The construction of the transaction graph and the enumeration of cycles is performed for each year separately, and the analysis is conducted jointly on both samples with proper adjustments. Figure 4a shows the model estimates. We observe that an increase of 1 s.d. of the number of cycles of length $k=2$ is associated with an increase of 5% in the edge transaction volume ($b=0.05$; 95% confidence interval (95% CI), 0.03, 0.07; $P<0.001$), and that this increases up to 12% for cycles of length $k=5$ ($b=0.10$; 95% CI, 0.07, 0.14; $P<0.001$). Interestingly, we find also that the betweenness centrality of the edges has a comparably positive impact on their transaction volume, meaning that edges participating in many chains of transactions (of any length, open or closed) have higher performance ($b=0.09$; 95% CI, 0.06, 0.12; $P<0.001$). Note that the impact of cycles of length $k=5$ was comparable to that of betweenness. As shown in Fig. 4a, a similar association was not observed for paths (that is, unclosed cycles) of the same respective length. The latter are defined as sequences of connected nodes where no node appears more than one time, and we have used only the paths that are non-nested, and, thus, independent of cycles and longer paths.

Finally, we explore the credit healthiness of the nodes that are involved in many cycles. We used an ordinary least-squares (OLS)

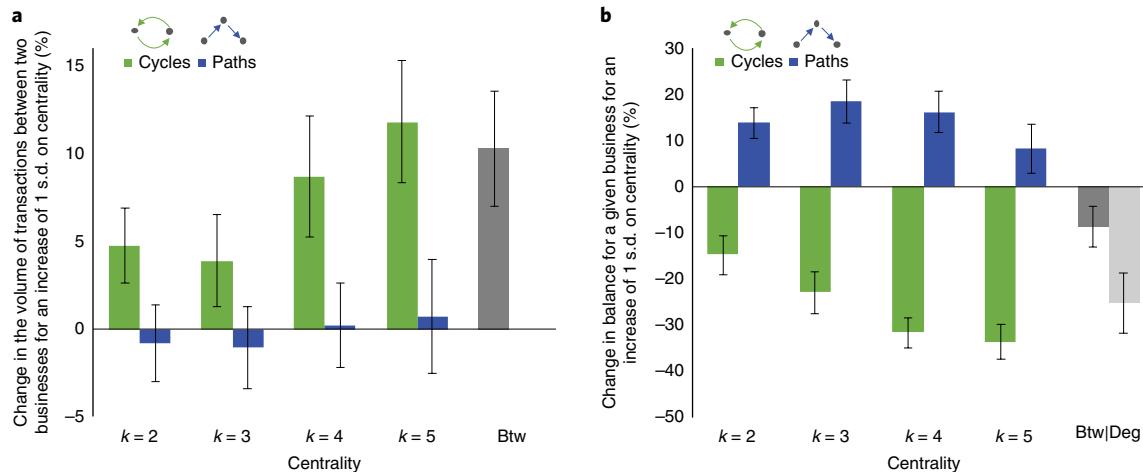


Fig. 4 | Statistical models' coefficients with 95% CIs ($N = 1,477$ businesses; $E = 13,753$ partnerships). **a**, Relationship between cycles and paths of length k , and edge betweenness (Btw) centrality and the edge transaction weight (dyadic generalized linear models with a random effect at the node level). **b**, Relationship between cycles and paths of length k , degree (Deg) and node betweenness (Btw) centrality and the node credit healthiness (OLS model). The figure has been created with Supplementary Code 7.

model with the credit healthiness of each business as the dependent variable (see Methods). Figure 4b shows the estimates of these models. We observe that businesses with many trading partners (degree) have a lower average absolute balance; that is, better credit healthiness. Businesses that are part of longer cycles also have lower average balance. While an increase of 1 s.d. in the number of cycles of length $k = 2$ is associated with an improvement of 14% in the trader's credit healthiness (by lowering the average absolute balance; $b = -0.17$; 95% CI, $-0.21, -0.12$; $P < 0.001$), this impact increases to 30% for cycles of size $k = 4$ ($b = -0.38$; 95% CI, $-0.44, -0.33$; $P < 0.001$) and size $k = 5$ ($b = -0.41$; 95% CI, $-0.47, -0.35$; $P < 0.001$). Similar to the edge-level analysis, we find that the betweenness centrality of the nodes improves substantially their credit healthiness ($b = -0.09$; 95% CI, $-0.14, -0.03$; $P = 0.003$). However, the betweenness effect is smaller than the impact of cycles; that is, nodes that are involved in many (and lengthier) cycles are healthier even compared to nodes with high betweenness centrality. As hypothesized, and shown in Fig. 2b, the number of non-closed paths passing through a node is associated with an increase in its average absolute balance; that is, a decrease in its healthiness.

The findings at the edge and node level resisted a long battery of robustness analyses accounting for dependencies and alternative modelling approaches (see Supplementary Note 3). In addition, we indirectly compared the impact of cycles of different lengths by using a stepwise statistical analysis (gradually adding the different cycles), and observed an increase in model fit that supports the importance of lengthier cycles. Identical conclusions are reached if we compare the coefficients and the fit of the different independent models. We also compared the cycles and the paths by considering the respective aggregate metrics—that is, number of cycles of all lengths and number of paths of all lengths that an edge belongs to—and the results were aligned with the above findings. Finally, we explored whether the observations about performance within cycles relate to the geographic proximity of the transacting nodes. Interestingly, we found that traders within cycles are not more frequently co-located than traders that participate in non-nested paths; and, more important, we found that edges across different cities (both for cycles and paths) have slightly higher weights (see Supplementary Note 2). Therefore, the association of cycle length and economic performance that we observe is not rooted in geographic co-location of the traders.

Alternative financial services are of increasing importance. They leverage recent technological advances and aspire to address

fundamental economic needs, especially in rural and under-developed areas where affordable banking services are often lacking. At the same time, they constitute promising laboratories for the study of salient aspects of social and economic life, such as how humans perceive and use money, or trade and collaborate with each other^{2,18}. And studying such monetary systems allows us to revisit fundamental questions regarding the emergence and performance of various monetary instruments^{19,20} using quantitative analysis of detailed digital transaction traces²¹. Furthermore, credit systems such as Sardex are increasingly relevant for emerging sharing-economy services⁴ that rely on decentralized trading and cooperation platforms. In these systems, more often than not, the exchange of goods and services takes place using some kind of a coupon-based system. In all of these cases there is a need for (and already use of) an accounting mechanism similar to Sardex, so as to facilitate the participants' interactions. Hence, the ideas and metrics proposed here are applicable elsewhere too.

Information from monetary instruments that trace out paths through an economy is not commonly available²². But the availability of data such as ours, with a panopticon view of a complete and defined economic system, which is increasingly possible²¹, allows us to examine the flow of money in a new way. We find that Sardex shares common features, such as the small-world property, with social or technological networks; but it also has distinct properties such as the prevalence of cycles. We show that cyclic motifs are increasingly over-represented in the Sardex economic network and that a subset of nodes plays a central role in the existence of cycles. This result verifies the common intuition about the importance of circular transactions in an economy, which is even more crucial (and structurally necessary) in these closed mutual-credit economies. We then quantified important economic metrics, namely edge weight (the transaction volume between traders) and node credit healthiness (a measure of net balances), for the traders involved in many cycles. Our findings suggest that cycles are positively associated with system performance at both global and individual levels and that these associations are larger for the longer cycles. Besides, such associations are not seen (or are much smaller) for linear paths of equal length to such cycles.

These findings resisted a battery of robustness checks. Nevertheless, we do not make any formal causality claim about the effects of cycles on economic performance because, from a statistical point of view, this cannot be strictly proved. Namely, a proper

randomized controlled trial is unfeasible in this setting and hence we would need to rely on either observational studies or natural experiments for leveraging instruments. With either of these methodological approaches, valid causal inference could be pursued at the level of granularity of the economy (having whole economies as units of analyses) or at the level of businesses (within a single economy). Unfortunately, there is only one economy of this type at this stage of development, and analysis at the business level raises treatment interference issues for which there are no straightforward solutions, though we use spatial regression models and other analytic tools here (see Supplementary Note 3).

Still, our network modelling and analysis approach reveals that complex types of embeddedness in economic systems—and not just dyadic (or even triadic) interactions between two traders engaged in direct exchange—are associated both with individual actors' economic performance and with the economic success of the system as a whole. Previous experimental work has shown that the structure of networks can affect the value individuals can extract from networks, whether their interactions are cooperative or monetary^{19,23–25}. In the economy, such studies have focused on the topological properties of the world trade web²⁶, or of interbank payment networks²⁷. Our data involve different types of economic actors, a much smaller geographic scale, and include detailed temporal and geodesic information. Of course, it is important to emphasize that fully disentangling the Sardex flows is practically impossible because the currency units do not have identification numbers²².

The significance of diverse motifs in complex systems has been evaluated in various situations in recent decades, and several techniques for quantifying them have been devised¹⁵. Cyclic motifs are of crucial importance in a variety of settings that range from bartering- and sharing-economy applications⁴ to gift exchange networks or even to kidney exchange systems²⁸. Nevertheless, cycles are uncommonly investigated, especially in directed graphs such as ours. One earlier study provided a closed-form expression that relates the number of expected cycles in a directed graph with its in/out-degree distribution, and it showed that cycles in food web, power grid, metabolic and other networks are under-represented compared with respective random graphs^{10,13}. Our analysis builds on these earlier works and provides a systematic approach not only for assessing the prevalence of cycles in observed networks, but also for quantifying which nodes contribute to this phenomenon, using the new k -cycle centrality and k -cycle coverage metrics.

The existence of cycles has been, mainly, associated with negative effects such as instability in dynamic systems. It has been shown that short or long feedback loops can destabilize systems by reinforcing oscillations and amplifying undesirable perturbations. For example, a study of biological and technological networks found that directed cycles are under-represented, and it hypothesized that this property emerged through an evolutionary process rewarding stable structures; extensive simulations supported this argument¹¹. Similarly, another study used an analytical model to prove that cyclic network motifs can amplify risk contagion effects in financial systems such as the interbank loan networks²⁹. In this case, a small asset devaluation in a bank might trigger a cascade of equity reassessments to its lenders, and, through a cyclic path of successive interbank links, further deteriorate the equity of this bank. Contrary to these results, we show here (using actual data) that the system's performance is improved within the cyclic transactions. This is not surprising if one realizes that Sardex cycles capture cooperative relationships, and the credit obligations are not bilateral but rather towards the community. Namely, each debtor can pay its debt by selling products and services to any other Sardex member, and, similarly, each creditor can spend its credit buying from any other member. Hence, perturbations and node failures can be amortized and accommodated by the system.

Furthermore, Sardex cycles perhaps reflect trust in the currency and in the ability of Sardex members to sustain this complementary

market, and hence the cycles reinforce this positive effect. It is possible to see the positive effect of cycles as due to the fact that money in Sardex is essentially a cooperation-fostering medium, while, in other financial networks, it is treated as a commodity. Finally, it is interesting to note that our findings suggest also that betweenness centrality plays a very important role in both of the considered performance metrics and mainly in increasing the edge weight. Intuitively, this might be related to the fact that, in such closed systems, the economic activity can be improved if certain links act as strong conduits transferring credit among distant groups of nodes, and this is, by definition, the role of these high-betweenness edges.

The existence and relevance of cycles also suggests possible strategies for policy makers to evaluate, for example, strategically fostering cycles by brokering introductions to trading partners. While outsiders to a system intervening in it may not be able, for example, to change the betweenness centrality of actors, they might be able to change cycle centrality to similar effect by brokering a few introductions to foster the creation of cycles. Moreover, with respect to the credit healthiness of firms, for nodes with the same turnover (that is, weighted degree), increasing the number of trading partners (degree) improves healthiness since it splits the load more evenly (more, smaller transactions are creating the same wealth). On the other hand, if we want to actually increase the turnover in the system, adding more transactions over the existing trading relationships (that is, the existing edges) deteriorates the balance unless these transactions are part of cycles.

Trust of the users towards a complementary currency such as Sardex is probably the most important element for its function, as it enables the currency's adoption and utilization beyond bilateral-only trading. The notion of trust here is more intricate than in classical macroeconomic models³⁰ as it includes the trust of the traders towards the Sardex network itself (which is not backed by any official state), the trust that there will be enough resources of interest to buy in the future in this closed economy and the trust that the other traders will repay their debts. Cycles in networks may be relevant to economic success in part because the existence of many—and, in particular, lengthy—cycles indicates the trust of the traders in this closed economy with respect to all of these factors. Hence, money itself may not be enough to induce trust or robustness of exchange among strangers in groups³¹ and cycles may also be needed. Indeed, from a structural, geodesic point-of-view, cycles play a crucial role in the flow of money in a mutual credit system. Moreover, and more generally, cycles may be relevant to other phenomena within graphs, such as the flow of germs or information, beneficially retarding the flow of the former and harmfully suppressing the flow of the latter. Cycles are also surely relevant to the controllability of a graph³². Understanding the practical relevance of cyclic motifs is an opportunity for further work. While individuals might control their bilateral economic ties, they have less control over their long-range interactions. Yet, such unseen network features may matter, both for individuals themselves and for their group as a whole.

Methods

Data sets and network model. Our data set includes detailed information for all Sardex transactions conducted from 1 January, 2013 until 31 December, 2014 (see Supplementary Note 1). Additionally, the data set includes information about each firm's category, capital commitment, location and date of joining the system. This time interval constitutes the fastest expansion period of Sardex, which, by the end of 2014, yielded an aggregate turnover of 38.81 million euros. We do not have data about the businesses' transactions in euros, nor about their overall economic activities (in euros). We have removed the transactions among businesses and their own employees since the latter have different qualitative and quantitative features. Namely, employees conduct much smaller transactions than businesses and, most importantly, they cannot have debt since they are not assigned a credit line. The entire population of the businesses in Sardex has been used, and therefore no statistical methods were employed to predetermine sample. The authors have the necessary approval to analyse and publish these data.

We have defined the Sardex network as follows: each node represents a Sardex business, and each directed weighted edge captures the aggregate flow of currency from the buyer business (source node) to the seller business (destination node) during the time period of interest. This currency flow can be the result of one or more transactions. We construct and analyse the Sardex network on a yearly basis. That is, we split the network into two distinct networks; one for the first year, and one for the second year. The reason for following this approach is twofold. First, we do not assume that the effect of transaction edges persists forever, and hence we consider only cycles within a certain time interval (12 months). Second, we have found that the cycles of length $k=5$ are created, on average, in approximately 230 d, which renders the annual separation more representative (Supplementary Fig. 5). Following this methodology, the Sardex network had 877 nodes and 5,962 edges in 2013, 1,353 nodes and 9,916 edges in 2014 and 1,477 nodes and 13,753 edges for the overall period of two years (2013–2014). Mathematically, Sardex is modelled as a weighted directed graph $G=(N,E)$, where N is the set of Sardex businesses and E the set of edges representing their transactions during the time period of interest. Every edge $(i,j)\in E$ with weight e_{ij} captures the total amount of currency that i has paid to j .

Formulae. Following previous modelling approaches³³, we define an elementary path p_{uv} of length k as an ordered set of nodes $p_{uv}=(u=u_1,u_2,\dots,u_{k+1}=v)$, such that $(u_i,u_{i+1})\in E$ for $i\in[1,k]$. Next, we define an elementary cycle c_{uw} as an elementary path in which the first and the last nodes are identical. A cycle of length $|c_{uw}|=k$ contains k edges and k different nodes. Two elementary cycles are considered distinct if the one is not a cyclic permutation of the other, and we have considered only such cycles in our study. We define the set of cycles of length k that include nodes as $P_k(n)=\{c_{uw}: n\in c_{uw}, u\in N, |c_{uw}|=k\}$, and we also define the set of all cycles of length k in the graph $P_k(G)=\{c_{uw}: u\in N, |c_{uw}|=k\}$. We introduce the k -cycle centrality of node n as the portion of k -cycles that node belongs to; that is, $C_y(n)=|P_k(n)|/|P_k(G)|$. By definition, k -cycle centralities, for all nodes and all values of k , lie in the interval [0,1]. If a graph does not have any cycles of length k , we set the k -cycle centrality equal to zero for all nodes. Finally, we define the k -cycle coverage $Cv_k(n)=M_k(n)/|P_k(n)|(k-1)$, where $M_k(n)$ is the number of different nodes in the k -cycles that traverse node n .

The k -cyclic density and (overall) cyclic density for a network $G=(N,E)$ are defined as follows:

$$C_k(G)=\log\left[\frac{P_k(G)}{\max\{1, E(|P_k(G_N)|)\}}\right], k=2, \dots, N,$$

and $C(G)=\sum_{k=2}^T \frac{1}{T} \log\left[\frac{P_k(G)}{\max\{1, E(|P_k(G_N)|)\}}\right]$

where T is the upper bound of the length of cycles we enumerate, and its value depends on the context. In this study, we use $T=5$; that is, we count up to 5 cycles, as the average creation delay of longer cycles expands beyond the duration of the data set. The cyclic density metric takes a negative value when the observed cycles are less than in the null model, a zero value when it is exactly the same and positive values for graphs that exhibit more cycles. Note that we use a logarithmic scale as the number of cycles grows exponentially with the number of nodes. Also, the ‘max’ operator in the denominator ensures the metric is properly defined even for small-sized graphs. The enumeration of cycles can be implemented using previous methods³³.

Models. For the edge-level analysis in Fig. 4a, we have used dyadic generalized linear models with a random effect at the node level¹⁷ predicting the volume of transactions between two businesses (that is, over a certain edge). Due to skewness of distribution, volume was log transformed and treated as a normal distribution. Confidence intervals were measured using two-tailed tests. We controlled for the type of businesses involved (using the business category) as well as for their lifespan (that is, the duration since they joined the network), and we compared the results with the number of paths of the same length crossing that edge. Our analysis is on an annual basis. That is, the construction of the transaction graph and the enumeration of cycles is performed for each year separately, and the statistical analysis is conducted jointly on both samples with proper adjustments. For the node-level analysis in Fig. 4b, we have used an OLS model with the credit healthiness of each business as the dependent variable. Due to skewness of distribution, credit healthiness was log transformed and treated as a normal distribution. Confidence intervals were measured using two-tailed tests. We controlled for the business type, its weighted degree (turnover) and its lifespan; and we again compared the results with the impact of paths of the respective length.

Several alternative modelling approaches have been employed, including a random effects model for the node-level analysis, that account for relational and temporal node dependencies with appropriate corrections for the standard errors (Supplementary Note 3). We have also implemented various measures for assessing and reducing multicollinearity issues. We repeated our study for each year separately, so as to consider the time dimension, and we also employed a lagged model that studies the impact of cycles created in year one on the performance metrics in year two, and reached the same conclusions about the transactions and credit conditions within cycles. We also tested the robustness of the above findings

when we employ only the time-sequential cycles; that is, those cycles with edges that are created in strict time sequence. We controlled for the k -cycle coverage of nodes and edges and did not observe any qualitative difference in our findings. Finally, we also employed a set of non-parametric, non-linear models, such as random forest and support vector machine models, which, however, did not produce interpretable results. Details for all models and methods are available in Supplementary Note 3.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Code for the main models and figures that are presented in the paper and the Supplementary Information is provided at the Supplementary Software report, and also available upon request from the corresponding author.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to approval by Sardex Spa and based on the confidentiality agreement of Sardex Spa with its clients.

