Applied Network Science

RESEARCH Open Access

Structural inequalities emerging from a large wire transfers network

Alfonso Semeraro[1]*, Marcella Tambuscio[1,2], Silvia Ronchiadin[3], Laura Li Puma[3] and Giancarlo Ruffo[1]

*Correspondence:
alfonso.semeraro@unito.it
[1]Dipartimento di Informatica, Università degli Studi di Torino, Turin, Italy
Full list of author information is available at the end of the article

Abstract

We aim to explore the connections between structural network inequalities and bank's customer spending behaviours, within an entire national ecosystem made of natural persons (i.e., an individual human being) and legal entities (i.e., private or public organisations), different business sectors, and supply chains that span distinct geographical regions. We focus on Italy, that is among the wealthiest nations in the world, and also an example of a complex economic system. In particular, we had access to a large subset of anonymised and GDPR-compliant wire transfer data recorded from Jan 2016 to Dec 2017 by Intesa Sanpaolo, a leading banking group in the Eurozone, and the most important one in Italy.

Intesa Sanpaolo wire transfers network exhibits a strong heavy-tailed behaviour and a giant component that grows continuously around the same core of the 1% highest degree nodes, and it also shows a general disassortative pattern, even if some ranges of degrees' values stand out from the trend. Structural heterogeneity is explored further by means of a bow-tie analysis, that shows clearly that the majority of relevant, in terms of transferred amount, transactions is settled between a smaller set of nodes that are associated to legal entities and that mostly belong to the strongly connected component. This observation brings to a more comprehensive inspection of differences between Italian regions and business sectors, that could support the detection and the understanding of the interplay between supply chains.

Our results suggest that there is a general flow of money that seems to stream down from higher degree legal entities to lower degree natural persons, crossing Italian regions and connecting different business sectors, and that is finally redistributed through expenses sharing within families and smaller communities. We also describe a reference dataset and an empirical contribution to the study on financial networks, focusing on finer-grained information concerned about spending behaviour through wire transfers.

Keywords: Complex networks analysis, Financial networks, Wire transfers, Assortativity, Bow-tie analysis, Spending Behaviour, Supply chains

Introduction

Complex Network Analysis (CNA) provides a set of quantitative measures that allows scientists and professionals to study the structure and the dynamics of a wide range of systems that can be represented and modelled by means of a graph. These measures have been proved to be exceptionally useful to capture, understand and predict emerging phenomena in many different complex systems, such as biological, social, physical, financial and economic. In fact, network based mathematical, computational, and visualisation tools can be seen as parts of a very powerful multi dimensional model of the system under study: as a "globe" is a model of Earth, showing in a blink of an eye water bodies, landmasses, and mountain chains, but also allowing for a deeper inspection through latitudes and longitudes to find nations and cities, a graph supports the data analyst to zoom in from global to local structures, unfolding hidden and significant patterns. Nevertheless, as pointed out in many introductory and more advanced textbooks (Newman 2010; Caldarelli and Chessa 2016; Barabási and et al 2016), spanning from theoretical aspects to more practical applications, the interpretation of graph measures is not always intuitive: when network science is applied to make sense of complex data sets that are produced massively every day, the connection from local to global quantities may remain obscure and hard to explain.

In our work, we aim to explore the connections between structural network inequalities and bank's customer spending behaviours, within an entire national ecosystem made of natural persons (i.e., individual human beings) and legal entities (i.e„ private or public organisations), different business sectors, and supply chains that span distinct geographical regions. We focus on Italy, presenting the results of our study on the Intesa Sanpaolo (ISP) wire transfers network, built from data described in "Dataset description" section, that exhibits a strong heavy-tailed behaviour and a giant component that grows continuously around the same core of the 1% highest degree nodes. The general trend of the degree correlation is clearly disassortative, showing a signal of a strong hierarchical behaviour where high degree nodes (big companies or institutions) pay salaries to natural persons or services to lower degree businesses. Such a disassortative behaviour changes locally when wire transfers are settled between natural persons: in such interactions expenses sharing inside families or smaller communities seem to better explain economic transactions. These structural heterogeneity is explored further by means of a bow-tie analysis, that shows clearly that the majority of relevant, in terms of transferred amount, transactions is settled between a smaller set of nodes that are associated to legal entities and that mostly belong to the strongly connected component. This observation brings to a more comprehensive inspection of differences between Italian regions and business sectors, that could support the detection and the understanding of the interplay between supply chains.

In the next section, we review the existing literature on this subject to highlight our specific contribution w.r.t. the state of the art.

Related work

The literature about the applications of network science to represent and analyse money flows is quite large and mainly involves three categories: (i) systemic risk in financial networks, (ii) money laundering and fraud detection and (iii) spending behaviour analysis.