Received: 12 September 2017; Accepted: 11 September 2018;

Published online: 22 October 2018

References

- Satoshi, N. Bitcoin: a peer-to-peer electronic cash system. *bitcoin* <https://bitcoin.org/en/bitcoin-paper> (2008).
- Suri, T., Jack, W. & Stoker, T. M. Documenting the birth of a financial economy. *Proc. Natl Acad. Sci. USA* **109**, 10257–10262 (2012).
- Seyfang, G. & Longhurst, N. Growing green money? Mapping community currencies for sustainable development. *Ecol. Econ.* **86**, 65–77 (2013).
- Sundararajan, A. *The Sharing Economy* (MIT Press, Cambridge, 2016).
- The Sardex factor. *Financial Times Magazine* (18 September 2015); <http://www.ft.com/intl/cms/s/2/cf875d9a-5be6-11e5-a28b-50226830d644.html>
- Littera, G., Sartori, L., Dini, P. & Antoniadis, P. From an idea to a scalable working model: merging economic benefits with social values in Sardex. *Int. J. Community Currency Res.* **21**, 6–21 (2017).
- Watts, J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
- Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Erdos, P. & Renyi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).
- Butts, C. T. *Cycle Census Statistics for Exponential Random Graph Models* (Institute for Mathematical Behavioral Sciences, UC Irvine, 2006); <https://escholarship.org/uc/item/4719h8d2>
- Maayan, A. et al. Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. *Proc. Natl Acad. Sci. USA* **105**, 19235–19240 (2008).
- Carvalho, V. M., Nirei, M., Saito, Y. U. & Tahbaz-Salehi, A. *Supply Chain Disruptions: Evidence from the Great East Japan Earthquake* Working Paper No. 2017-01 (Becker Friedman Institute for Research in Economics, 2017).
- Bianconi, G., Gulbahce, N. & Motter, A. E. Local structure of directed networks. *Phys. Rev. Lett.* **100**, 11870 (2008).
- Opsahl, T., Colizza, V., Panzarasa, P. & Ramasco, J. J. Prominence and control: the weighted rich-club effect. *Phys. Rev. Lett.* **101**, 168702 (2008).
- Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Networks* **1**, 215–239 (1978).
- Snijders, T. A. B. & Kenny, D. A. The social relations model for family data: a multilevel approach. *Pers. Relat.* **6**, 471–486 (1999).
- Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
- Guyer, J. J. Soft currencies, cash economies, new monies: past and present. *Proc. Natl Acad. Sci. USA* **109**, 2214–2221 (2012).
- Simmel, G. *The Philosophy of Money* (Routledge, Cambridge, 2011).
- Lazer, D. et al. Life in the network: the coming age of computational social science. *Science* **323**, 721–723 (2009).
- Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
- Nishi, A., Shirado, H., Rand, D. & Christakis, N. A. Inequality and visibility of wealth in experimental social networks. *Nature* **526**, 426–429 (2015).
- Nishi, A., Shirado, H. & Christakis, N. A. Intermediate levels of social fluidity amplify economic growth and mitigate economic inequality in experimental social networks. *Sociol. Sci.* **2**, 544–557 (2015).
- Rand, D. G., Nowak, M., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proc. Natl Acad. Sci. USA* **111**, 17093–17098 (2014).

26. Fagiolo, G., Reyes, J. & Schiavo, S. On the topological properties of the world trade web: a weighted network analysis. *Phys. A* **387**, 3868–3873 (2008).
27. Soramaki, K., Bech, M. L., Arnold, J., Glass, R. J. & Beyeler, W. E. The topology of interbank payment flows. *Phys. A* **379**, 317–333 (2007).
28. Abraham, D. J., Blum, A. & Sandholm, T. Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. In *Proc. 8th ACM Electronic Commerce* 295–304 (ACM, 2007).
29. Bardoscia, M., Battiston, S., Caccioli, F. & Caldarelli, G. Pathways towards instability in financial networks. *Nat. Commun.* **8**, 14416 (2017).
30. Kiyotaki, N. & Moore, J. Evil is the root of all money. *Am. Econ. Rev.* **92**, 62–66 (2002).
31. Camera, G., Casari, M. & Bigoni, M. Money and trust among strangers. *Proc. Natl Acad. Sci. USA* **110**, 14889–14893 (2013).
32. Lu, Y. Y., Slotine, J. J. & Barabasi, A. L. Control centrality and hierarchical structure in complex networks. *PLoS ONE* **7**, e44459 (2012).
33. Johnson, D. B. Finding all the elementary circuits of a directed graph. *SIAM J. Comput.* **4**, 77–84 (1975).

Acknowledgements

We thank D. G. Alvarez, F. Fu, J. Horton and A. Oswald for helpful comments. Support for this research was provided by a grant from the Robert Wood Johnson Foundation and the Star Family Foundation. Also, G.I. acknowledges that this publication has emanated from research supported in part by a research grant from Science Foundation Ireland under grant 16/IA/4610; and E.M.A. acknowledges the support by the National

Science Foundation under grant IIS-1409177 and by the Office of Naval Research under grants YIP N00014-14-1-0485 and N00014-17-1-2131 to E.M.A. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

G.I., Y.C., L.T. and N.A.C. designed the project. G.I. and G.L. prepared the data. G.I. and Y.C. performed the statistical analyses. G.I., Y.C., E.M.A., L.T. and N.A.C. analysed the findings. G.I., Y.C., E.M.A., G.L., L.T. and N.A.C. wrote the manuscript.

Competing interests

G.L. is one of the founders and currently an employee of Sardex Spa.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41562-018-0450-0>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to G.I.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018

In the format provided by the authors and unedited.

Cyclic motifs in the Sardex monetary network

George Iosifidis ^{1,11*}, Yanick Charette^{2,11}, Edoardo M. Airoldi^{3,4}, Giuseppe Littera⁵, Leandros Tassiulas^{6,7} and Nicholas A. Christakis ^{6,8,9,10}

¹School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland. ²School of Social Work and Criminology, Laval University, Quebec City, Quebec, Canada. ³Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, USA. ⁴Institute for Quantitative Social Sciences, Harvard University, Cambridge, MA, USA. ⁵Sardex Spa, Sardinia, Italy. ⁶Yale Institute for Network Science, Yale University, New Haven, CT, USA. ⁷Department of Electrical Engineering, Yale University, New Haven, CT, USA. ⁸Department of Sociology, Yale University, New Haven, CT, USA. ⁹Department of Biomedical Engineering, Yale University, New Haven, CT, USA. ¹⁰School of Organization and Management, Yale University, New Haven, CT, USA. ¹¹These authors contributed equally: G. Iosifidis and Y. Charette. *e-mail: iosifidg@tcd.ie

Supplementary Information for

Title: Cyclic Motifs in the Sardex Monetary Network

Authors: George Iosifidis^{1†}, Yanick Charette^{2†}, Edoardo M. Airoldi³, Giuseppe Littera⁴, Leandros Tassiulas^{5,6}, and Nicholas A. Christakis^{5,7,8,9}

Affiliations:

¹School of Computer Science and Statistics, Trinity College Dublin, College Green, Dublin 2, Ireland.

²School of Social Work and Criminology, Laval University, Quebec, G1V0A6, Canada.

³Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA 19122, USA

⁴Sardex Spa, Sardinia, Italy.

⁵Yale Institute for Network Science, Yale University, New Haven, Connecticut 06520, USA.

⁶Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520, USA.

⁷Department of Sociology, Yale University, New Haven, Connecticut 06520, USA.

⁸Department of Biomedical Engineering, Yale University, New Haven, Connecticut 06520, USA.

⁹School of Organization and Management, Yale University, New Haven, Connecticut 06520, USA.

[†] George Iosifidis and Yanick Charette are jointly first authors.

- To whom correspondence should be addressed: George.Iosifidis@tcd.ie

Supplementary Note 1

1.1 Background Information

Sardex was initially used, at its introduction in 2010, to facilitate transactions among small- and medium-sized local enterprises so as to address the money liquidity problem in Sardinia due to the 2008 financial crisis⁸⁻¹⁰. It was subsequently used by these businesses to pay part of their employees' salaries, as a way to avoid cutting salaries or laying off employees. By the end of 2015, Sardex comprised 3,150 members (businesses and individuals) and had an annual turnover (aggregate volume of transactions) of approximately 50 Million Euros.

Sardex is a complementary currency, designed to operate in parallel with Euro (not to substitute it), and aiming to increase market liquidity (or, money velocity). The businesses join Sardex mainly for the following two reasons: First, since their customers were (or, still are) facing liquidity shortages, they were unable to sell their goods and services, having the so-called "untapped potential" problem. By joining Sardex, the businesses are trying to sell this stock in this complementary market. Second, when joining Sardex, businesses are granted a small overdraft in Sardex credits by the managing entity of the network. This helps them to overcome investment problems (not being able to borrow in Euro), and hence continue their production of commodities and services which then they can sell within the Sardex network, repaying this way their debt to the community.

When a business enters the Sardex network, it signs a contract, committing a certain amount of Euros (or products of commensurate value) and obtains access to a credit line. This is used essentially as a collateral for the case the member does not comply with its contractual obligations. While in Sardex, each business is free to trade with other businesses not in Sardex, using Euro. The businesses cannot directly exchange Euros with Sardex, as the latter is a closed economic system. However, it is obviously possible to have indirect exchanges among the two currencies: for instance, a company may buy various materials with Sardex and build products that later sells in the Euro market. In this paper, we do not have access to this type of indirect exchanges (nor Sardex Spa has), and hence we do not quantify or analyze them. The Sardex network has zero interest rate, as its main purpose is to incentivize circulation of money and serve as a means of exchange.

Sardex Spa is a member of the Sardex economy and can sell or purchase items and services from other members. Its main responsibilities are to ensure that Sardex members comply with their contractual terms (e.g., selling using their Euro-market prices), and maintain the information system that is necessary for the operation of Sardex. The latter comprises the servers, the electronic accounts and electronic directory of the system, the digital ledger that keeps track of the transactions, the electronic transaction mechanism, the mobile payment platform and so on. Furthermore, Sardex provides technical support to members upon request. For example, small businesses in rural areas might need help to use the system or find partners in the network, learn how to complete a transaction, etc. It is important to emphasize that Sardex is an open system, and all legal and economically-healthy businesses are admitted; in other words, Sardex Spa is not filtering the businesses that wish to enter the system and certainly cannot predict if they will be involved in cyclic transactions, be highly active in trading or not, etc. Finally, each Sardex member can leave the network, as long as it brings its credit balance to zero.

It is clear from the above that Sardex, and other similar insulated mutual-credit systems, cannot be analyzed through the lens of typical (macro-)economic models³⁴ for many reasons. For instance, in typical economies, the prolonged accumulations of capital at some members is not considered a disadvantage for them or the economy. Clearly this is not the case for Sardex where the goal is to have very high circulation of money (high liquidity) but at the same time avoid excessive credit and debts. Moreover, in such closed economic systems there might be uncertainty about the lack of resources (items, services, etc.) that will be of interest to its members, and this amplifies the problem of double coincidence of needs and wants. Besides, Sardex and similar community currencies are fundamentally different from other IOU models where each trader issues its own currency³⁴, because here all transactions involve multilateral commitment since Sardex can be exchanged between any two members.

Finally, it is interesting to note the relevance of Sardex-like systems (and hence the relevance of our analysis) to the emerging and increasingly important sharing economy applications. The term “sharing economy” is broadly used today to describe systems where humans exchange directly services and resources in a distributed fashion or through third parties. Using this definition, alternative currencies can be seen as a sharing economy paradigm. Moreover, many other sharing economy systems need accounting support (so as to keep track of exchanges and owed credit) that does not involve mainstream currency (e.g., time-banks, community-based labor exchange, etc.). These closed economies are essentially mutual credit systems, where the effort and resources of one member are credited to her and create an equal debt to other members. Therefore, mechanisms such as Sardex can be directly applied in sharing economy - and we believe our analysis sheds light on the operation of these systems as well. We refer the interested reader to references^{7, 38, 39} for a detailed taxonomy of such platforms.

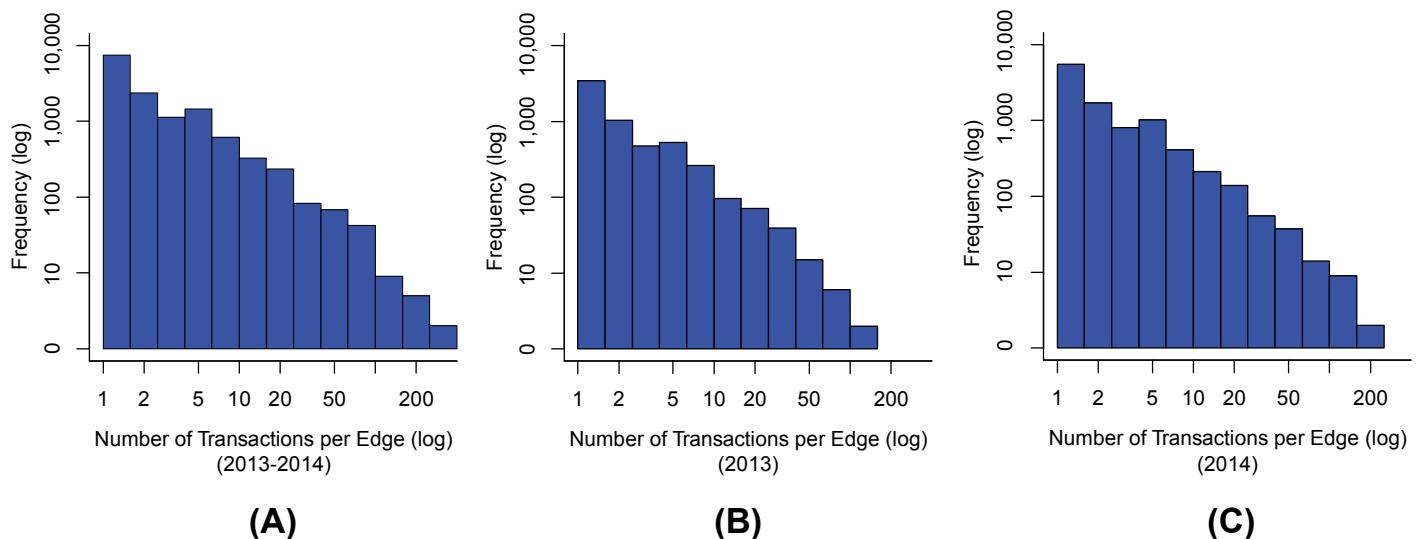
We use the transaction data of Sardex for the time period starting in January 1, 2013 until December 31, 2014, i.e., 24 months in total. This time period was the fastest-expanding interval for Sardex. It increased from 434 members and 1,852 cumulative transactions in February 2013, to 2,122 members (businesses and employees) and 65,529 cumulative transactions in December 2014. As shown in the sample below, the dataset comprises two basic tables: (i) the Transaction Matrix and, (ii) the Members Attributes Table. Note that the IDs of the buyers and sellers have been replaced to enhance the dataset’s privacy. Towards this direction, we only present aggregate population statistics, and have only used and analyzed past transactions (until December 31, 2014) so as to keep a time-window separation which is longer than one year, from the current transactions.

Supplementary Table 1. Transaction Matrix Sample and Members Attributes

Transfer.ID	BuyerID	SellerID	Amount	Date & Time		
SRD088	1	2	27.47	2/4/2013 14:15		
SRD956	3	4	100.00	2/1/2013 18:49		
Node	Activation Date	Capacity	Category	Subcategory	City	Province
1	9/2/2013	1000	Servizi	Associazione culturale	Arzana	Ogliastra
3	19/5/2013	30000	Ho.Re.Ca	Lavorazione marmi e graniti	Cagliari	Cagliari

1.2 Network Model

We have defined the Sardex network as follows: each node represents a Sardex member, and in particular a business, and every directed weighted edge captures the aggregate flow of currency from the buyer (source node) to the seller (destination node) during the time period of interest (the 24 months altogether or each year separately). This currency flow can be the result of one or more transactions. We have removed the transactions among businesses and their own employees since the latter have different qualitative and quantitative features ($n = 754$ employees). Namely, employees conduct smaller transactions than businesses and, most importantly, they cannot have debt since they are not assigned a credit line. For the resulting business-to-business (B2B) network, we have used all edges, even those comprising a single transaction. Namely, from the total of 13,753 edges, 7,439 capture a single transaction among the buyer-seller pair. Supplementary Figure 1 depicts the histogram of the number of transactions per edge in Sardex for each year separately, as well as for the 24 months altogether. The 1-transaction-only edges have a higher mean and median volume (mean = 512.0, median = 200.0) than the rest of the transactions (mean = 289.2, median = 97.2), and hence we have kept all of them for the network representation. Besides, removing edges in order to simplify representation (as those with 1 transaction only) would impact the credit balance of the nodes, and hence it is not a valid option.



Supplementary Figure 1. Number of Transactions Per Edge. **(A):** Distribution for the cumulative transactions of the 2 years ($E=13,753$ edges; 48,170 transactions). **(B):** Distribution for the transactions of the 1st year (2013; $E=5,962$ edges; 16,968 transactions). **(C):** Distribution of the transaction of the 2nd year (2014; $E=9,916$ edges; 31,202 transactions).

1.3 Descriptive Analysis of Sardex

Supplementary Table 2 below summarizes the features of the Sardex network. We provide information for each year separately, as we follow this annual separation in our statistical analysis. Supplementary Table 3 provides additional information for the nodes of the networks. The Sardex members are classified in six business categories, i.e.: 1) Construction (“Edilizia”), 2) Services (“Servizi”), 3) Retail (“Commercio al Dettaglio”), 4)

Primary Sector Industry (“Industria, artigianato e produzione”), 5) Hotel- Restaurant- Café (“Ho. Re. Ca”), 6) Wholesale, (“Commercio all’ingrosso”).