Systemic risk in financial networks

Researchers have used networks to understand relations among financial institutions (primarily banks, but also firms or organisations) (Schweitzer et al. 2009). Most of these works focus on the description of networks structure, highlighting characteristics commonly found in other empirical complex networks, such as a scale-free degree distribution (Boss et al. 2004; Soramäki et al. 2007; Inaoka et al. 2004). Nevertheless, there are some contradicting results, showing the inherently complexity of this domain: for instance, in Boss et al. (2004) the authors find that the Austrian inter-bank network exhibits low clustering coefficient, a relatively short average shortest path length and a community structure that matches with the regional organisation, while in Soramäki et al. (2007) the inter-bank payments transferred between US commercial banks exhibits high clustering coefficient and point out the dynamic evolution of the structure, since they observed that the properties of the network changed considerably after terrorist attacks on 2001. Moreover, there has been considerable growing interest in *financial contagion*: the default of a bank may lead to default of its creditor banks, and so on, triggering *failure cascades* (Caccioli et al. 2018). Preliminary work in this field focused on exploring the stability of the whole system and investigated in which is the role of the network structure in this epidemic process in order to reduce *systemic risk* (May and Arinaminpathy 2009; Haldane and May 2011; Lublóy 2006). It has been found that diversification (more counterparts) initially enhances the network connectivity, virtually helping a cascading failure, but higher diversification leads to more stable systems (Elliott et al. 2014). On the other hand, in Acemoglu et al. (2015) the authors show that this contagion exhibits a phase transition: if failures are relatively small, a more connected network guarantees stability, but beyond a certain level, big failures can propagate more easily if there are many links among banks, leading to more fragile configurations. Moreover, different methods have been proposed to identify the most influential nodes (Battiston et al. 2012; Amini et al. 2016), recovering missing links (Caldarelli et al. 2013) and planning investments in strategic locations (Pozzi et al. 2013). A multi-layer network approach has been proposed in Battiston et al. (2016) to better represent the complexity of the system: the authors found that a partial knowledge of the contracts (links) among financial institutions can affect the estimating default probabilities and the ability to mitigate systemic risk.

Money laundering and fraud detection

Complex network analysis has been exploited to develop approaches (based on literature about criminal networks) in order to prevent and detect malicious financial activities. An interactive tool that combines visual analysis with k-core clustering has been proposed in Didimo et al. (2011): visual tools can be very helpful in detecting fraud in a small network, but they are hardly useful in large collections of transactions. A similar approach has been presented in Dreżewski et al. (2015), where centrality measures have been used to evaluate the role of individual nodes. In Colladon and Remondi (2017) the authors try to characterise the risk profile of customers of a factoring company: the nodes with higher values are less central in the transactions network and usually deal with larger amounts operations.

Spending behaviour analysis

Researchers have studied human economic relations revealed by payments and credit cards networks. In Dong et al. (2018) authors study community level human purchase behaviour, giving empirical evidence that the number of 'social bridges' between urban communities is a much stronger indicator of similarity in individual purchase behaviour than other factors such as income and socio-demographic variables. Such individuals acting as social bridges are people that live in some given communities, and that work at close-by locations.

Also recently, the proliferation of e-commerce platforms allowed a comparison of spending data and social interactions as face-to-face and mobile phone logs (Singh et al. 2013) or online social profile information (Zhang and Pennacchiotti 2013) that revealed that social behaviour can be useful to predict purchasing activities (propensity to buy, business diversity, loyalty). Comparison among communication data and bank transactions have also highlighted the presence of inequalities: in Leo et al. (2016) the researchers observed that social structure is strongly stratified and exhibits homophily according to socioeconomic classification. A similar analysis over credit card transactions (Sobolevsky et al. 2017) highlighted clear correlations between individual spending behaviour and social socioeconomic indexes denoting quality of life. The comparison among financial data (money transactions) and social media data can be used also to measure the attractiveness of places and cities (Sobolevsky et al. 2015). Moreover, such analysis could also be used to assess the individual risk and credit scores, traditionally computed from the user's past financial history. An extensive analysis of credit cards transactions data pointed out that features like diversity of shopping patterns, loyalty to the same vendors and regularity in payments are better predictors for possible financial difficulties than demographic indicators (Singh et al. 2015). Based on similar concepts, MobiScore provides credit scores when no financial history is available, using mobile phone network data patterns and supervised machine learning methods (San Pedro et al. 2015).

Finally, non conventional data sources as mobile phone network records can also be used to analyse social diversity in large urban areas; in particular, in Beiró et al. (2018) researchers have proved that people tend to choose a profile of malls more in line with their own socio-economic status and the distance from their home to the mall, and that higher mixing does positively contribute to the process of choosing a mall.

Our contribution

Networks are a great tool to make sense of the strong interdependence between many different actors (natural persons and legal entities), distinct business sectors that fuel the engine of national supply chains, and geographic regions that show quite heterogeneous spending behaviours. The network model allows us to capture high level dynamics, but also to zoom into smaller communities and to spot local dynamics, whose importance for the entire ecosystem may be harder to assess. Our results suggest that network science may be used successfully to observe structural heterogeneity that are easy to contextualise in the complex Italian financial system, also highlighting profound inequalities and strong interdependence between different business sectors. Also, we found a bigger picture of the general flow of money that seems to stream down from higher degree legal entities to lower degree natural persons, and that are finally redistributed through expenses sharing within families and smaller communities. Our study provides a framework that could be useful to other researchers to explore similar data sets and compare results: a deep understanding of the interplay among money exchange network structure and spending behaviour is indeed of great interest for economists, policy makers, sociologists and of course, network scientists.

Research case's description

In this section we describe our research case, providing a brief overview of the Italian financial system, our data provider, and the wire transfer dataset that we analysed.

Italian financial system

Italy is among the wealthiest nations in the world, and also an example of a complex economic system (Pugliese et al. 2019): according the International Monetary Fund (IMF) (World Economic Outlook Database 2018), Italy is the 3rd-largest national economy in the Eurozone, the 8th-largest by nominal Gross Domestic Product (GDP) in the world, and the 12th-largest by Purchasing Power Parity (PPP). Italian economy is considered a major advanced one, and Italy is a founding member of the European Union, the Eurozone, the Organisation for Economic Co-operation and Development (OECD), the G7 and the G20.

Nevertheless, like many other European countries, also Italy is characterised by large inequalities in the distribution of wealth and income. The gap between the North and the South of the peninsula is notorious (Ciani and Torrini 2019), and a heterogeneous distribution is usually observed also at finer granularity such as regions, provinces, and municipalities.

In the Istituto Nazionale di Statistica (ISTAT) 2017 report on regional economy (Report ISTAT 2017) there is the full detail of the imbalance between richest and poorest regions in Italy. For instance, it is stated that the GDP per capita of the whole South Italy in 2017 is 18.500€, 45% lower than the Centre-North GDP per capita.

Intesa Sanpaolo Innovation Center and wire transfer transactions data

The main driver of this work is a joint collaboration framework between University of Turin and Intesa Sanpaolo Innovation Center that has been established in 2015. One of the main outcomes of this agreement is the collection and the availability of a significant subset of wire transfer transactions recorded from Jan 2016 to Dec 2017, that we will describe below.

Please, notice that ISP is a leading banking group in the Eurozone, and the most important one in Italy, with 15% of national market share and more than 12% in 17 out of 20 Italian regions (Intesa Sanpaolo 2020).

Intesa Sanpaolo Innovation Center is part of ISP group, and its mission is exploring business models of the future to discover new assets and skills that support the long-term competitiveness of ISP group and its customers. ISP has established the Innovation Center Labs to respond to the complex needs of the bank and the market, determined by the evolution of market trends and exponential growth technologies.

Anonymised and General Data Protection Regulation (GDPR)-compliant data have been provided by ISP, through Intesa Sanpaolo Innovation Center; any researcher interested to access this dataset to replicate the analyses described here, as well as to propose a new line of experiments on these data, should contact directly Intesa Sanpaolo Innovation Center co-authors of this paper to start their internal data release procedure.

Generally speaking, a *wire transfer* is method to exchange money among people or organisations, typically from one bank account to another. Wire transfer is a quite common payment method in Italy: even if it represents almost 30% of overall non-cash transactions, where payments made with credit/debit cards are the most used with more than 50% of non-cash transactions, wire transfer is the preferred method alternative to cash in terms of total amount, with approximately the 80% of the overall transferred money volumes. Latest official Bank of Italy statistics on payments can be found in the 'Payment System - March 2019-September 2019' report (Bank of Italy 2019).

A wire transfer can be made in three different ways: in cash by the payer at a physical office, using an Automatic Teller Machine (ATM) or or through internet banking. It may require a small commission charge for the payment to be sent and generally it can take a few days. For these reasons, the wire transfer protocol is impractical for fast purchases, while it is a standard for fixed or large amounts payments, such as rents or salaries, because it is direct, traceable and can be scheduled for recurring operations.

In our analysis, by using wire transfers only, we tried to catch those payments which underlie some kind of social and civic relationship between the payer and the payee, ruling out extemporaneous expenditures.