Supplementary Table 2. Description of Sardex network characteristics

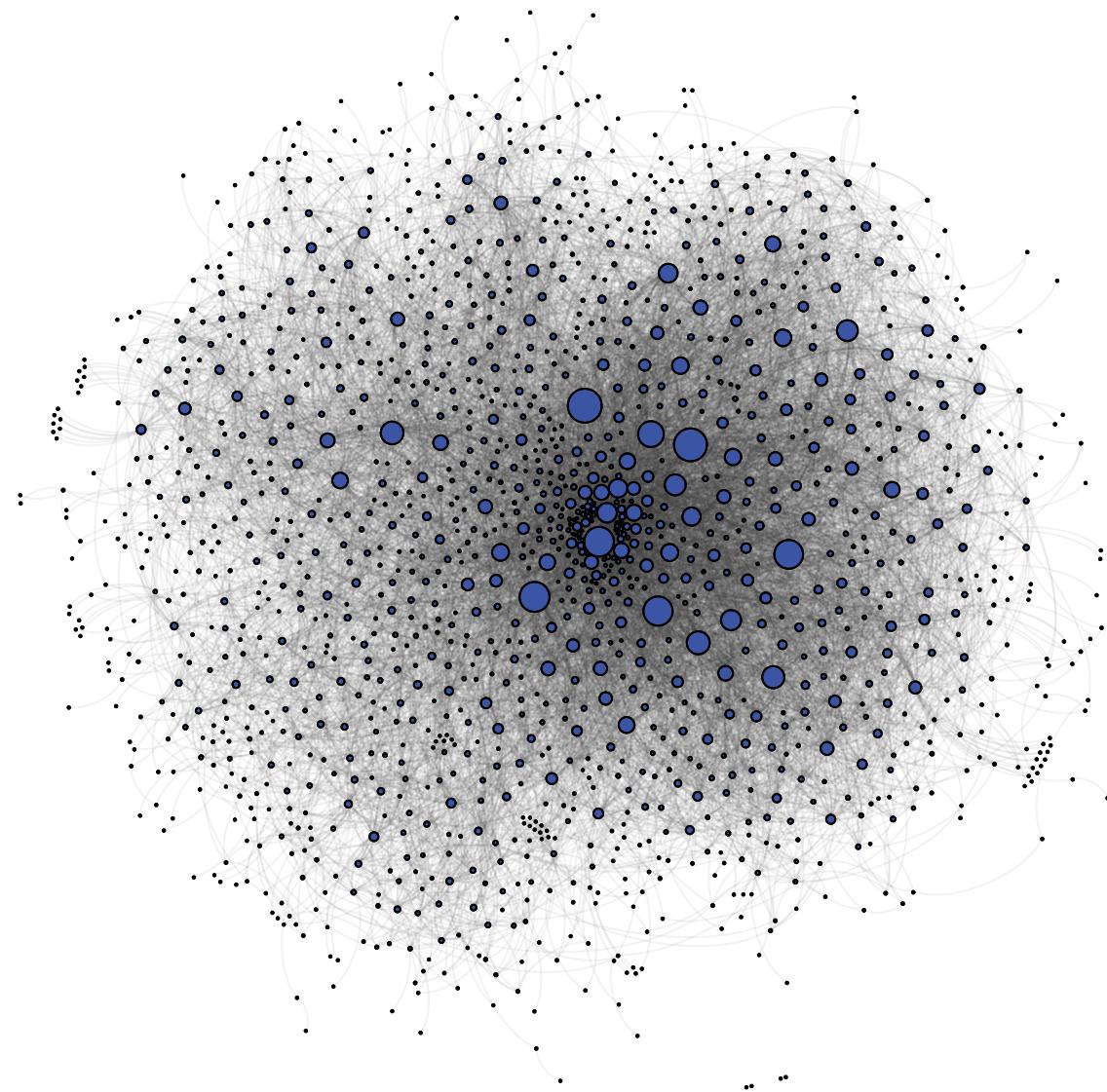
Network characteristics	Year 1 (2013)	Year 2 (2014)
Number of nodes	877	1,353
Number of buyers	763	1,163
Number of sellers	788	1,195
Number of edges	5,962	9,916
Density	0.78%	0.54%
Transitivity	0.11	0.10
Degree assortativity	-0.06	-0.06
Degree centralization	0.09	0.09
Number of cycles		
Length 2	59	1,170
Length 3	1,695	6,678
Length 4	31,583	106,578
Length 5	436,021	1,993,046
Num. of non-nested paths		
Length 2	179	284
Length 3	2,656	5,226
Length 4	42,628	93,604
Length 5	1,708,584	2,582,776

Supplementary Table 3. Description of Sardex Nodes

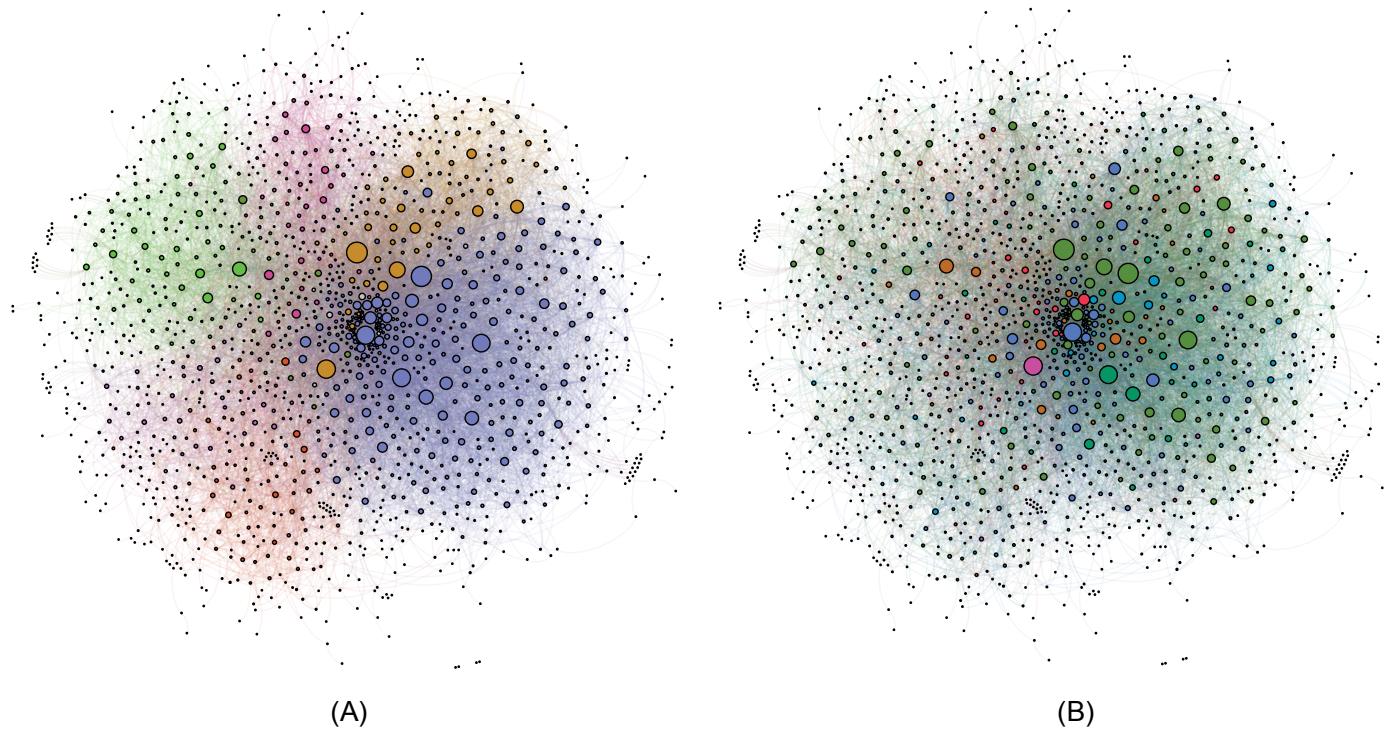
Node characteristics	Year 1 (2013)		Year 2 (2014)			
	(n=877)	n	%	(n=1,353)	n	%
Category						
Retail	194	22.1%		327	24.2%	
Wholesale	54	6.2%		78	5.8%	
Construction	67	7.6%		117	8.7%	
Ho. Re. Ca	111	12.6%		161	11.9%	
Industry	123	14.1%		189	14.0%	
Services	328	37.4%		481	35.5%	
	M	SD		M	SD	
Lifespan in the network (days)	241.20	99.70		249.06	119.91	
Average number of paths crossing						
Length 2	0.61	2.79		0.63	2.25	
Length 3	12.11	37.89		15.45	51.58	
Length 4	243.03	705.27		345.91	1,092.36	
Length 5	7,576.87	23,602.01		17,670.08	59,855.29	
Average number of cycles crossing						
Length 2	1.00	2.28		1.16	2.75	
Length 3	5.80	18.93		7.19	24.06	
Length 4	78.84	268.25		109.50	405.72	
Length 5	1,162.74	3,960.76		1,905.31	7,274.49	
Betweenness	1,726.31	4,878.06		2,553.19	7,579.00	
Unweighted degree	13.60	18.17		14.66	20.66	
Weighted degree	12,837.09	25,229.52		14,760.48	31,240.14	
Average absolute daily balance	2,444.49	4,162.54		3,832.77	8,622.26	
<hr/>						
Edge characteristics	(n=5,962)			(n=9,916)		
	M	SD		M	SD	
Volume (Sardex)	950.97	2,597.70		1,005.00	3,019.06	
Average number of paths crossing						
Length 2	0.06	0.52		0.06	0.40	
Length 3	1.35	5.06		1.60	6.16	
Length 4	28.92	78.53		38.29	106.51	
Length 5	874.66	2,052.20		2,198.41	9,023.96	
Average number of cycles crossing						
Length 2	0.16	0.36		0.17	0.37	
Length 3	0.87	1.68		1.00	1.90	
Length 4	11.84	21.14		15.25	28.05	
Length 5	174.54	309.41		265.34	497.51	
Betweenness	351.61	430.98		481.02	556.16	

In Supplementary Figure 2, we depict the Sardex network, where we have ignored the edges' directionality so as to improve visualization. Besides, Supplementary Movie 1 depicts the creation of edges over 1-year period (2013) in the city of Sassari. Additionally, in Supplementary Figure 3, we plot the network where the node size represents its weighted degree, and the node color its province location (Panel A) or its business category (Panel B). We observe that the nodes present assortativity in terms of province location ($V = 0.38$) but less in terms of business category ($V = 0.08$).

Supplementary Movie 1. Edges creation in city of Sassari during Year 2013. The nodes represent businesses that are located in Sassari, and the edges depicted their trading relationships (one or more transactions). The video presents the edges in a temporal sequence.



Supplementary Figure 2. Sardex network representation (2013-2014; N=1,477 businesses; E=13,753 partnerships). Node size is proportional to degree, edges represent trading relationships for the period of study.



Supplementary Figure 3. Sardex network (2013-2014) with category and location information. Node size is proportional to weighted degree ($N=1,477$ businesses; $E=13,753$ partnerships). **(A)**: Node color denotes location (9 different provinces: Nuoro (green), Ogliastra (dark green), Oristano (violet), Medio Campidano (orange), Carbonia-Iglesias (light blue), Quartucciu (grey), Cagliari (dark blue), Sassari (red), Olbia-Tempio (purple)). **(B)**: Node color denotes business category (6 different categories: serviced (dark blue), retail (green), industry (red), Ho. Re. Ca. (turquoise), construction (light blue), wholesale (orange)).

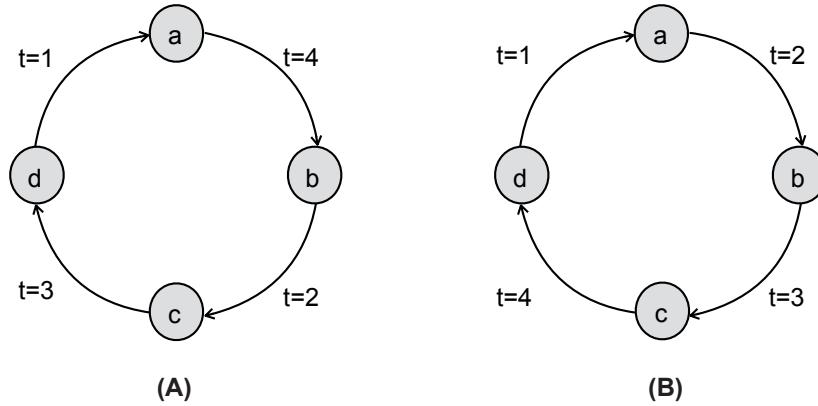
By the end of the time period of our study (December 2014), the top 20% of Sardex businesses with the largest number of partners (highest degree) were responsible for 67.9% of the overall trading edges (partnerships). Also, the top 20% of the businesses with the highest monetary value of transactions made 98.4% of the total monetary value of transactions.

Supplementary Note 2

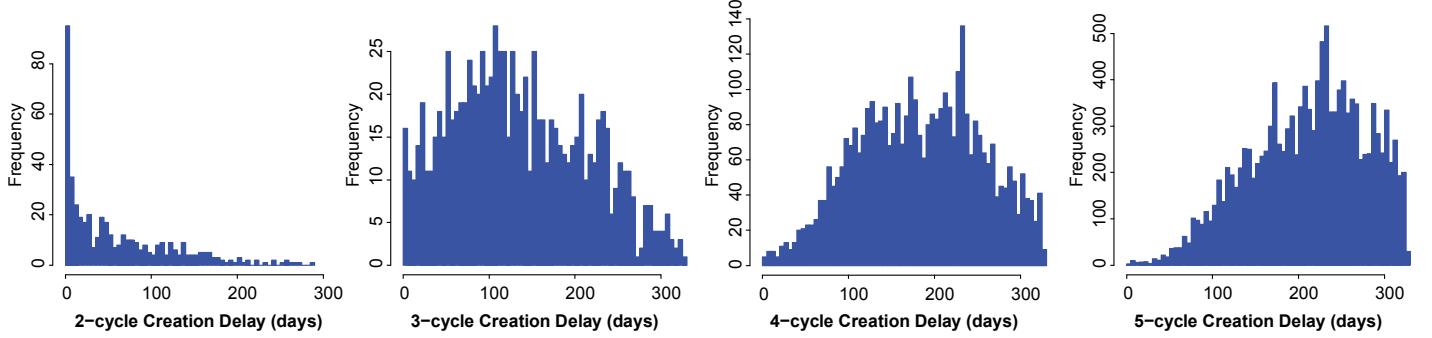
2.1 Alternative Definitions about Cycles and Examples

We provide additional information about the cycles and paths, clarifying the differences between sequential and non-sequential cycles, as well as between nested and non-nested paths. As stated above, a cycle is a succession of nodes where the first and the last nodes are identical, and they are pairwise connected through directed edges. By this definition, cycles do not contain repeated edges. Also, two cycles are considered distinct if one is not a cyclic permutation of the other.

It is clear that, in many cases, the edges of the cycles are not created in time order (sequentially). This is clarified with an example in Supplementary Figure 4 where we present the difference between *non-sequential* and *sequential* cycles. Then, one can measure the actual time it takes for a sequential cycle of a certain length to be created. The distribution of these time intervals (delays) is depicted in Supplementary Figure 5 and their basic statistics in Supplementary Table 4. Based on these results, we decided to use up to 5-cycles, since longer cycles have larger delay that practically exceeds the duration of our dataset. More importantly, the statistical analysis results are essentially independent of whether we use the entire population of cycles (sequential and non-sequential) or only the sequential cycles. Therefore, we used the largest population, i.e., both sequential and non-sequential cycles combined (see Supplementary Table 5 for population size comparisons).



Supplementary Figure 4. Distinction among sequential and non-sequential cycles. (A): Cycle of length 4, the edges of which have not been created in a time-sequential fashion. (B): Cycle of length 4 with time-sequence in the creation of its edges. The times shown are hypothetical and can represent, for example, different days.



Supplementary Figure 5. Delay in cycle formation for year 2013. The results for year 2014 are similar. Cycles of small length have skewed distribution, while cycles of larger length an almost normal distribution. The mean value of 5-cycles creation delay is 211.2 days (233.3 for 2014), and hence this was the largest size of cycles that we considered.

Supplementary Table 4. Mean and median values of (sequential) cycle formation delay.

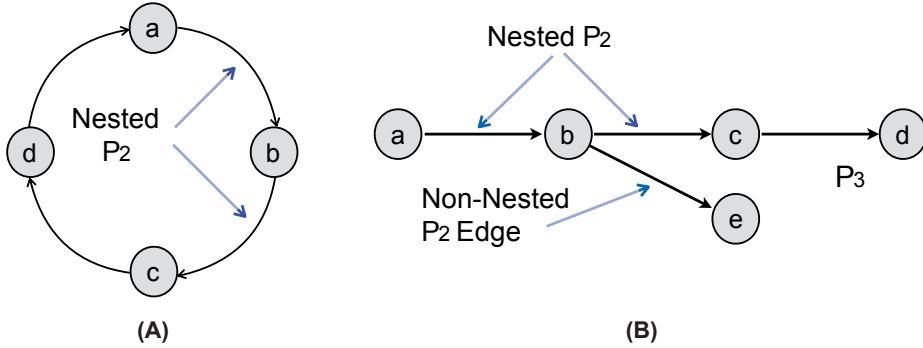
Year	2- Cycle	3- Cycle	4-Cycle	5-Cycle
2013 (Mean/Median)	60.2 (41)	135.4 (126)	183 (184)	211.2 (219)
2014 (Mean/Median)	72.4 (44)	157.3 (150)	202.2 (207)	233.3 (244)

Supplementary Table 5. Number of cycles of different length in Sardex by year

Year	Nodes	Edges	Type of Cycles	2- Cycle	3- Cycle	4-Cycle	5-Cycle
2013	877	5,962	All Cycles	59	1,695	31,583	436,021
			Sequential Only	59	939	3,786	13,556
2014	1,353	9,916	All Cycles	1,170	6,678	106,578	1,993,046
			Sequential Only	1,170	1,665	7,289	32,321

Another important concept for our analysis is the *non-nested* paths. Recall that a path is a sequence of connected nodes, where no node appears more than one time. Essentially, paths can be considered as “unclosed” cycles. We use hereafter the term non-nested paths to describe the paths that are not fully embedded in longer paths or cycles. This definition allows partial overlapping, as it is shown in Supplementary Figure 6(B). For the purposes of the statistical analysis it was crucial to be able to assess the impact of edges that belong only to paths of a certain length, i.e., do not belong concurrently in many paths of different length, or concurrently in paths and cycles. Therefore, we used only the non-nested paths in our analysis. Please note in Supplementary Table 2 that the population sizes of non-nested paths and cycles are of the same order of magnitude. Finally, note that in order to identify

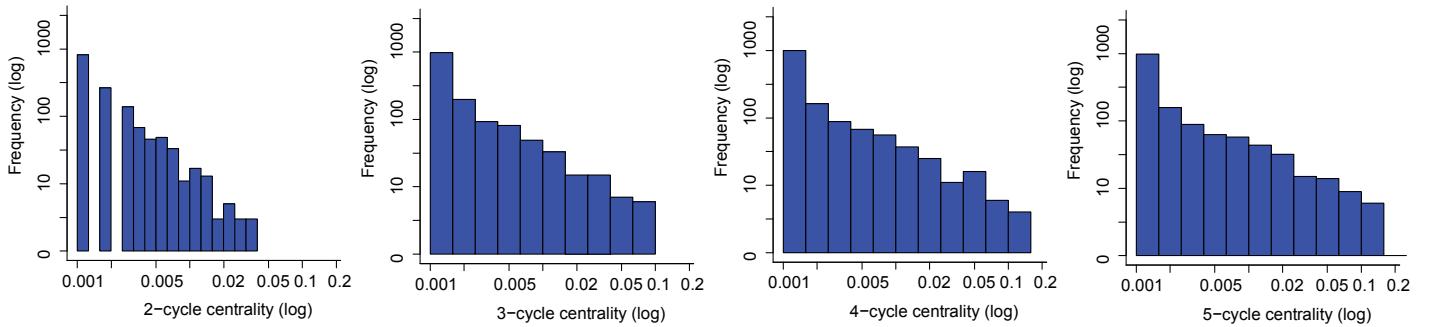
the non-nested paths of length 5, we had to enumerate and identify the paths up to length 6. Given that paths lengthier than 6 hops certainly include the paths of length 6, this methodology ensures that the non-nested paths of length 2, 3, 4, and 5 that we use, are indeed not part of paths of lengths 3, 4, 5, 6 or more hops.



Supplementary Figure 6. Nested paths. (A): Path of length 2 nested in a 4-cycle. (B): Path of length 2 nested in a path of length 3.