Dataset description

In this study[1], we considered the whole set of wire transfer transactions recorded in 24 months by Intesa Sanpaolo from Jan 2016 to Dec 2017. According to GDPR (Regulation (EU) 2016/679 of the European Parliament and of the Council 2016), Intesa Sanpaolo provided only data of customers that gave their consent for data treatment (90.73% of the total). Moreover, the dataset has been anonymised because of the sensitive nature of the data: each customer and each transaction have been identified by a randomly generated *ID*, and sensitive information (i.e. name, address, phone number. . .) has been removed. Customers accounts' geographical location has been inferred from the office branch in which the contract have been stipulated, and not retrieved, for instance, by the last personal address the customers notified to the bank. Although this method is prone to errors, as one customer may have opened an account in city A and then moved to city B, it allows a better privacy-preservation policy to protect individual's sensitive information. Observe also that in the remainder of the document, we differentiate 'persons' as 'natural persons' and 'legal entities'; in fact, in jurisprudence, not all persons are people: some are companies or organisations that are legal persons, but not in the ordinary sense. We prefer to make such distinction with the less ambiguous definition of *legal entity* when we refer to a private or public organisation, as opposed to *natural person* when we refer to an actual human being that opened a bank account.

After the above mentioned preliminary privacy preservation operations, the dataset still contains a lot of exploitable information about customers and wire transfers. Details are given below:

- *Customers*: the registry of bank financial products holders, all the active accounts in the period of observation;

[1] For reproducibility purposes, please observe that we have run our analyses with the following system configuration: OS and Hardware: Linux 2.6.32-696.el6.x86_64, 8 CPUs, Intel(R) Xeon(R) CPU E5-2697 v2 @ 2.70GHz, Disk 516GB, RAM 64GB; Languages, packages and software: R 3.3.3, Python 3.5.1, igraph, networkx, Gephi.

  - Observations: more than 12M bank account owners (11.5M natural person, 500K legal entity), for 12.7M different bank accounts
  - Data Collection Period: January 2016 - December 2017
  - Customer Dataset contains information on Intesa Sanpaolo bank account owners. Information collected includes *anonymised internal ID*, *Associazione Bancaria Italiana (ABI) code*, that is the national bank code; *Codice di Avviamento Bancario (CAB) code*, that is the office branch code; user's *gender*; *birth year*; *ATtività ECOnomiche (ATECO) code*, that briefly describes the economic activity's type, and *legal status*, i.e., natural person or legal entity. From ABI + CAB codes we can retrieve the *office* in which the contract has been signed, and of course the corresponding *geographical location*. As stated above, please notice that the dataset contains only information for customers that allowed their data treatment. Moreover, each customer can have more bank accounts, and each bank account can be shared by multiple customers.

- *Wire transfers*: each wire transfer transaction involves one payer and one payee, both of them corresponding to an ISP *customer*, as already defined. Details are given below:

  - Observations: 100 million transactions, worth over 1000 billion euros
  - Data Collection Period: January 2016 - December 2017
  - Collected information includes wire transfer *anonymised ID*, *payer* and *payee ABI/CAB codes*, *payer* and *payee anonymised International Bank Account Number (IBAN) codes*, *payer* and *payee internal anonymised ID*, *channel*, *timestamp*, *amount*, and a free text note stating the *purpose of the payment* edited by the payer.

Representativity of the dataset: customers data

Data analysed here are representative of the Italian case, because of the leading position of Intesa Sanpaolo in the Italian financial market. Therefore it likely hosts a substantial part of the financial person to person transactions among all kinds of customers across the country. In addition to the overall good geographical distribution, the dataset reflects the Italian population age distribution quite well: in Fig. 1 we show the comparison among two age frequency histograms representing age distributions, one calculated from ISP customers' data, and the other retrieved from official 2017 data provided by ISTAT (Statistiche Demografiche ISTAT 2017). Apart from significant differences in the first two age ranges, that can be explained with the fact that children and teenagers are less likely to open a bank account, similar proportions are observable in the two plots.

In the dataset, natural persons are the most represented, making 88% of the total 12.7 millions of bank accounts, while legal entities own the remaining 12% accounts. However, the number of wire transfers is more than twice for legal entities than for natural persons, with a total amount of an order of magnitude higher. In Fig. 2 we show that the highest overall amount is transferred from legal to legal entities, even if the majority of wire transfers are from legal to natural persons.

An account is usually owned by a singular person (69%), but it can be shared by two owners (23.9%) or more (the remaining 7.1%). Accounts with more than 20 owners are just a few, created mainly for business purposes.

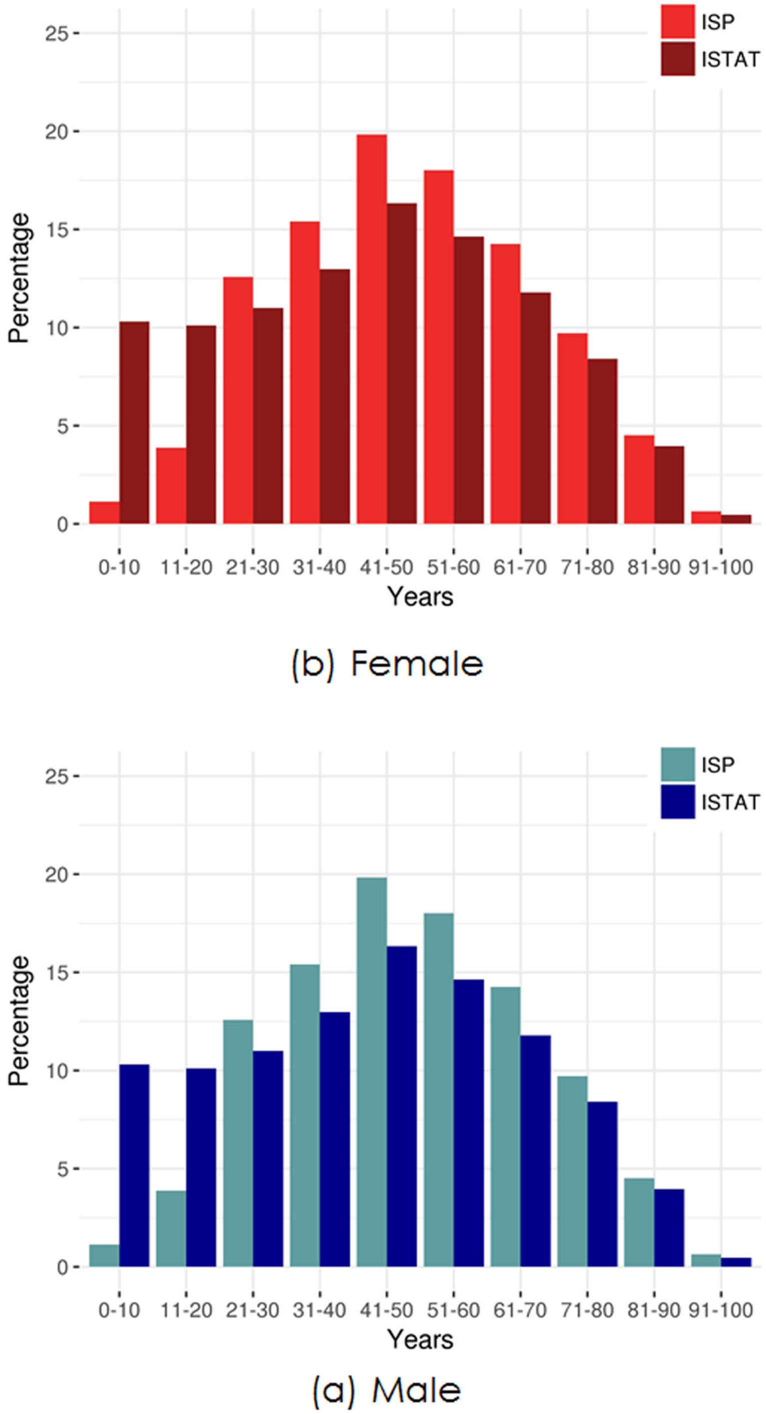
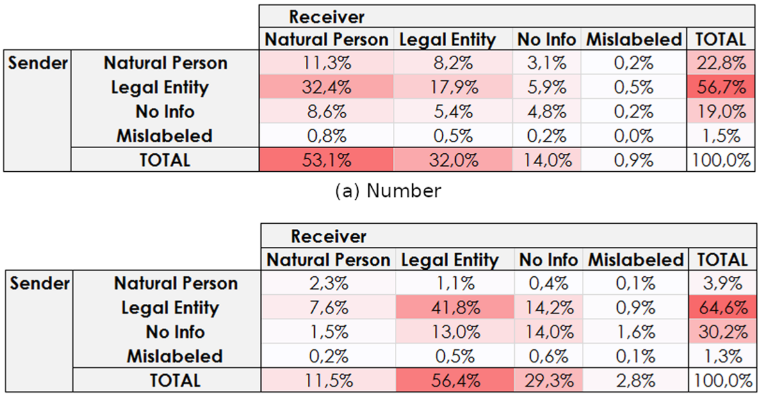


Fig. 1 Age distribution: Comparison among age frequency histograms for ISP bank account holders and Italian population, both for female (above) and for male (below). We can see that ISP dataset reflects Italian population, with the exception of elderly and underage citizens. Official Italian population by age have been retrieved from the ISTAT on-line data base (Statistiche Demografiche ISTAT 2017)

Representativity of the dataset: wire transfers data

We integrated the analysis with some information from the Customer Registry (as the distinction between legal entities and natural person and the ATECO code). From ABI