2.2 Distribution and Correlations of Cycle Centralities in Sardex and Random Graphs

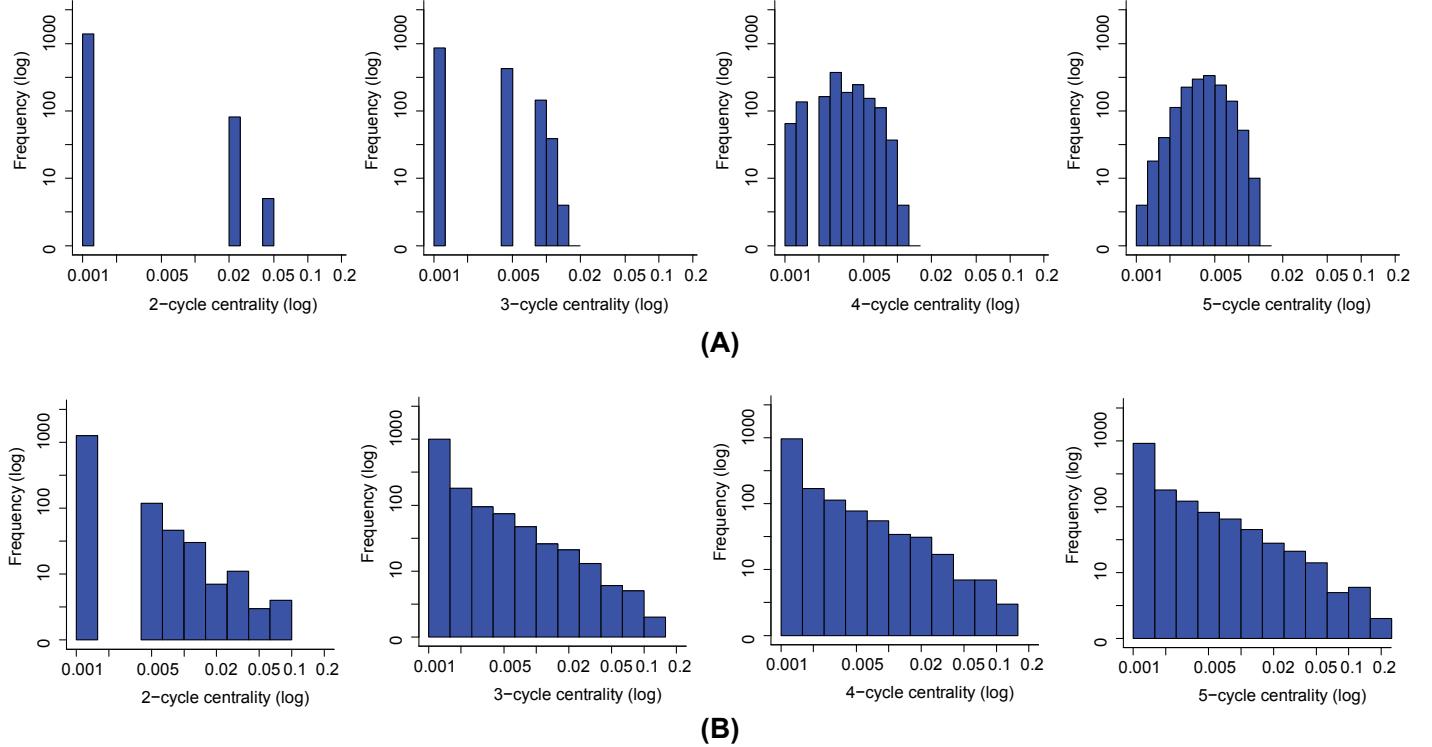
We provide in this subsection additional information about the cycle centrality metrics that we introduced, and their relation to other typical centrality measures. The distribution of node k-cycle centralities in Sardex (for both years) is presented in Supplementary Figure 7. We observe the long tail of the distributions, especially for 4-cycle and 5-cycle centralities. Also, a large subset of nodes has k-cycle centrality equal to zero.



Supplementary Figure 7. Distribution of k-cycle node centrality for the Sardex network (2013-2014). Note that 825 nodes have 2-cycle centrality (Cy_2) equal to zero; 681 nodes have zero 3-cycle centrality ($Cy_3=0$; 454 nodes have $Cy_4=0$; and 366 nodes have $Cy_5=0$, out of total $N=1,477$ nodes. The figure has been created with Supplementary Code 1.

Moreover, we compare these results with the distributions of cycle centralities in two classes of null models, namely the R-D graph model (random graph derived from Sardex with edge reshuffling such

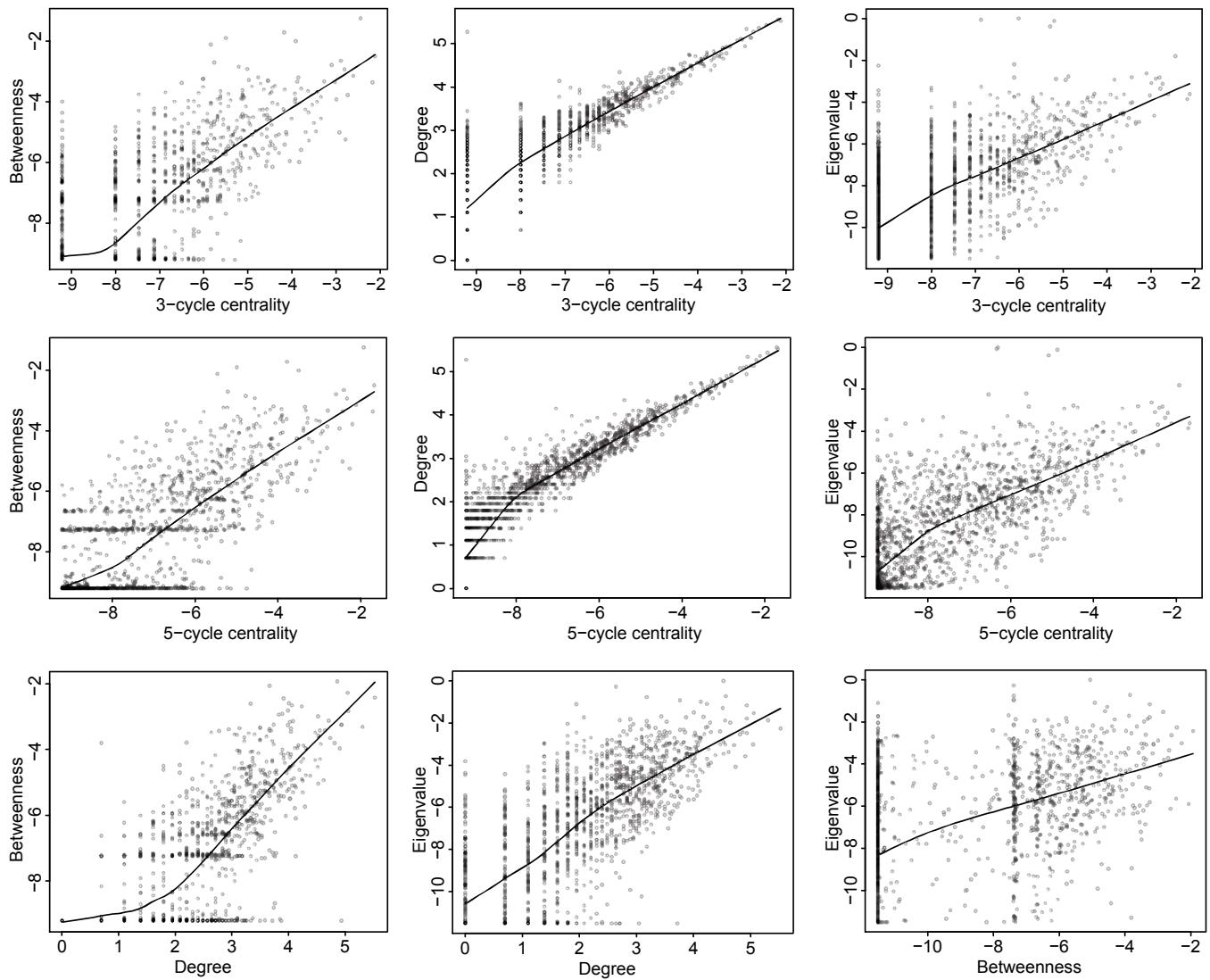
that it preserves the in/out-degree distribution) and the Erdos-Renyi (E-R) graph. Supplementary Figure 8 shows the distributions for the null models. First, we observe that the distribution is fundamentally different in the E-R graph (Supplementary Figure 8(A)), i.e., very few nodes have non-zero 2-cycle centrality, and the 5-cycle centralities are (almost) normally distributed. On the other hand, the R-D graph has a k-cycle distribution that is more similar to the observed Sardex graph (Supplementary Figure 8(B)), although the number of nodes having zero k-cycle centralities is significantly smaller in Sardex. The results have been verified using multiple instances of the null models.



Supplementary Figure 8. Distributions of k-cycle centrality in random graphs. (A): Distribution of k-cycle node centrality for a E-R graph, which has the same dimension ($N=1,477$ nodes; $E=13,753$ edges). The distribution of nodes with zero k-cycle centrality is: 1272 with $Cy_2=0$; 839 with $Cy_3=0$; 55 with $Cy_4=0$; and 2 with $Cy_5=0$. (B): Distribution of k-cycle node centrality for a random R-D graph (reshuffled graph that preserves the in/out degree distribution of the Sardex network). The distributions are similar with the ones in the observed Sardex graphs, but the population sizes of nodes with zero cycle centrality is substantially different: 1118 with $Cy_2=0$; 714 with $Cy_3=0$; 383 with $Cy_4=0$; and 294 with $Cy_5=0$. The figure has been created with Supplementary Code 2 and 3.

Finally, we study the correlation of cycle centralities with other typical centrality measures. Namely, in Supplementary Figure 9, we present the detailed relation of the 3-cycle and 5-cycle centralities (omitting 2-cycle and 4-cycle for brevity) with the betweenness, degree and eigenvector centrality of nodes in Sardex. We observe that the correlations are low in regions where the cycle centrality values are low to medium, but are higher when the cycle centralities are very high (see right corner of the plots). In every case, these correlations are comparable (and even lower in most cases) than

the respective correlations among other centrality metrics in Sardex or in other networks^{40, 41}, and they are similar to the correlation of other novel and recently introduced centralities with degree and betweenness. The detailed correlation coefficients are shown in Supplementary Table 6.



Supplementary Figure 9. Centrality Correlations. Correlation (log-scale) of Betweenness, Degree, Eigenvalue and Cycle Centrality, for businesses in Sardex network (2014; N=1,353 businesses).

Supplementary Table 6. Correlations among various Centrality Metrics, Year 2014 (N=1353 businesses).

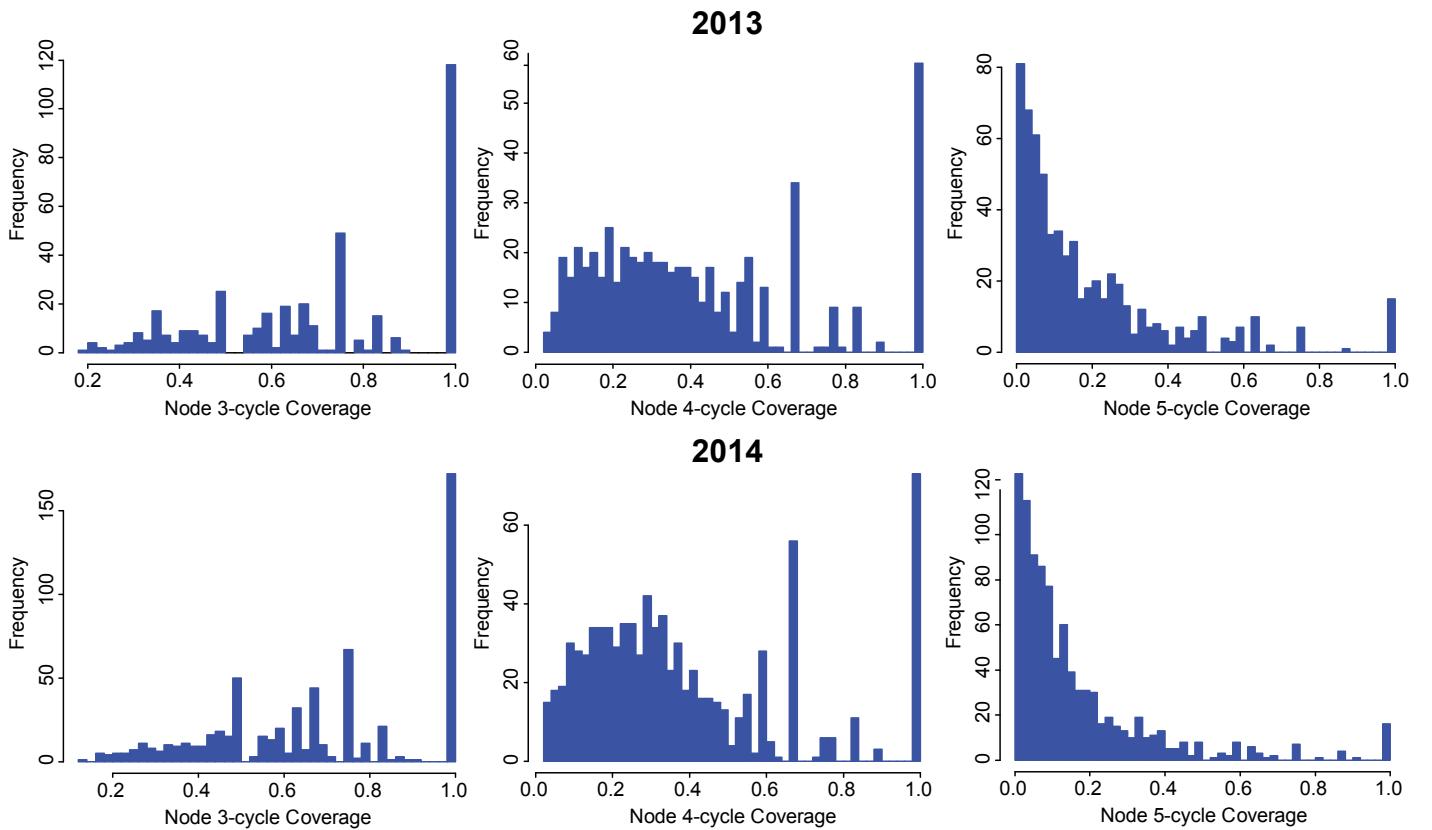
k-cycle	Betweenness $\rho(r)$	Degree centrality $\rho(r)$	Other Comparisons	Centrality $\rho(r)$
2-cycle	.66 (.53)	.76 (.89)	Betweenness / Degree	.77 (.54)
3-cycle	.70 (.55)	.88 (.87)	Eigenvalue / Degree	.74 (.14)
4-cycle	.70 (.55)	.92 (.86)	Eigenvalue / Betweenness	.48 (.08)
5-cycle	.70 (.55)	.92 (.85)	Eigenvalue / 5-cycle	.79 (.10)
T-cycle	.74 (.55)	.92 (.87)	Eigenvalue / 3-cycle	.70 (.11)

ρ : Spearman's rho ; r : Pearson's r

It is important to emphasize at this point that for low values of the k-cycle centrality the correlation with degree and betweenness centrality is relatively low. For example, if we consider the 70% of the nodes with the lowest 3-cycle centrality (936 out of 1353 nodes for year 2014), then the Spearman correlation between 3-cycle centrality and degree is just .54. Similarly, the correlation between 4-cycle centrality and degree is .74 (correlation of the 70% of nodes having lower 4-cycle centralities), and the respective value for 5-cycle centrality is .77. These results reveal that the seemingly high correlation values of Supplementary Table 6 are due to the strong correlation between the high k-cycle centralities with the corresponding degree and betweenness centralities. This is evident from Supplementary Figure 9.

2.3 Distribution of Cycle Coverage in Sardex Graphs

We present in this subsection the distribution of cycle coverage values for the nodes in Sardex, for each year. The results are depicted in Supplementary Figure 10. We observe that 3-cycle and 4-cycle coverage values span almost uniformly across the entire range, with a high pick in value 1. This means that a large subset of the nodes has non-overlapping cycles. On the other hand, 5-cycle coverage values are concentrated in low ranges, which indicates a large overlapping of 5-cycles.



Supplementary Figure 10. Distribution of Cycle Coverage values. Upper row of histograms presents the results for the 1st year (2013; N=877 businesses); and the lower row of histograms the results for the 2nd year (2014; N=1,353 businesses).

2.4 Definitions of Credit Healthiness

We formally introduce here the credit healthiness metric for each node n and for the time interval t_n (first day) to t_G (last day) as follows:

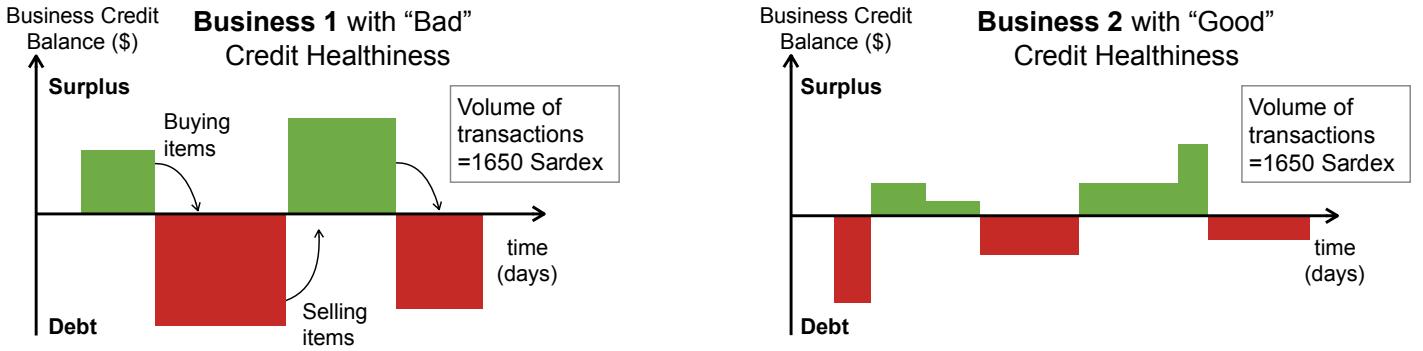
$$H(n) = \frac{1}{t_G - t_n} \sum_{t=t_n}^{t_G} |B(n, t)|$$

where the quantity:

$$B(n, t) = \sum_{t=t_n}^{t_G} \left(\sum_{i=1}^N e_{in} - \sum_{j=1}^N e_{nj} \right)$$

is the credit balance of node n, positive or negative, at the end of each day t. Note that this balance is calculated using the aggregate incoming (due to selling) and outgoing (due to buying) edge weights (i.e., e_{in} and e_{nj} , respectively) that have been realized up to day t. Also, please note that the credit healthiness metric takes into account the day each node has joined the system, i.e., it is normalized by the node lifespan.

An illustrative explanation of the node credit healthiness metric is shown in Supplementary Figure 11. Intuitively, if two nodes n and m have the same volume of transactions during a certain time interval, but one of them has higher credit healthiness, i.e., $H(n) > H(m)$ then we consider node m performing higher than n . Essentially, the credit healthiness is inversely proportional to the average absolute balance of the node, if we omit the negative sign of the negative balances (debts).



Supplementary Figure 11. Credit healthiness example. A hypothetical example of the credit balances of two nodes that are of the same type (i.e., trade the same product), have the same capacity, and equal turnover (1650 SRD). The node on the left has bad ("unhealthy") credit balance as it spends prolonged time periods with very high credit surpluses or debts, unlike the node on the right.

We also consider an alternative definition for the credit healthiness metric denoted $H_B(n)$ which is more sensitive on the time dimension. That is, we use the expression for $H(n)$ but do not normalize with the time duration ($t_G - t_n$). This alternative metric succeeds in identifying nodes that have been idle with a non-zero credit surplus (or debt) for a very long time period. Namely, these nodes will have a low (and hence desirable) value of $H(n)$ but a very high $H_B(n)$ value. The latter metric essentially captures the aggregate (accumulated) absolute balance of the node during a certain time interval. At this point, it is important to emphasize that the statistical analysis is robust in using both these different credit healthiness metrics.

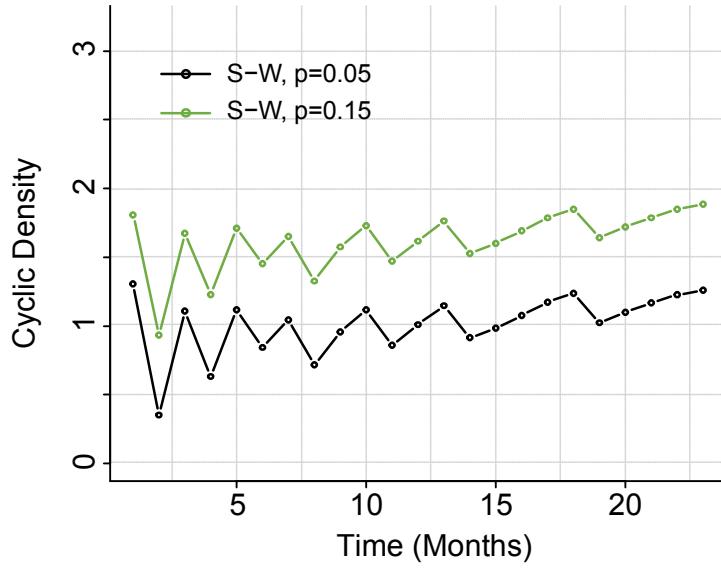
2.5 Discussion about Cyclic density

We provide here the closed form expressions for the expected number of k -cycles in a R-D graph. These quantities were used above in order to define the cyclic density in Fig. 3A. The result is obtained from¹⁷. Specifically, a directed graph G_N that has D_{in} and D_{out} in- and out-degree distributions respectively, has an expected number $P_k(G_N)$ of k -cycles:

$$E(|P_k(G_N)|) = \frac{1}{k} \left(\frac{\langle D_{in} D_{out} \rangle}{\langle D_{in} \rangle} \right)^k$$

This result holds for relatively small cycle lengths which, for the graphs here, is up to 10-cycles. Similar analytical formulas have been proposed for undirected scale-free networks⁴². Furthermore, for the

calculations of cyclic density in Fig. 3A we have used a population of 200 representative Erdos-Renyi graphs that have the same number of edges and nodes, and also a population of 200 Small-World (S-W) graphs, using the `watts.strogatz()` function in `igraph`⁴³. The “neighborhood” parameter for the S-W has been selected so as to yield a comparable number of edges with the observed Sardex graphs. Moreover, the rewiring probability has been set to .15, but we have tested also other values and the obtained conclusions do not change. For example, we present in Supplementary Figure 12 the cyclic density for S-W with rewiring probability .05. In all cases, Sardex is found having a larger number of cycles compared to all null models.



Supplementary Figure 12. Sardex Cyclic Density. Comparison with the Small-World graph, $p=0.05$ and $p=0.15$; presented results are averaged for a population size of 200 graphs, in both cases. The figure has been created with Supplementary Code 5.

2.6 Node co-location and cycles

Finally, we study the role of location in the creation of cycles. In particular, we explore whether there is a co-location effect whereby nodes that are in proximity (within the same city) are more likely to participate in cycles and create transactions of higher value. This would imply that the co-location might be a more important driving factor than the cycles per se. However, we observe that this is not the case. We benchmark all findings with the respective results for the non-nested paths.

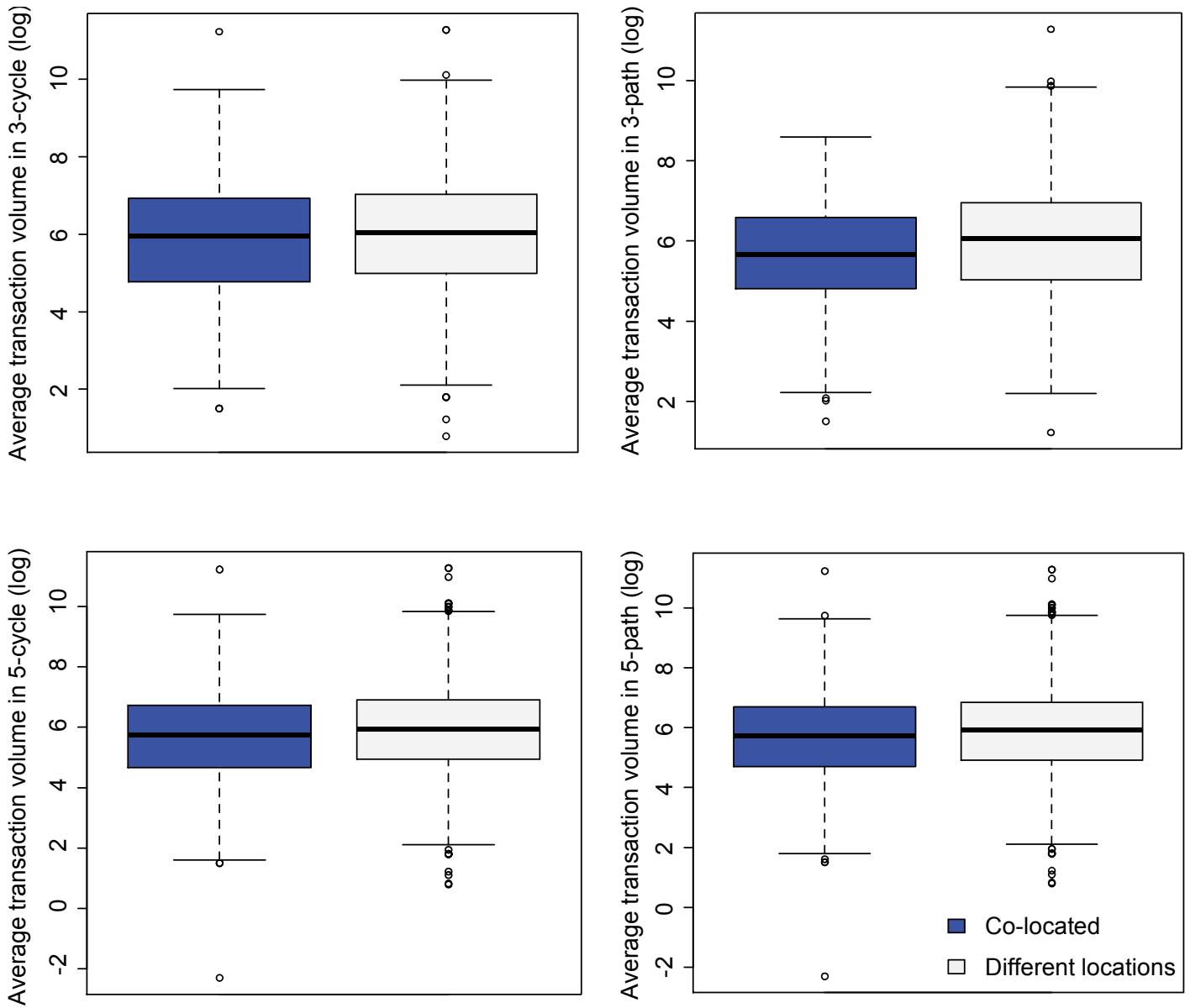
Supplementary Table 7. Portion of co-located nodes that participate in edges in cycles or paths (2013: N=877; 2014: N=1,353).

Collocated Edges (%)	2-Cycle	2-Path (P2)	3-Cycle	3-Path (P3)	4-Cycle	4-Path (P4)	5-Cycle	5-Path (P5)
Year 2013	32.95	22.10	24.76	20.00	21.18	20.50	21.25	21.47
Year 2014	38.01	23.61	31.59	24.83	27.24	25.56	26.55	26.25

First, Supplementary Table 7 shows the percentage of edges in cycles and paths for which the trading nodes are co-located in the same city. We see, for example, that only 21.2% of the edges in cycles of length 4 formed during 2013 involve nodes that are co-located. Clearly, the majority of cycle edges are formed across difference cities; while the co-location percentages are similar to those observed for the paths. In other words, there is no strong co-location effect and, in any case, the observations in cycles are comparable to the paths. Supplementary Table 8 presents the weights of edges among co-located nodes and among nodes in different cities. We provide the mean, the median, the results of a t-test, and the respective box-plots (Supplementary Figure 13). It can be observed that most cycle and path edges are not co-located, and also that the non-co-located edges “perform” slightly better. The above results hold for both years and persist if we control for the business category. Next, we also quantify the effect of the actual geodesic distance among the nodes.

Supplementary Table 8. Rounded weights of collocated and non-co-located edges in cycles and paths (2013: N=5,962)

Mean (Median)	2-Cycle	P2	3-Cycle	P3	4-Cycle	P4	5-Cycle	P5
Collocated	1092 (434)	652 (293)	996 (384)	551 (289)	841 (316)	871 (328)	822 (310)	790 (302)
Different City	1597 (551)	857 (342)	1249 (423)	1209 (424)	1104 (400)	1115 (400)	1040 (382)	994 (369)
p-value (t-test)	0.050	0.130	0.060	<0.001	0.006	0.030	0.009	0.005



Supplementary Figure 13. Box plots of Edge Weights for Co-located Nodes Edge weights for co-located and non-co-located nodes in 3-cycle (A), P3 (B), 5-cycle (C) and 5-path centrality (D). Details of the statistics and populations are shown in Supplementary Table 8. (2013: N=5,962)

Supplementary Note 3

In this section, we present the generalized linear models we used in the main paper and their diagnostics, and we describe a suite of auxiliary models we used to evaluate the robustness of our findings to model specification. This includes various lagged models and alternative auto-regression models, as well as models with random or fixed effects. The results of the sensitivity analyses with these auxiliary models support the claims reported in the main text. As noted above, we construct and analyze two instances of the Sardex network, namely one for the transactions of each year. In our statistical analysis, we used these two networks in a unique model as panel data.

3.1 Brief data description and variables for models

We use the transaction data for the Sardex businesses in years 2013 and 2014, as well as the features of these businesses. In particular, for each business, the dataset contains the date it joined the system, its location (province), and the type of business (6 basic categories). For each transaction, we have the exact day and time that it was committed, along with the buyer ID, the seller ID, and the amount of transferred money. The available variables are summarized in Supplementary Table 9.

Supplementary Table 9. Summary of the variables used in the models

Control Variables

$t \in \{1, 2\}$: time dimension (slots), $t=1$ for year 2013, and $t=2$ for year 2014.

N_t : number of nodes in the system at the end of year 2013 ($N_1=877$) and 2014 ($N_2=1353$).

SN_t : set of nodes at the end of each year; $N_t = |SN_t|$, $SN_1 \cap SN_2 \neq \emptyset$, $SN_1 \cup SN_2 = SN$, $N = |SN| = 1477$. We also define $N_T = N_1 + N_2 = 2230$ (total observations at the node level).

Node Variables

$x_{C2}^t = \{x_{iC2}^t : i = 1, \dots, N_t\}$: number of cycles of length 2 crossing node i in slot t .

$x_{C3}^t = \{x_{iC3}^t : i = 1, \dots, N_t\}$: number of cycles of length 3 crossing node i in slot t .

$x_{C4}^t = \{x_{iC4}^t : i = 1, \dots, N_t\}$: number of cycles of length 4 crossing node i in slot t .

$x_{C5}^t = \{x_{iC5}^t : i = 1, \dots, N_t\}$: number of cycles of length 5 crossing node i in slot t .

$x_{P2}^t = \{x_{iP2}^t : i = 1, \dots, N_t\}$: number of paths of length 2 crossing node i in slot t .

$x_{P3}^t = \{x_{iP3}^t : i = 1, \dots, N_t\}$: number of paths of length 3 crossing node i in slot t .

$x_{P4}^t = \{x_{iP4}^t : i = 1, \dots, N_t\}$: number of paths of length 4 crossing node i in slot t .

$x_{P5}^t = \{x_{iP5}^t : i = 1, \dots, N_t\}$: number of paths of length 5 crossing node i in slot t .

$x_6^t = \{x_{i6}^t : i = 1, \dots, N_t\}$: total weighted degree of node i in slot t (end of the year).

$x_7^t = \{x_{i7}^t : i = 1, \dots, N_t\}$: duration of node i in slot t (i.e., 365 minus the day it joined).

$q = \{q_i : i \in SN\}$: category of business i ; categorical variable with 6 values.

$y^t = \{y_i^t : i \in N_t, y_i^t \in \mathbf{R}^+\}$: average absolute balance of node i , i.e., average (absolute) daily balance in a year. (see Sec 4.4 for definition).

Edge Variables

$r_{C2}^t = \{r_{ijC2}^t : i, j = 1, \dots, N_t\}$: number of cycles of length 2 crossing edge (i, j) in slot t .

$r_{C3}^t = \{r_{ijC3}^t : i, j = 1, \dots, N_t\}$: number of cycles of length 3 crossing node (i, j) in slot t .

$r_{C4}^t = \{r_{ijC4}^t : i, j = 1, \dots, N_t\}$: number of cycles of length 4 crossing node (i, j) in slot t .

$r_{C5}^t = \{r_{ijC5}^t : i, j = 1, \dots, N_t\}$: number of cycles of length 5 crossing node (i, j) in slot t .

$r_{P2}^t = \{r_{ijP2}^t : i, j = 1, \dots, N_t\}$: number of paths of length 2 crossing node (i, j) in slot t .

$r_{P3}^t = \{r_{ijP3}^t : i, j = 1, \dots, N_t\}$: number of paths of length 3 crossing node (i, j) in slot t .

$r_{P4}^t = \{r_{ijP4}^t : i, j = 1, \dots, N_t\}$: number of paths of length 4 crossing node (i, j) in slot t .

$r_{P5}^t = \{r_{ijP5}^t : i, j = 1, \dots, N_t\}$: number of paths of length 5 crossing node (i, j) in slot t .

$z^t = \{z_{ij}^t : i, j \in N_t, z_{ij}^t \in \mathbf{R}^+\}$: weight of edge (i, j) at the end of slot t .

3.2 Edge Models Definition

Random Effects Model

We used a random effects model with varying intercept for the analysis of the edge weights. In particular, the model for cycles (and paths) of length k can be written:

$$(M1): \log(z_{ij}^t) = a_{ij}^t + \beta_{ck} \log(r_{ijk}^t) + \beta_{pk} \log(r_{ijk}^t) + \beta_{db} \log(x_{i7}^t) + \beta_{ds} \log(x_{j7}^t) + \beta_{cb} q_i + \beta_{cs} q_j + \beta_\tau t + U_i + U_j + \varepsilon_{ijt}$$

where U_i is the between node-buyer errors, U_j is the between node-seller errors, and ε_{ijt} is the within edge errors (for the different years). We assume that $U_i, U_j \sim N(0, \sigma^2)$. Coefficients β_{ck} and β_{pk} capture the k-cycle and k-path effects, respectively. Also, please note that we have different predictors and coefficients for the category (β_{cb}, β_{cs}) and duration (β_{db}, β_{ds}) of the buyer ("b") and seller ("s"). We have log-transformed both the dependent and the independent variables since they have skewed distributions. We define the models for the impact of 2-cycles, 3-cycles, 4-cycles and 5-cycles according to equation (M1) above (setting the value of k accordingly). For fitting the model we employed the *lmer()* function from *lme4* package⁴⁴ in *R*. The results of the random effect model for the edges are reported in Supplementary Table 10.

Fixed Effects Model

We also used a fixed effect model to assess sensitivity of the results with the previous model to the Normal assumptions on the node-buyer and node-seller errors. In this case, the model is described by eq. (M1) where the variables U_i and U_j are being substituted as follows:

$$U_i = \gamma_{2b} I_{2b} + \gamma_{3b} I_{3b} + \dots + \gamma_{Nb} I_{Nb}, \quad U_j = \gamma_{2s} I_{2s} + \gamma_{3s} I_{3s} + \dots + \gamma_{Ns} I_{Ns}$$

where $I_{nb}, I_{ns} \in \{0,1\}$, $n = 2, \dots, N$, are the $N - 1$ binary (dummy) variables, per node-buyer and node-seller respectively, and γ_{nb}, γ_{ns} , $n = 2, \dots, N$, the respective coefficients. For fitting the model we employed the *lm()* *R* function. The results of the fixed effect model for the edges are reported in Supplementary Table 20.

Time-Lagged Model

Finally, we also used a lagged model to assess the impact of independent variables calculated in year 2013 ($t_1 = 1$) on the edge weight in year 2014 ($t_2 = 2$). The model equation in this case can be rewritten as follows:

$$(M2): \log(z_{ij}^{t_2}) = a_{ij}^{t_2} + \beta_{ck} \log(r_{ijk}^{t_1}) + \beta_{pk} \log(r_{ijk}^{t_1}) + \beta_{db} \log(x_{i7}^{t_1}) + \beta_{ds} \log(x_{j7}^{t_1}) + \beta_{cb} q_i + \beta_{cs} q_j + \varepsilon_{ijt_2}$$

We used the *lmer()* function in *R*⁴⁴. The results of the time-lagged model for edges are reported in Supplementary Table 16.

3.3 Node Models Definition

The general models for the analysis of the impact of k -cycles on node credit healthiness is the following:

$$(M3): \log(y_i^t) = a_i^t + \beta_{Ck} \log(x_{iCk}^t) + \beta_{Pk} \log(x_{iPk}^t) + \beta_w \log(x_{i6}^t) + \beta_d \log(x_{i7}^t) + \beta_c q_i + \beta_\tau t + U_i + \varepsilon_{it}$$

Coefficients β_{Ck} and β_{Pk} capture the k -cycle and k -path effects, respectively. U_i is the between-nodes error and ε_{it} the within-node error (for the different years). We define the models for the impact of 2-cycles, 3-cycles, 4-cycles and 5-cycles according to (M3) (setting properly the value of k). Note that we again have log-transformed both the dependent and the independent variables since they have skewed distributions. The auxiliary node models below stem from equation (M3) by defining properly the error effect U_i .

The specific models we employed differ in the full specifications of the terms U_i , as follows.

Ordinary Least-Squares Model

For the ordinary least-squares (OLS) model we used eq. (M3) by setting $U_i = 0$. For fitting the model we employed the *lm()* function in *R*. The results of the OLS model for the nodes are reported in Supplementary Table 21.