		Receiver				
		Natural Person	Legal Entity	No Info	Mislabeled	TOTAL
Sender	Natural Person	11,3%	8,2%	3,1%	0,2%	22,8%
	Legal Entity	32,4%	17,9%	5,9%	0,5%	56,7%
	No Info	8,6%	5,4%	4,8%	0,2%	19,0%
	Mislabeled	0,8%	0,5%	0,2%	0,0%	1,5%
	TOTAL	53,1%	32,0%	14,0%	0,9%	100,0%

(a) Number

		Receiver				
		Natural Person	Legal Entity	No Info	Mislabeled	TOTAL
Sender	Natural Person	2,3%	1,1%	0,4%	0,1%	3,9%
	Legal Entity	7,6%	41,8%	14,2%	0,9%	64,6%
	No Info	1,5%	13,0%	14,0%	1,6%	30,2%
	Mislabeled	0,2%	0,5%	0,6%	0,1%	1,3%
	TOTAL	11,5%	56,4%	29,3%	2,8%	100,0%

(b) Amount

Fig. 2 Distribution of wire transfers by number (top) and amount (bottom): the highest overall amount is transferred from legal to legal entities, even if the majority of wire transfers are from legal to natural persons

and CAB codes we could extract the geographical information about the sender and the receiver: the bank provided a list of ABI, CAB, addresses of its branches, from which we retrieved the geographical coordinates. As ISP's market share may vary on a local basis, meaning that there are localities underrepresented in the data due to a lower percentage of ISP customers, we grouped the wire transfers at a regional level to check mutual Pearson correlations between different distributions, i.e., regional GDP, incoming/outgoing wire transfers number, incoming/outgoing wire transfers total amount (see Fig. 3). All the mutual correlations are very high. For example, the outgoing wire transfers total amount positively correlates with the GDP of Italian regions (Gross Domestic Product (GDP) at Current Market Prices by NUTS 2 Regions 2019) ($r = 0.97$, p-value $= 4.1 * 10^{-13}$).

The dataset covers a two years period of time, which allows us to observe how money circulates in a regular week, in a regular month, within the whole year or in the proximity of holidays and events, and to compare different time windows. Results are showed in the calendar plots in Fig. 4. As expected, there are few transfers during the weekend and Italian holidays, and generally we can observe an increment, both in number of transactions and total amount, in the last days of each month (probably due to the salaries payments) and in particular in June and December (probably bonus salaries and taxes). Despite the spikes at the end of the month, the number of transfers seems to be generally more regular than the amount, which fluctuations can be heavy and hard to explain. Christmas time is the most active in the year, while August is the most economically depressed month in the year, maybe because of vacations: it is likely that during vacations, bank customers prefer to use credit cards or cashes over wire transfers, that could partially explain why this payment method is underused. Moreover, and probably mostly significant, many companies are closed for summer break; in fact, recall that the total amount is strongly dependent on legal entities activities.

While the above considerations hold for both 2016 and 2017, there are several punctual differences between the two years, due to fluctuations in the economic activity of all the players involved. Explaining such a variance is non trivial, and it would necessitate the application of natural language processing tools on the purpose of the payments as declared by the payers.