Random Effects Model

We used a random effects model with varying intercept where the two time observations (level 1) are nested at the nodes (level 2). Note that for some nodes we only have one observation. In particular, of the total 1,477 nodes at the end of year 2014, the 877 were present during year 2013. For the random effect we assumed a normal distribution, i.e., $U_i \sim N(0, \sigma^2)$. For fitting the model we employed the *lmer()* function from *lme4* package⁴⁴. The results of the random effect model for the nodes are reported in Supplementary Table 21.

Fixed Effects Model

We used a fixed effects model by including one fixed predictor for each node. Namely, this model stems from eq. (M3) by substituting:

$$U_i = \gamma_2 I_2 + \gamma_3 I_3 + \cdots + \gamma_N I_N ,$$

where $I_n \in \{0,1\}, n = 2, \dots, N$, are the $N - 1$ binary (dummy) variables and $\gamma_n, n = 2, \dots, N$, the respective coefficients. For fitting the model we employed the *lm()* function.

As explained below, the fixed effects model for the node analysis does not fit the data properly (having very high standard errors and AIC values), and hence we have only included it as a possible candidate for the sensitivity analysis; that is, we deem these results unreliable and do not report them.

Distance Error Model

We used two network autocorrelation models. The first one is the distance error model that employs the node distances in the distribution of the error term. Namely, the value of U_i in eq. (M3) for this model, is a component of the $N_T \times 1$ matrix \mathbf{U} which is recursively defined as follows:

$$\mathbf{U} = \lambda \mathbf{W} \mathbf{U}$$

where \mathbf{W} is the $N_T \times N_T$ social distance matrix for which the following hold:

- Every component w_{ij} represents the weighted shortest-path distance from node i to node j .
- For $i, j = 1, \dots, N_1$, w_{ij} is the shortest-path from i to j in the Sardex graph that is constructed using the transactions of year 2013.
- For $i, j = N_1 + 1, \dots, N_1 + N_2$, w_{ij} is the shortest-path from i to j in the Sardex graph that is constructed using the transactions of year 2014.
- For $i = 1, \dots, N_1$, $j = N_1 + 1, \dots, N_2$ it is, by definition, $w_{ij} = 0$ if j is the different time instance (observation) of node i , and $w_{ij} = \infty$ otherwise.

Finally, λ is the relational autoregressive parameter. For fitting the model we employed the *errorsarlm()* function from the *spdep* package⁴⁵. For more details about the methodology for this model we kindly refer the reader to⁴⁶. The results of the distance error model for the nodes are shown in Supplementary Table 21.

Eigenvector Filtering Model

This is the second autocorrelation model that we used. For the eigenvector filtering model, parameter U_i is different for each time slot and is given by the following equation:

$$U_i^t = \gamma_1 V_1^t + \gamma_2 V_2^t + \cdots \gamma_K V_K^t$$

where $V_k^t, k = 1, \dots, K$ are the eigenvector filters computed from the matrix \mathbf{W} defined in the previous paragraph, and are used to “remove” the social distance from such linear models; γ_k^t are the respective coefficients. We used the *ME()* function from package *spdep*⁵⁴ to compute the eigenvector filters, and the *lm()* function to apply the filters to the models. We refer the reader to⁵³ for additional details. The results of this model are reported in Supplementary Table 21.

Robust Standard Errors Model

We used an additional model, the Robust SE model, for post correction of the standard errors in the OLS model. The main idea is to use heteroskedacity and autocorrelation-consistent (HAC) estimation of the covariance matrix of the coefficient estimates in such a model⁴⁷. We used the *rlm()* function from the *MASS* package⁴⁸. The results of this model are reported in Supplementary Table 21.

Time-Lagged Model

We also used a lagged model to assess the impact of independent variables calculated in year 2013 ($t_1 = 1$) on the node credit healthiness in year 2014 ($t_2 = 2$). In this case, the model equation changes significantly, i.e., we cannot employ (M3), and we rewrite it as follows:

$$(M4): \log(y_i^{t_2}) = a_i^{t_2} + \beta_{Ck}\log(x_{iCk}^{t_1}) + \beta_{Pk}\log(x_{iPk}^{t_1}) + \beta_w\log(x_{i6}^{t_1}) + \beta_d\log(x_{i7}^{t_1}) + \beta_c q_i + \varepsilon_{it_2}$$

We used the *lm()* R function. The results of the time lagged model are reported in Suppl. Table 16.

3.4 Extended results with control variables

In this section, we discuss the extended versions of the random effects models predicting edge volume and ordinary least square models predicting node credit healthiness that we presented in Figure 4 of the main paper. Please note that the latter are obtained after a proper transformation ($e^x - 1$) of the coefficients that are shown below.

In detail, Supplementary Table 10 presents the extended results of the random effects models of Fig. 4(A) in the main paper, with all the covariates. These models assess the volume of money that has been transacted between two businesses. Since the numbers of cycles of different lengths crossing one edge were highly correlated ($r = .25$ to $.96$; correlation matrices in appendix), and in order to avoid multicollinearity issues, we used independent models for each cycle length. Each model assesses the effect of a different cycle length and controls for the number of paths of same length (used as a comparison point). No problematic signs of multicollinearity were observed in these models. We have used all the key available covariates in our dataset, and we have also calculated the actual geodesic distance among the nodes of each edge. The effect of the distance is not significant ($p < 0.05$).

Supplementary Table 10. Random effect models predicting **edge volume** (13,753 edges; 1,262 buyers 1,317 sellers).

	Length of paths and cycles									
	2		3		4		5		Betweenness	
	b	(95% CI)	b	(95% CI)	b	(95% CI)	b	(95% CI)	b	(95% CI)
Fixed effects										
(Intercept)	5.98	(5.91, 6.06)	5.99	(5.92, 6.07)	6.02	(5.94, 6.09)	6.03	(5.95, 6.10)	5.99	(5.92, 6.06)
2nd Year (ref=1st year)	-0.17	(-0.22, -0.11)	-0.17	(-0.22, -0.11)	-0.17	(-0.22, -0.12)	-0.17	(-0.22, -0.11)	-0.19	(-0.25, -0.14)
Number of paths†	0.00	(-0.02, 0.02)	-0.01	(-0.03, 0.02)	0.00	(-0.02, 0.03)	0.00	(-0.03, 0.03)	-	-
Number of cycles†	0.05	(0.03, 0.07)	0.04	(0.02, 0.07)	0.08	(0.04, 0.11)	0.10	(0.07, 0.14)	-	-
Geographic distance†	0.02	(0.00, 0.05)	0.02	(0.00, 0.05)	0.02	(0.00, 0.04)	0.02	(-0.01, 0.04)	0.01	(-0.01, 0.03)
Betweenness†	-	-	-	-	-	-	-	-	0.09	(0.06, 0.12)
Duration of buyer	0.09	(0.07, 0.12)	0.09	(0.06, 0.12)	0.08	(0.06, 0.11)	0.08	(0.05, 0.11)	0.09	(0.07, 0.12)
Duration of seller	0.09	(0.06, 0.12)	0.08	(0.05, 0.11)	0.08	(0.04, 0.11)	0.07	(0.04, 0.10)	0.09	(0.06, 0.12)
Category of buyer (ref = Retail)										
Wholesale	0.37	(0.21, 0.53)	0.37	(0.22, 0.53)	0.37	(0.22, 0.53)	0.37	(0.21, 0.53)	0.37	(0.21, 0.53)
Construction	0.13	(0.00, 0.26)	0.14	(0.00, 0.27)	0.15	(0.01, 0.28)	0.15	(0.02, 0.28)	0.16	(0.03, 0.30)
Ho. Re. Ca.	0.19	(0.06, 0.33)	0.20	(0.07, 0.33)	0.20	(0.07, 0.34)	0.21	(0.07, 0.34)	0.18	(0.05, 0.32)
Industry	0.04	(-0.08, 0.17)	0.05	(-0.07, 0.17)	0.06	(-0.07, 0.18)	0.05	(-0.07, 0.18)	0.05	(-0.07, 0.18)
Services	0.09	(0.00, 0.18)	0.10	(0.01, 0.19)	0.11	(0.02, 0.20)	0.11	(0.02, 0.20)	0.11	(0.02, 0.20)
Category of seller (ref = Retail)										
Wholesale	0.64	(0.43, 0.85)	0.65	(0.43, 0.86)	0.65	(0.44, 0.86)	0.65	(0.44, 0.86)	0.62	(0.41, 0.82)
Construction	1.14	(0.94, 1.33)	1.14	(0.94, 1.34)	1.14	(0.94, 1.34)	1.14	(0.94, 1.33)	1.11	(0.91, 1.30)
Ho. Re. Ca.	-0.51	(-0.68, -0.35)	-0.51	(-0.67, -0.34)	-0.49	(-0.66, -0.33)	-0.49	(-0.65, -0.32)	-0.51	(-0.67, -0.35)
Industry	0.43	(0.27, 0.59)	0.44	(0.27, 0.60)	0.45	(0.28, 0.61)	0.45	(0.29, 0.62)	0.43	(0.27, 0.59)
Services	0.44	(0.31, 0.56)	0.44	(0.32, 0.57)	0.45	(0.32, 0.57)	0.45	(0.33, 0.58)	0.42	(0.30, 0.55)
Random effects										
Buyer	Var	SD	Var	SD	Var	SD	Var	SD	Var	SD
Buyer	0.18	0.43	0.18	0.42	0.18	0.42	0.18	0.42	0.19	0.43
Seller	0.47	0.68	0.47	0.68	0.47	0.68	0.47	0.68	0.44	0.67
AICc	49985		49995		49986		49974		50855	

† Values are logged

Since we are using dyadic data, in order to respect the assumption of independence of the observations, we need to account for the fact that some nodes appear more than once in the multiple dyads of the network. To consider the dependencies of these observations, a possible method is suggested to include random effects at each of the node level⁴⁹. Unlike a hierarchical structure, in network data, the nodes are interrelated. Thus, random effects need to be crossed-classified. This structure allows us to account for covariates at the node level (i.e., type of business and lifespan), but also at the edge level (i.e., number of paths and cycles). Confidence intervals are estimated using 1,000 bootstrap simulations. Beyond the effect of paths and cycles (Fig. 4), we observe that nodes that were present for a longer period in the network conducted more transactions. The average amount of Sardex transferred between two businesses did not vary from the 1st to 2nd year. We also note that businesses of different categories will have different transaction volume, particularly on the side of the seller.

Supplementary Table 11 presents the extended results of the second class of models, predicting the credit healthiness at the node level with the coefficients for all covariates. Using the nodes that were present in both time periods of the panel data⁵⁰ (i.e., nodes entered the system during the 1st year and being active in the 2nd year), we added a random effect at the node level to account for dependencies related to the network structure (alternative models are presented below). Beyond the effect of the paths and cycles observed in Fig. 4, we present in Supplementary Table 11 the effect of other covariates. We see that (i) nodes with higher weighted degree have higher average absolute balance (worst credit

healthiness), (ii) the duration of node observation for each time period does not influence its credit healthiness, (iii) Businesses involved in wholesale and food industry have worst credit healthiness, and (iv) the average absolute balance was higher in the second year than in the first year.

Supplementary Table 11. OLS models predicting **node credit healthiness** (N=2,230 observations; 1,477 nodes).

	Length of paths and cycles				Betweenness and Degree					
	2		3		4					
	b	(95% CI)	b	(95% CI)	b	(95% CI)				
Fixed effects										
(Intercept)	6.94	(6.83, 7.05)	6.94	(6.83, 7.05)	6.91	(6.80, 7.02)	6.85	(6.74, 6.97)	6.88	(6.77, 6.99)
2nd Year (ref=1st year)	0.24	(0.17, 0.32)	0.24	(0.16, 0.31)	0.24	(0.17, 0.32)	0.30	(0.23, 0.38)	0.25	(0.18, 0.33)
Number of paths†	0.13 (0.10, 0.17)		0.17 (0.13, 0.21)		0.15 (0.11, 0.19)		0.07 (0.03, 0.12)		-	-
Number of cycles†	-0.17 (-0.21, -0.12)		-0.27 (-0.32, -0.22)		-0.38 (-0.44, -0.33)		-0.41 (-0.47, -0.35)		-	-
Degree†	-	-	-	-	-	-	-	-	-0.29	(-0.38, -0.20)
Betweenness†	-	-	-	-	-	-	-	-	-0.09	(-0.14, -0.03)
Weighted degree†	1.24	(1.19, 1.29)	1.29	(1.23, 1.34)	1.36	(1.31, 1.42)	1.41	(1.35, 1.47)	1.41	(1.34, 1.48)
Duration of business	0.03	(-0.11, 0.17)	0.04	(-0.10, 0.18)	0.08	(-0.06, 0.22)	0.10	(-0.04, 0.24)	0.11	(-0.03, 0.26)
Category of business (ref = Retail)										
Wholesale	0.21	(0.03, 0.38)	0.17	(0.00, 0.35)	0.18	(0.01, 0.35)	0.19	(0.02, 0.36)	0.15	(-0.02, 0.33)
Construction	0.12	(-0.04, 0.27)	0.08	(-0.08, 0.23)	0.06	(-0.09, 0.21)	0.06	(-0.09, 0.21)	0.03	(-0.13, 0.18)
Ho. Re. Ca.	0.18	(0.05, 0.32)	0.16	(0.03, 0.29)	0.18	(0.05, 0.31)	0.19	(0.06, 0.32)	0.17	(0.04, 0.31)
Industry	0.00	(-0.13, 0.12)	-0.02	(-0.14, 0.11)	-0.01	(-0.13, 0.12)	0.00	(-0.12, 0.13)	-0.09	(-0.22, 0.03)
Services	0.10	(0.00, 0.20)	0.09	(-0.01, 0.19)	0.08	(-0.01, 0.18)	0.08	(-0.01, 0.18)	0.03	(-0.07, 0.13)
AICc	5869.3		5829.2		5793.3		5786.4		5845.1	

† Values are logged

3.5 Increase in the influence of cycle by length

Previously, we used independent models to observe the effect of cycles for each length. The results of these models can be also employed to indirectly compare the impact of the cycles of different length, i.e., to assess whether there is an increase of the effect. In particular, inspecting the coefficients (*b*) and the model fit (AICc) for the cycles of the different length on the two performance metrics, i.e., the edge volume and the node credit healthiness (Supplementary Tables 10 and 11), we can see a general trend for both metrics that the lengthier cycles have a larger impact (i.e., the absolute value of the respective coefficient (*b*) is larger) and the model fit is better (i.e., lower AICc values). This difference of the cycle impact is more evident among cycles of length 5 (or 4) than cycles of length 2 (or 3).

To further investigate this issue, we used stepwise statistical models, adding the covariates (i.e., the number of cycles of different length) one after the other, and compared the model fits. Supplementary Table 12 presents the estimates at each step. Since the covariates are highly collinear, we cannot easily interpret their coefficients. However, the overall model fit is interpretable. We observe, for both the node and the edge analysis, that the total model fit (AICc) increases from model 1 (which includes only cycles involving two edges) to model 4 (which includes cycles involving 2 to 5 edges). At each step, the number of larger cycles crossing one edge increases the fit of the model. This underlines an increase of

quality of explanation achieved through the addition of the larger cycles. In other words, cycles of different length carry different information regarding the economic activity in the system.

Supplementary Table 12: Stepwise models estimating the simultaneous effect of all the cycles

Node credit healthiness (n=2,230 observations; 1,477 nodes)								
	Model 1		Model 2		Model 3		Model 4	
	b	(95% CI)	b	(95% CI)	b	(95% CI)	b	(95% CI)
Number of cycles								
k=2	-0.15	(-0.20, -0.11)	-0.07	(-0.14, -0.01)	-0.06	(-0.12, 0.01)	-0.07	(-0.13, -0.01)
k=3	-	-	-0.18	(-0.26, -0.11)	0.05	(-0.06, 0.16)	-0.03	(-0.15, 0.09)
k=4	-	-	-	-	-0.33	(-0.44, -0.22)	-0.03	(-0.19, 0.27)
k=5	-	-	-	-	-	-	-0.29	(-0.46, -0.12)
AICc	5834	>	5811	>	5794	>	5786	
Model comparison vs [Chi ² (df)]								
Model 2	37.3(2); p<0.001		-		-		-	
Model 3	67.5(4); p<0.001		30.2(2); p<0.001		-		-	
Model 4	87.0(6); p<0.001		49.7(4); p<0.001		19.5(2); p<0.001		-	

Edge volume (n=13,753 edges; 1,262 buyers; 1,317 sellers)								
	Model 1		Model 2		Model 3		Model 4	
	b	(95% CI)	b	(95% CI)	b	(95% CI)	b	(95% CI)
Number of cycles								
k=2	0.05	(0.03, 0.07)	0.04	(0.02, 0.06)	0.04	(0.02, 0.06)	0.04	(0.02, 0.06)
k=3			0.03	(0.00, 0.06)	0.00	(-0.03, 0.03)	0.00	(-0.03, 0.03)
k=4					0.08	(0.04, 0.12)	-0.04	(-0.11, 0.02)
k=5							0.16	(0.10, 0.24)
AICc	51845	=	51844	>	51835	>	51817	
Model comparison vs [Chi ² (df)]								
Model 2	5.2(2); p=0.076		-		-		-	
Model 3	18.6(4); p<0.001		13.4(2); p=0.001		-		-	
Model 4	40.1(6); p<0.001		35.0(4); p<0.001		21.5(2); p<0.001			

3.6 The general effect of Cycles

Another way to compare the effects of different cycles and paths without facing the issue of multicollinearity, is to extract the common variance of cycles of different length and the common variance of paths of different length as two independent factors. Using principal component analysis with eigenvalue decomposition, we extracted two factors from cycles and paths covariance. This analysis doesn't allow us to compare cycles of different size in the same model, but is a compromise for comparing the general effect of all cycles of different sizes, all paths of different size, including degree and betweenness in the same model, without major multicollinearity issues. Supplementary Table 13 presents models predicting node credit healthiness and edge volume including cycles and paths factors, and degree and betweenness. We observe that, similarly to our main models, cycles increase node healthiness, while paths decrease it. This effect is stronger than betweenness, but similar to that of degree. For the model predicting edge volume, cycles increase volume, while paths have no effect. The effect of cycles and betweenness are comparable. The fit of these models is the best among all presented models, showing the importance of both cycles and other network features in this economic system.

Supplementary Table 13: Models predicting node credit healthiness and edge volume including cycles and paths factors, and degree and betweenness (n=2,230 observations; 1,477 nodes)

	Node credit healthiness		Edge volume	
	b	(95% CI)	b	(95% CI)
Cycles factor	-0.18	(-0.26, -0.09)	0.10	(0.07, 0.14)
Paths factor	0.17	(0.12, 0.21)	-0.01	(-0.03, 0.02)
Betweenness	-0.03	(-0.08, 0.03)	0.09	(0.06, 0.12)
Degree	-0.27	(-0.39, -0.16)	-	-
AICc	5772.5		49943	

3.7 Effects of betweenness and degree in different models

In the main models, we included both degree and betweenness. Supplementary Table 14 presents alternative models as sensitivity test for the independent influence of these measures, including all others control variables. We observe that the effect size of betweenness of model 3 ($b = -0.20$) is higher than the model 1 including both degree and betweenness ($b = -0.09$), but still lower than most of the cycle effects (except 2-cycle; Supplementary Table 11). Moreover, the fit of model 3 is worse ($AICc = 5885.5$) than model 1 which includes both degree and betweenness ($AICc = 5845.1$). Note that model 2, excluding betweenness and considering only degree, results also in a worse fit ($AICc = 5852.0$).

Supplementary Table 14: Models predicting node credit healthiness including degree and betweenness independently (n=2,230 observations; 1,477 nodes)

	Model 1: With both betweenness and degree		Model 2: With degree only		Model 3: With betweenness only	
	b	(95% CI)	b	(95% CI)	b	(95% CI)
Degree	-0.29	(-0.38, -0.20)	-0.37	(-0.44, -0.30)	-	-
Betweenness	-0.09	(-0.14, -0.03)	-	-	-0.20	(-0.24, -0.15)
AICc	5845.1		5852.0		5885.5	

3.8 Stability of the effect through the observation periods and lagged models

As we mentioned earlier, the network was split in two different observation periods. Supplementary Table 15 shows the results from the models in Supplementary Tables 10 and 11, presenting each year independently to ensure that these results are mostly comparable (note that in our previous analysis we used both populations concurrently). We observe that, both for the edges and the nodes, the coefficients' direction and their respective values are qualitatively identical (considering the confidence intervals) in the two observation periods. The effect of cycles and paths then seems constant for these two years.

Supplementary Table 15. Comparison of Results Among the Two Years

	Node Credit Healthiness				Edge volume			
	1st year (n = 877)		2nd Year (n = 1353)		1st year (n = 5962)		2nd Year (n = 9916)	
	b	(95% CI)	b	(95% CI)	b	(95% CI)	b	(95% CI)
Cycle length								
k=2	-0.19	(-0.25, -0.13)	-0.07	(-0.21, 0.07)	0.07	(0.03, 0.10)	0.04	(0.02, 0.07)
k=3	-0.31	(-0.38, -0.25)	-0.26	(-0.36, -0.16)	0.07	(0.03, 0.11)	0.03	(-0.00, 0.07)
k=4	-0.44	(-0.52, -0.37)	-0.39	(-0.48, -0.31)	0.13	(0.08, 0.19)	0.07	(0.02, 0.11)
k=5	-0.49	(-0.56, -0.41)	-0.44	(-0.52, -0.35)	0.12	(0.06, 0.18)	0.11	(0.06, 0.15)
Paths length								
k=2	0.12	(0.08, 0.17)	0.06	(-0.05, 0.17)	0.01	(-0.02, 0.05)	-0.01	(-0.04, 0.01)
k=3	0.19	(0.14, 0.24)	0.19	(0.11, 0.27)	0.00	(-0.04, 0.04)	0.00	(-0.03, 0.03)
k=4	0.18	(0.13, 0.23)	0.14	(0.08, 0.20)	-0.01	(0.06, 0.03)	0.05	(0.00, 0.08)
k=5	0.08	(0.02, 0.14)	0.10	(0.05, 0.16)	0.02	(-0.04, 0.08)	0.02	(-0.02, 0.06)

In the main models (Tables 10 and 11 above), the dependent and the independent variables are observed during the same time period (Year 1 vs Year 1; Year 2 vs Year 2). Although we do not assume causality in any of our models, we add a one-year lag in order to increase the independence of our measures. The results of the independent lagged models for each cycle length are shown in Supplementary Table 16. The independent variables (i.e., cycles, paths, degree, betweenness) are measured at Year 1 and predict the dependent variables (i.e., node credit healthiness and edge volume) at Year 2, for those nodes that were part of both networks (i.e., n=877). We observe that there are very few differences among the coefficients in the main models and the lagged models.

Supplementary Table 16. Models with Lagged Cycles and Paths

	Node Credit Healthiness (Year 2014, n = 1353)		Edge volume (Year 2014, n = 9916)	
	b	(95% CI)	b	(95% CI)
Cycles length (Year 2013)				
k=2	-0.03	(-0.14, 0.07)	0.09	(0.03, 0.15)
k=3	-0.11	(-0.23, 0.01)	0.08	(0.01, 0.16)
k=4	-0.34	(-0.48, -0.21)	0.14	(0.05, 0.23)
k=5	-0.48	(-0.63, -0.33)	0.12	(0.03, 0.24)
Paths length (Year 2013)				
k=2	0.21	(0.12, 0.30)	0.00	(-0.06, 0.07)
k=3	0.26	(0.17, 0.35)	-0.02	(-0.09, 0.05)
k=4	0.28	(0.18, 0.37)	0.08	(-0.01, 0.16)
k=5	0.19	(0.08, 0.30)	0.04	(-0.06, 0.14)

3.9 Cycle sequentiality and overlap

In this subsection, we evaluate the robustness of our models by (i) using an alternative definition of cycles which takes into account only the sequential cycles, and (ii) controlling for node k-cycle coverage in addition to k-cycle centrality.

In particular, as it was explained above (Supplementary Figure 4), there are cycles in which the links are created in increasing time sequence (sequential cycles) and others that do not conform to this condition (non-sequential cycles). In our models, so far, we employed both classes of cycles using a broader definition (which does not consider the time sequence of edges). Therefore, a question that arises naturally is whether sequential cycles have the same impact on the two performance metrics (edge

weight and node credit healthiness). To address this issue, we re-run the models using only sequential cycles. Supplementary Table 17 shows the estimates of these models (independent models for each cycle length). The observed effects are qualitatively identical but smaller compared to the effects of the main models (presented in the paper), and possibly less precise. This could be due, in part, to the smaller population of sequential cycles (compared to the population of non-sequential cycles) and this decreases our explanatory power. On the other hand, this might be also due to the fact that cycles do not need to be sequential to be efficient. Further analysis of these dynamic processes could increase our understanding of the effect of monetary cycles, but this is beyond the scope of this paper.

Supplementary Table 17 : Model with sequential cycles

	Node Credit Healthiness (n=2,230 observations; 1,477 nodes)		Edge Volume (n=13,753 edges; 1,262 buyers; 1,317 sellers)	
	b	(95% CI)	b	(95% CI)
Sequential cycle length				
k=2	-0.16	(-0.21, -0.11)	0.05	(0.03, 0.07)
k=3	-0.22	(-0.27, -0.17)	0.03	(0.00, 0.05)
k=4	-0.29	(-0.35, -0.23)	0.06	(0.04, 0.09)
k=5	-0.29	(-0.35, -0.23)	0.10	(0.07, 0.13)

Regarding cycle coverage, it is important to emphasize that - by definition - different cycles can be partially overlapping. Namely, as it was explained in the main paper, two nodes with the same k-cycle centrality might reach with these k-cycles different number of nodes, depending on the extent to which their cycles are overlapping. In order to account for this effect, and differentiate among such nodes, we repeated our main model analysis by using for each node (or edge), instead of the k-cycle centrality, the number of different nodes it reaches by its k-cycles. As we observe from Supplementary Table 18, the results are qualitatively identical with the results when we employ only the k-cycle centrality.

Supplementary Table 18: Model with k-cycle coverage

	Node Credit Healthiness (N=2,230 observations; 1,477 nodes)		Edge Volume (N=13,753 edges; 1,262 buyers; 1,317 sellers)	
	b	(95% CI)	b	(95% CI)
Cycle coverage length				
k=2	-0.15	(-0.20, -0.10)	0.05	(0.03, 0.07)
k=3	-0.25	(-0.30, -0.20)	0.04	(0.01, 0.06)
k=4	-0.35	(-0.41, -0.29)	0.08	(0.05, 0.11)
k=5	-0.37	(-0.43, -0.31)	0.10	(0.07, 0.14)

3.10 Alternative measure of node credit healthiness

The measure of credit healthiness $H(n)$ removes the effect of time by dividing the aggregate absolute daily balance by the number of days that the business was in the network. One could consider that the longer an entity had a bad balance (e.g., a very high surplus), the worst the credit healthiness should be. Supplementary Table 19 shows the regression coefficients when we employ this alternative credit healthiness metric $H_B(n)$, i.e., we do not divide the balance by time. We observe that the estimates are very close to the models that employ the first (previous) credit healthiness metric. This result is not surprising, as we were controlling in those models for the lifespan of the business in the network (hence already taking into account the time dimension).

Supplementary Table 19: Models with alternative measure of node credit healthiness

Alternative Node Credit Healthiness (N=2,230 observations; 1,477 nodes)		
b	(95% CI)	
Cycle length		
k=2	-0.19	(-0.25, -0.13)
k=3	-0.28	(-0.35, -0.21)
k=4	-0.35	(-0.42, -0.28)
k=5	-0.34	(-0.42, -0.27)

3.11 Alternative models to account for dependencies

In order to further verify the robustness of the above results, we have fit a suite of alternative models that account for dependencies between the observations. Indeed, in our complex dataset there are several multicollinearity issues and we took a multitude of approaches to ensure these issues do not affect our analysis and findings. First, an important collinearity effect arises between the cycles of different length. We addressed this issue by employing a different model for each different type (i.e., length) of cycles, and then we compared their effect indirectly, i.e., through the respective coefficients and model fit metrics (AICc values). Second, within each of these models there is still a certain level of collinearity between the different covariates. However, this effect is at an acceptable level. Namely, we did make sure that there were no problematic signs of multicollinearity in the presented models according to the Variance Inflation Factor (VIF) metric. For the models at the edges level (Supplementary Table 10), all VIF were smaller than 1.5. For the models at the nodes level (Supplementary Table 11), the highest VIF (5.21) was observed for the model including the degree measure. The other covariates showed even less signs of multicollinearity (VIF < 3). Even though the interpretation of this scale is subjective, the standard practice is to consider VIF values smaller than 10 as non-problematic⁵¹.

Supplementary Table 20 shows the results from an alternative modeling technique to control for the node dependencies when predicting edge volume. Supplementary Figure 14(B) shows the distribution of predicted vs actual values for these models. In particular, for this model we used fixed effect instead of random effect at the node level. Despite being stricter, this model shows that the results remain consistent with the random effect model presented in the main paper. In particular, the fixed effect model fit is comparable to the fit of the random effect model, and the respective coefficients are almost identical (considering the CI values). Since random effects models present a better fit (AICc smaller), we favor this model in the main paper.

Supplementary Table 20. Alternative model for **edge volume** with fixed affects to account for edges dependencies (n=13,753 edges; 1,262 buyers; 1,317 sellers)

	Random effect				Fixed effect			
	b	95% CI	SE/ b	AICc	b	95% CI	SE/ b	AICc
Cycles								
k=2	0.05	(0.03, 0.07)	0.22	51942	0.04	(0.02, 0.06)	0.28	52576
k=3	0.04	(0.01, 0.06)	0.34	51954	0.02	(-0.01, 0.05)	0.79	52589
k=4	0.08	(0.05, 0.12)	0.20	51938	0.07	(0.02, 0.11)	0.33	52580
k=5	0.11	(0.08, 0.15)	0.16	51922	0.11	(0.06, 0.16)	0.24	52571
Paths								
k=2	-0.01	(-0.03, 0.01)	1.30	51942	0.00	(-0.03, 0.02)	7.15	52576
k=3	-0.01	(-0.03, 0.01)	1.07	51954	-0.01	(-0.03, 0.02)	1.98	52589
k=4	0.00	(-0.02, 0.03)	6.37	51938	-0.01	(-0.04, 0.02)	1.56	52580
k=5	0.01	(-0.02, 0.04)	2.15	51922	0.00	(-0.04, 0.04)	8.34	52571

We adopted OLS models as a main model to predict node healthiness as it has the best fit for the data. Supplementary Table 21 shows the results from alternative modeling techniques that control for the node dependencies when predicting the node credit healthiness. For each model we present the coefficients, the coefficient of variation (defined as SE/|b|, i.e., a relative measure of deviation) and the AICc. Supplementary Figure 14(A) shows the distribution of predicted vs actual values for the models.

The first model of the table shows the Ordinary Least Square (OLS) regression model as a comparison point. This model did not account for dependencies between the observations. The second model shows the results when we correct for standard errors according to the correlation between individuals and using heteroskedastic and autocorrelation robust standard error estimation⁵². This model is not optimal, and is less strict since we do not specify the dependency structure (i.e., the network) between the observations. The fixed effect model would be a more restrictive model, since, by adding the fixed effect for 877 businesses, we end up with very small degree of freedom. However, it exhibits very large standard errors and essentially it does not fit well the data (the respective AICc values are 11447, 11442, 11449, 11445, almost twice as high as the other models showed). Therefore, we do not include it as a comparison model.

The last two models in Supplementary Table 21 are autoregressive and capture the dependency among the nodes using the network structure. That is, these models use a distance matrix to weight the relation between each observation. The first one uses a social distance matrix based on the weighted shortest path between each pair of nodes in order to correct standard errors⁵³. The second is using eigenvector filtering to identify vectors of autocorrelation in the observations, minimizing Moran's I⁵⁴. Considering that there was no substantial difference in the effect size (b) between each of these strategies, we opted to present in the main paper the simplest OLS model. For the above analyses we used the R software packages “lme4”⁴⁴, “sandwich”⁴⁷, and “spdep”⁴⁵.

Supplementary Table 21: Alternative models for **node credit healthiness** to account for node dependencies
(n=2,230 observations; 1,477 nodes)

	OLS				Robust SE				Random effect			
	b	95% CI	SE/ b	AICc	b	95% CI	SE/ b	AICc	b	95% CI	SE/ b	AICc
Cycles												
k=2	-0.17	(-0.21, -0.12)	0.14	5869	-0.20	(-0.24, -0.15)	0.11	5893	-0.15	(-0.20, -0.10)	0.16	5861
k=3	-0.27	(-0.32, -0.22)	0.10	5829	-0.31	(-0.36, -0.27)	0.08	5855	-0.25	(-0.30, -0.19)	0.11	5831
k=4	-0.38	(-0.44, -0.32)	0.08	5793	-0.44	(-0.49, -0.39)	0.06	5823	-0.35	(-0.41, -0.29)	0.09	5802
k=5	-0.41	(-0.47, -0.35)	0.07	5786	-0.49	(-0.54, -0.43)	0.05	5818	-0.38	(-0.43, -0.32)	0.08	5801
Paths												
k=2	0.13	(0.10, 0.17)	0.14	5869	0.14	(0.11, 0.18)	0.12	5893	0.12	(0.08, 0.16)	0.16	5861
k=3	0.17	(0.13, 0.21)	0.12	5829	0.18	(0.15, 0.22)	0.10	5855	0.15	(0.11, 0.19)	0.13	5831
k=4	0.15	(0.11, 0.19)	0.14	5793	0.16	(0.12, 0.19)	0.11	5823	0.14	(0.10, -0.17)	0.15	5802
k=5	0.07	(0.03, 0.12)	0.30	5786	0.08	(0.04, 0.12)	0.24	5818	0.07	(0.02, 0.11)	0.33	5801
Distance error				Eigenvector filtering								
	b	95% CI	SE/ b	AICc	b	95%CI	SE/ b	AICc				
Cycles												
k=2	-0.14	(-0.18, -0.09)	0.18	5872	-0.12	(-0.17, -0.07)	0.21	5832				
k=3	-0.23	(-0.29, -0.17)	0.12	5839	-0.22	(-0.27, -0.16)	0.13	5801				
k=4	-0.37	(-0.43, -0.31)	0.08	5809	-0.35	(-0.41, -0.29)	0.09	5778				
k=5	-0.41	(-0.47, -0.35)	0.08	5803	-0.39	(-0.45, -0.32)	0.08	5782				
Paths												
k=2	0.14	(0.10, 0.18)	0.14	5872	0.14	(0.11, 0.18)	0.13	5832				
k=3	0.18	(0.14, 0.22)	0.11	5839	0.17	(0.14, 0.21)	0.11	5801				
k=4	0.15	(0.11, 0.19)	0.14	5809	0.16	(0.12, 0.20)	0.13	5778				
k=5	0.08	(0.03, 0.12)	0.28	5803	0.09	(0.05, 0.14)	0.23	5782				

Finally, we have also used (both for nodes and edges) a set of non-parametric non-linear models as sensitivity test. First, we fit random forest (RF) regression models⁵⁵ in our data using the R software package “randomForest”⁵⁶. For the edge model the RF model explains 17% of the variance, which is higher than the respective linear models (approximately 13%), and for the node models the RF model explains 63% of the variance, which is similar to its respective linear model (64%). Also, we have used a Support Vector Machine (SVM) as an alternative non-linear non-parametric model⁵⁷ using R software package “e1071”⁵⁸. The SVM explains 70% of the variance of healthiness (for the node model), and 25% of the variance of the transaction volume (for the edge model) which are slightly higher than their respective linear models.

As expected, many of these models are more predictive than the linear models. However, the downside is that it becomes really hard to parse out and compare the effect of each covariate in terms of size and/or direction. Moreover, each of these models uses a different metric of influence. Supplementary Table 22 presents the ranking of the t-value for the linear model, the “% increase” in mean squared error for the RF model, and the weight for the SVM model.