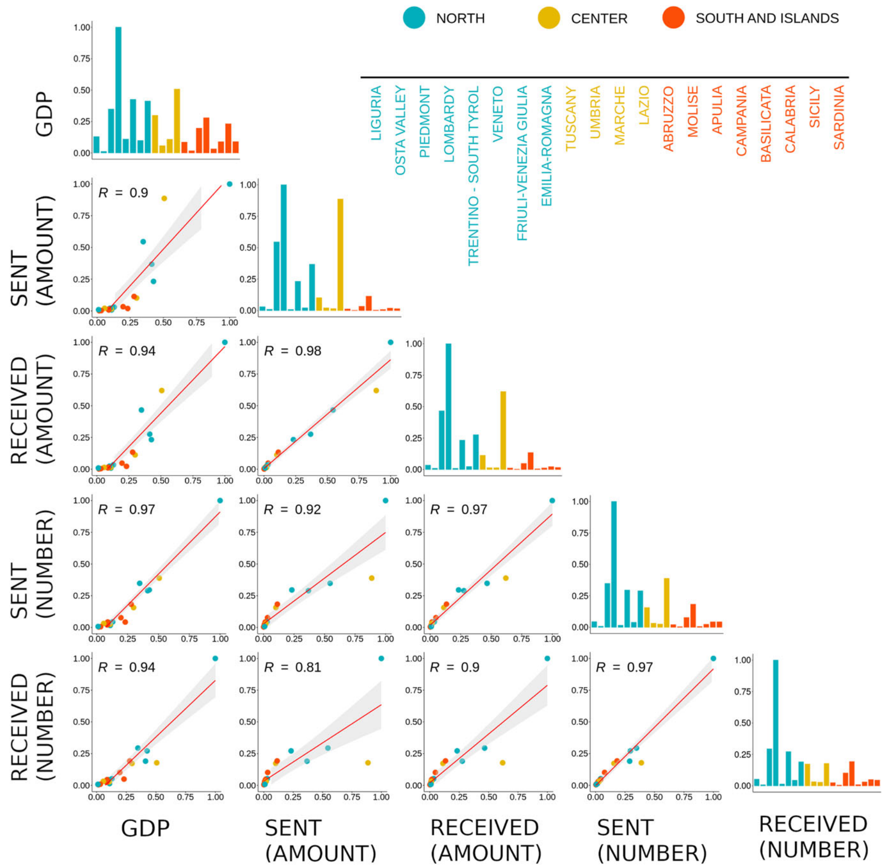


Fig. 3 Correlation scatter plots matrix: Mutual correlations between regional GDP, incoming/outgoing wire transfers number, and incoming/outgoing wire transfers total amount, all grouped by regions, are very high. Wire transfers number correlates better with GDP, also because wire transfers total amount exhibits some peculiar outliers (such as Lazio) that behave differently along the distributions. Notice that all the p-values are much lower than 10^{-4}. On the diagonal the plots for GDP, sent/received transfers (number and amount) reveal big inequalities between regions. Values have been normalised according to the maximum value (Lombardy) while on the x-axis regions follow the order shown in the legend. Please observe that due to the bank's policy, we cannot show the actual values of wire transfers numbers and amounts, but only their normalisation

The dataset used for our analysis is very informative and allows us to shed light on a substantial part of the Italian financial ecosystem, even considered that the data may cover differently the customer base depending on geographical region, age or business sector of the customers. If not specified, results showed in the following sections are related to 2017 data only. Results on 2016 data are generally consistent with those on 2017, and will not be displayed but for comparison.

Methods

In this section, we describe the techniques applied in our analysis: network analysis and bow-tie analysis. The main objective of such a section is to discuss the usefulness of these methods to address our research questions.

Network analysis

A wire transfer is a direct method of payment between two bank accounts, and a directed graph is an intuitive way to capture such an interaction. Network science can provide

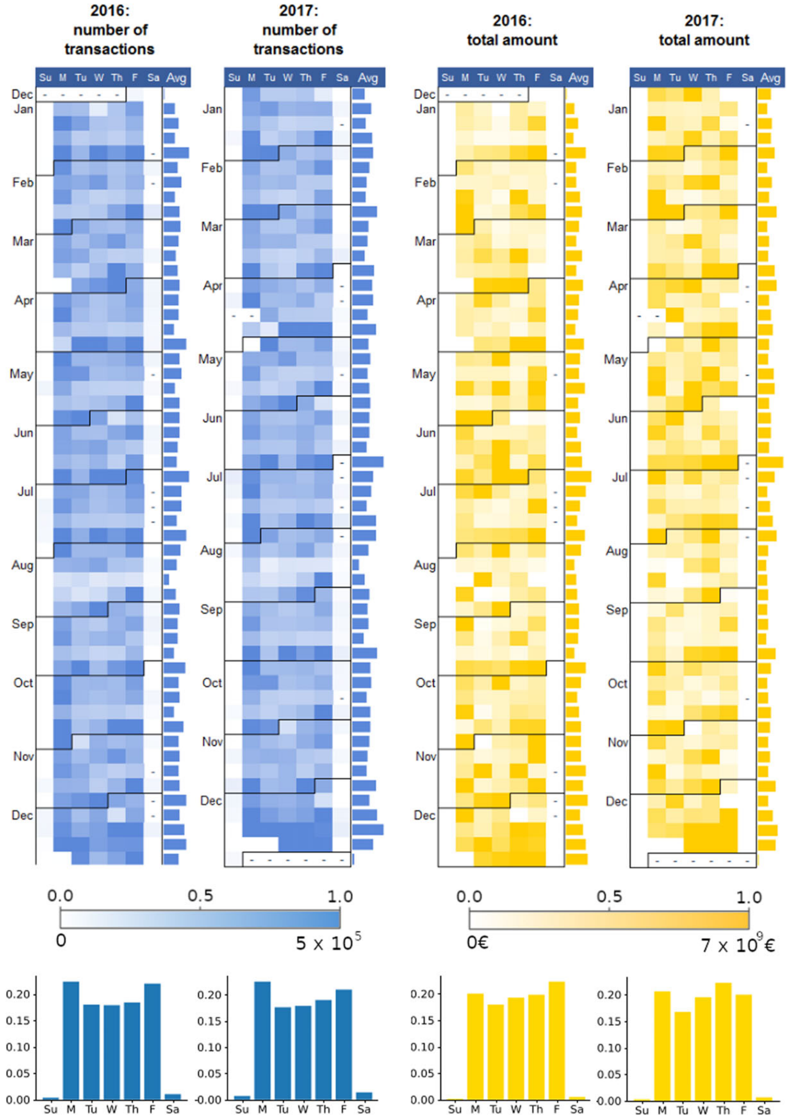


Fig. 4 Calendar maps: we show how the total daily number (blue) and amount (yellow) of transactions change between 2016 (first and third column) and 2017 (second and fourth column). We can see that there is an increment, both in number of transactions and total amount, in the last days of each month (likely due to the salaries payments) and in particular in June and December (likely because of bonus salaries and taxes)