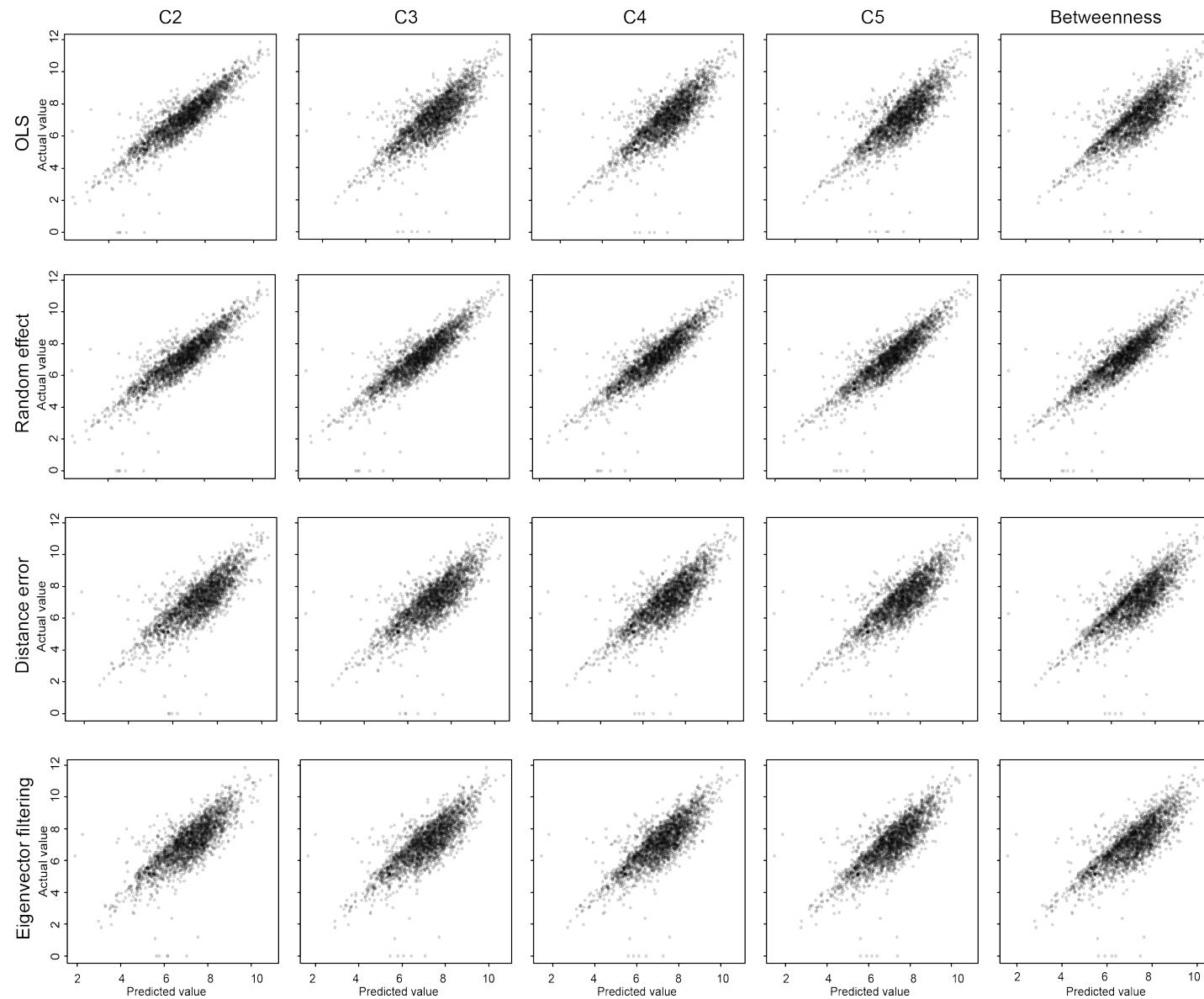
Supplementary Table 22: Ranks of covariates' importance for non-linear non-parametric models predicting node credit healthiness and edge volume in comparison to linear models

Node models (n=2,230 observations; 1,477 nodes)					Edge models (n=13,753 edges; 1,262 buyers; 1,317 sellers)									
	Support vector machine		Random forest models		OLS			Support vector machine		Random forest models		OLS		
	Weight	Rank	%IncMSE	Rank	t-value	Rank		Weight	Rank	%IncMSE	Rank	t-value	Rank	
Betweenness	14.4	10	13.3	10	3.2	11	Betweenness	455.2	1	407.8	1	31.7	1	
Degree	26.2	3	14.4	8	5.7	10	C2	2.7	19	22.3	16	4.7	11	
Weighted degree	105.1	1	89.6	1	45.6	1	P2	8.5	13	1.4	23	0.2	21	
C2	19.3	4	14.0	9	6.1	9	C3	0.2	23	24.8	15	3.3	14	
P2	6.5	13	9.5	12	6.3	8	P3	14.3	5	30.8	12	0.6	20	
C3	16.8	6	12.3	11	8.8	4	C4	0.3	22	34.7	10	4.4	13	
P3	15.5	8	20.4	3	7.7	6	P4	2.6	20	38.9	9	0.2	22	
C4	18.0	5	15.0	7	11.3	3	C5	9.4	11	46.1	7	5.5	9	
P4	14.8	9	15.6	6	6.7	7	P5	1.9	21	58.3	3	0.1	23	
C5	16.7	7	17.4	5	12.0	2	Distance	33.4	3	55.5	4	2.0	16	
P5	11.7	11	17.9	4	3.2	12	BuildingB	6.7	15	16.8	19	1.9	18	
Building	0.7	18	4.9	13	1.6	16	BuildingS	35.6	2	61.8	2	11.3	2	
Food	3.3	17	2.5	18	2.3	14	FoodB	11.0	10	18.2	18	2.9	15	
Industry	3.8	16	3.8	16	0.2	17	FoodS	25.8	4	54.7	5	6.1	6	
Service	5.6	14	3.6	17	2.1	15	IndustryB	7.3	14	13.4	21	0.7	19	
Wholesale	5.4	15	4.9	14	2.5	13	IndustryS	2.9	18	30.1	14	5.1	10	
Time	26.9	2	28.8	2	0.2	18	ServiceB	11.7	9	22.2	17	1.9	17	
Year	9.7	12	4.1	15	8.0	5	ServiceS	12.0	8	44.0	8	6.8	3	
							WholesaleB	13.7	7	16.3	20	4.6	12	
							WholesaleB	6.1	17	30.5	13	6.0	7	
							TimeB	13.9	6	32.1	11	6.8	4	
							TimeS	9.2	12	51.4	6	5.6	8	
							Year	6.6	16	8.3	22	6.1	5	

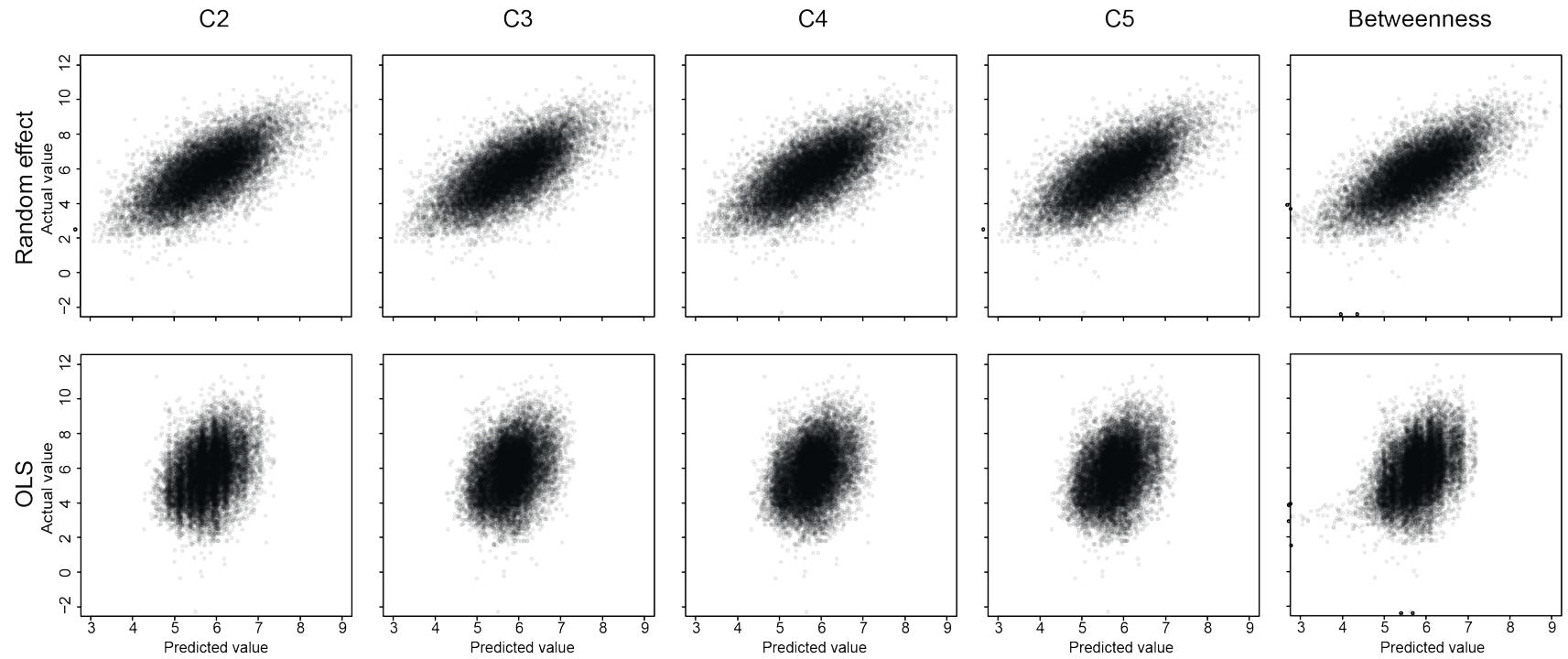
A representative example demonstrating the difficulties in interpreting the results of these models is the following. Using the linear model for credit healthiness, we found that cycles of size 3 (C3) have the 4th rank in terms of effect, while paths of the same size (P3) had the 6th rank (same ordering as the linear models). The respective SVM model gives the same qualitatively result, namely C3 has the 6th rank in terms of effect while P3 has the 8th rank. Inconsistently however, in the RF model, C3 were ranked 11th in terms of influence while P3 were ranked 3rd. More importantly, however, in the non-parametric non-linear models, the weight or influence metrics do not inform us on the direction of this influence. This is very important in that particular case, as paths and cycles have an opposite effect on node credit healthiness. If we want to recommend interventions on the network, we would want to know if we need to increase or decrease cycles and paths in order to improve our system. Therefore, these non-parametric models offer little information for interpretation and for designing a potential intervention: should we increase or decrease cycles in this closed economy?

Finally, let us clarify that the reason we have used a linear model is that this allows us to test our specific hypothesis. We agree that non-linear non-parametric models could be of interest here and would have a better predictive efficiency (indeed this is the case), but, as explained above, these models do not allow as easily the interpretation for the evaluation of hypotheses. These models favour a more inductive process (What is influencing healthiness), while we favoured a deductive reasoning (Are loops important for healthiness?).

Supplementary Figure 14(A): Distribution of predicted values vs actual values for each credit healthiness model



Supplementary Figure 14(B): Distribution of predicted values vs actual values for each model predicting edge volume



Supplementary Note 4

4.1 Limitations of the Statistical Analyses

The above suite of models reveals the high performance of the economic transactions that are involved in cycles, and the high credit robustness of the nodes within cycles. **However, this evidence albeit compelling is not sufficient to support a formal causality claim about this relation (i.e., effects of cycles on economic performance) because, from a statistical point of view, this cannot be strictly proved.** Namely, a proper randomized controlled trial is unfeasible in this setting. Thus, for a causal analysis we would have to rely on either observational studies, or natural experiments leveraging instruments. With either of these methodological approaches, valid causal inference could be pursued at two levels of granularity: (a) whole economies as the units of analysis, in which case we could compare many closed economic systems (at similar stage of development) directly to see if those with less cycles present lower economic performance; or (b) businesses within a single economy as the units of analysis, in which case we need to account for treatment interference (i.e., the fact the treating unit i may affect the outcome of units other than i) among the businesses.

Unfortunately, we only have access to one economy, thus analysis (a) is not feasible for lack of access. Analysis (b) is also unfeasible, for lack of methodology. Methodology for estimating causal effects in observational studies and natural experiments would boil down to a valid matching procedure, for instance, by using propensity score models in the presence of treatment interference and in a dynamic network context. While there has been some progress on matching methods causal analyses in the presence of treatment interference due to a network⁵⁹, and, separately, on instrumental variable approaches in the context of dynamic networks^{60, 61}, there are no provably valid methods at present that can help us carry out analysis (b) to convincingly establish causality. These are the reasons why we limited our claims to be non-causal and descriptive and rely on a suite of statistical models.

Supplementary Table 23

Correlation matrix at the nodes level (n=1,477)		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cycles																					
1.	Length 2	-	.83	.78	.71	.02	.28	.41	.44	.75	.62	.59	.03	.36	.17	.04	.03	-.03	-.05	-.07	.39
2.	Length 3		-	.92	.84	.00	.30	.48	.53	.83	.65	.65	.03	.41	.18	.05	.02	-.04	-.05	-.07	.43
3.	Length 4			-	.96	-.10	.17	.41	.52	.89	.70	.73	.03	.46	.15	.06	.02	-.04	-.01	-.06	.48
4.	Length 5				-	-.18	.05	.28	.45	.89	.70	.76	.03	.49	.13	.06	.02	-.03	.01	-.05	.49
Paths																					
5.	Length 2					-	.66	.52	.38	.01	-.12	-.04	.02	-.10	-.01	.02	.01	-.03	-.02	-.02	.05
6.	Length 3						-	.85	.66	.22	.08	.10	.05	.04	.05	.04	.03	-.02	-.06	-.05	.16
7.	Length 4							-	.84	.40	.24	.26	.07	.14	.06	.05	.04	-.03	-.05	-.06	.24
8.	Length 5								-	.52	.31	.39	-.22	.22	.05	.05	.03	-.03	-.03	-.04	.28
9.	Degree centrality									-	.75	.83	.02	.53	.18	.07	.02	-.02	-.05	-.07	.57
10.	Betweenness										-	.55	.00	.37	.21	.04	-.02	.02	-.05	-.10	.33
11.	Weighted degree											-	-.01	.51	.07	.11	.08	-.09	.02	-.03	.78
12.	2nd Year												-	.03	.00	.00	.00	.00	.00	.00	.07
13.	Duration of the business													-	.02	.05	.01	.01	.01	.03	.39
Category																					
14.	Retail													-	-.13	-.16	-.19	-.21	-.39	.00	
15.	Wholesale														-	-.07	-.09	-.10	-.18	.11	
16.	Construction															-	-.11	-.12	-.21	.07	
17.	Ho. Re. Ca.																-	-.14	-.27	-.05	
18.	Industry																	-	-.29	.00	
19.	Services																		-	-.01	
20.	Healthiness																			-	

Correlation matrix at the edges level (n=13,573)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Cycles																									
1. Length 2	-	.32	.29	.25	-.08	-.01	.10	.12	.14	.02	.05	.08	.01	-.02	.01	-.03	-.05	.04	.01	.00	.06	-.06	.00	.01	.08
2. Length 3	-	.76	.66	-.12	.03	.30	.41	.24	.03	.16	.17	.01	-.04	.01	-.06	-.09	.15	.10	-.01	.05	-.08	-.04	-.02	.07	
3. Length 4	-	.93	-.21	-.07	.30	.51	.30	.05	.22	.17	.03	-.05	.02	-.04	-.11	.22	.11	.02	.04	-.09	-.04	-.02	.09		
4. Length 5	-	-.30	-.21	.16	.42	.31	.07	.23	.14	.04	-.04	.02	-.02	-.10	.24	.11	.02	.04	-.07	-.03	-.02	.10			
Paths																									
5. Length 2		-	.50	.32	.20	-.01	.01	-.07	.01	-.02	-.01	-.01	-.02	.01	-.09	-.04	.01	-.01	-.01	.02	-.01	-.01			
6. Length 3			-	.70	.47	.12	.02	-.01	.03	-.01	-.02	.04	-.05	-.01	-.04	-.03	.03	.00	-.03	-.01	-.01	.02			
7. Length 4				-	.72	.27	.06	.09	.08	.01	-.03	.04	-.06	-.05	.05	.01	.05	.00	-.09	-.02	-.01	.07			
8. Length 5					-	.23	-.30	.23	.08	.02	-.06	.04	-.06	-.03	.24	.02	.05	.01	-.06	-.04	.00	.08			
9. Betweenness						-	.14	.03	.12	.07	-.09	.04	-.01	-.09	-.02	-.02	.07	.06	-.11	.06	.07	.21			
10. 2nd Year							-	-.24	.04	-.02	.02	-.02	.03	-.05	-.28	.04	-.02	.02	-.02	.01	-.04	-.02			
11. Duration of the seller								-	-.01	.08	-.05	.01	.01	.00	.17	-.03	.00	.00	.01	-.03	.05	.09			
Category of the seller									-																
12. Retail										-	-.18	-.22	-.19	-.22	-.44	-.04	.01	.00	-.01	-.04	.08	-.03	-.02		
13. Wholesale											-	-.10	-.08	-.10	-.20	.01	-.06	.00	.00	-.01	.04	.05	.07		
14. Construction												-	-.10	-.12	-.24	-.02	.00	-.02	.10	.01	-.04	-.02	.02		
15. Ho. Re. Ca.													-	-.10	-.21	.03	.01	.04	.00	-.05	.05	-.02	.05		
16. Industry														-	-.24	.00	.00	.03	-.01	-.03	.03	-.01	-.01		
17. Services															-	.03	.01	-.02	-.04	.08	-.12	.04	-.06		
18. Duration of the buyer															-	-.02	.07	-.05	.04	-.02	.00	.06			
Category of the buyer																-	-.22	-.19	-.29	-.23	-.45	-.11			
19. Retail																	-	-.22	-.19	-.29	-.23	-.45	-.11		
20. Wholesale																		-	-.07	-.11	-.09	-.17	.09		
21. Construction																			-	-.10	-.08	-.15	.16		
22. Ho. Re. Ca.																				-	-.12	-.23	-.19		
23. Industry																					-	-.18	.09		
24. Services																						-	.07		
25. Volume																								-	

Supplementary References

38. Web Portal, The Sharing Economy in Amsterdam, URL:
<https://www.iamsterdam.com/en/business/news-and-insights/sharing-economy>
39. P2P Foundation, Collaborative Consumption Wiki,
http://wiki.p2pfoundation.net/Collaborative_Consumption
40. Bolland J. M. Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, 10: 233-253, (1988). doi:10.1016/0378-8733(88)90014-7
41. Valente T. W., Coronges K., Lakon C., Costenbader E. How correlated are network centrality measures? *Connections (Toronto, Ont)*, 28:16-26, (2008).
42. Bianconi G., Capocci A. Number of Loops of Size h in Growing Scale-free Networks, *Phys. Rev. Lett.* 90, 078701, (2003) doi: 10.1103/PhysRevLett.90.078701
43. Csardi G, Nepusz T, Airolidi EM, Statistical Network Analysis with iGraph, Springer, 2018, Forthcoming.
44. Bates D., Maechler M., Bolker B., Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1-48, (2015). doi:10.18637/jss.v067.i01
45. Bivand R., Piras G. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63:1-36, (2015). Url <http://www.jstatsoft.org/v63/i18/>
46. Ord J. K. Estimation methods for models of spatial interaction. *Journal of American Statistical Association*, 70:120-126, (1975).
47. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11:1-17, (2004). Url: <http://www.jstatsoft.org/v11/i10/>.
48. Venables W. N., Ripley B. D. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0, (2002).
49. Duijn M.A. J. V., Snijders T. A. B., Zijlstra B. J. H. P2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58:234–254, (2004). doi:10.1046/j.0039-0402.2003.00258.x
50. Laird N. M., Ware J. H. Random-effects models for longitudinal data. *Biometrics*, Wiley, International Biometric Society, 38(4): 963–74, (1982). doi:10.2307/2529876
51. Kutner M. H., Nachtsheim C. J., Neter J. Applied Linear Regression Models, McGraw-Hill Irwin, ISBN 0-07-238688-6, (2004).

52. Andrews D. W. K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, (1991).
53. Leenders R. T. A. J. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24:21-47, (2002). doi:10.1016/S0378-8733(01)00049-1
54. Griffith D. A., Peres-Neto P. R. Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology*, 87(10):2603–2613, (2006). doi:10.1890/0012 9658(2006)87[2603:SMIETF]2.0.CO;2
55. Breiman, L. Random Forests, *Machine Learning* 45 (1), 5-32, (2001).
56. Liaw A., Wiener M. Classification and Regression by Random Forest, *R News*, 2 (2), pp. 18-22, (2002).
57. Cortes C., Vapnik V. Support-vector Network. *Machine Learning*, 20, 1-25, (1995).
58. Meyer D. Support Vector Machines. The Interface to libsvm in package e1071. Online Documentation of the package e1071 for R (2001). Retrieved from <https://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>
59. Forastiere L., Mealli F., VanderWeel T. J. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification. *Journal of American Statistical Association*, 111: 510-525 (2016). doi:10.1080/01621459.2015.1125788
60. O’Malley A. J., Elwert F., Rosenquist J. N., Zaslavsky A. M., Christakis N. A. Estimating Peer Effects in Longitudinal Dyadic Data Using Instrumental Variables. *Biometrics* 70: 506-515, (2014). doi: 10.1111/biom.12172
61. Phan T. Q., Airolid E. M. A Natural Experiment of Social Network Formation and Dynamics”, *PNAS* 112(21): 6595-6600 (2015). doi: 10.1073/pnas.1404770112