powerful tools for understanding how money flows in this huge dataset of payments and basic network's structural properties, like connectivity, degree distribution and assortativity, can synthesise the complexity of such data into simpler measures. Networks are also useful in order to highlight regularities in the way money is exchanged between different kinds of users, or groups of users, or between local components. Moreover, by mixing the information related to the payment itself (time, amount) with the additional information taken from the Customer Registry about the network nodes (see "Dataset description" section), we could not only reconstruct the network structure, but we also address the issues about the time and location of this cash exchange, taking into account the differences in the types of customers and business categories involved. The goal of our analysis is combining these two perspectives: we captured a snapshot of the entire network by structural measures and at the same time we seek for a comprehensive global picture of the interactions among the bank customers. In particular, we aim to shed some light on the dynamics that regulate the exchange of money among natural persons, among legal entities, and between the two categories of costumers.

Following the relational nature of the data, we set the bank accounts $v_i \in V$ as the nodes of our directed graph $G = (V, E)$ and we add a edge $e_{ij} \in E$ between two nodes i and j if there is a wire transfer from account i to account j. Moreover, we labelled the edges with amount a_{ij} (in euros) and the date of the transaction d_{ij}. Observe that it is a *multigraph*, and many of these payments (like rents or salaries) are repeated on a monthly basis between the same couple of nodes. Finally, for each node i we have a degree k_i, that is the sum of the in-degree k_i^{in} and the out-degree k_i^{out}, and a strength s_i that is the weighed degree, i.e., the sum of the in-strength s_i^{in} and the out-strength s_i^{out}, s.t., $s_i^{in} = \sum_{\forall j: e_{ji} \in E}(k_i \cdot a_{ij})$ $s_i^{out} = \sum_{\forall j: e_{ji} \in E}(k_i \cdot a_{ij})$.

Degree heterogeneity is often observed in many real networks (Barabási and Bonabeau 2003; Newman 2001; Wagner 2003). In particular, such networks are said to be *scale-free* because of their degree distribution, that follows a power-law $p_k \approx k^{-\gamma}$, where γ is the degree exponent that in many real networks has been observed to be among 2 and 3. Exploring degree heterogeneity could be the first step toward the understanding of emerging inequalities in customers' behaviours and statuses.

Bow-Tie analysis

Bow-tie analysis is very useful in directed graphs (such as the Web Broder et al. (2000)) to identify the main core of the network and to follow the direction of flows entering to and departing from it. In our wire network, this analysis allows us to understand better the role of the highly connected accounts that are involved into most of the economic activities, and also which are the nodes that inject money into the system or the ones that mostly receive payments. More specifically, if we keep a distinction between natural persons and legal entities, we can look for the role of each kind of players in the principal components.

A bow-tie decomposition ends with finding the following components:

- Strongly Connected Component (SCC) contains the nodes of the network that have a path to every other in SCC, and vice versa;
- In-Component (IN) contains the nodes of the network from which it is possible to find a path leading to every node in SCC; however, it is impossible for nodes in SCC to reach upstream nodes in IN.
- Out-Component (OUT) contains the nodes of the network to which there exists a path from every node in SCC, but there is no route back.
- **Tubes and Tendrils** contain the nodes of the network that are reachable from IN, or reaching nodes in OUT, however never reaching SCC.
- **Disconnected** are all the other smaller components that contain nodes that would never reach the central SCC, even ignoring the directions of the edges.

Experiments and analyses

In this section we describe and discuss the results of our experiments, based on the techniques that we briefly introduced in "Methods" section.

Hubs and connectivity

As already discussed in "Dataset description" section, the majority of the network is composed by natural persons (see Fig. 2) that exchange relatively small amounts of money a few times per year. On the other side, there is a smaller set formed by big organisations and businesses involved in hundreds of thousands of transactions per year, connected with both natural persons (like customers or employees) and other legal entities. This unbalance between the two kinds of actors makes the degree distribution fat-tailed: Fig. 5 shows the power law given by the total degree distribution, along with the in-degree and the out-degree distributions, plotted on double logarithmic axis (log-log plots): indeed $\log p_k$ is expected to depend linearly on $\log k$, then in a log-log plot the power-law follows a straight line whose slope is the degree exponent γ. In Fig. 5 the black points correspond to the empirical data and the red line represents the power-law fit, in which the slope is

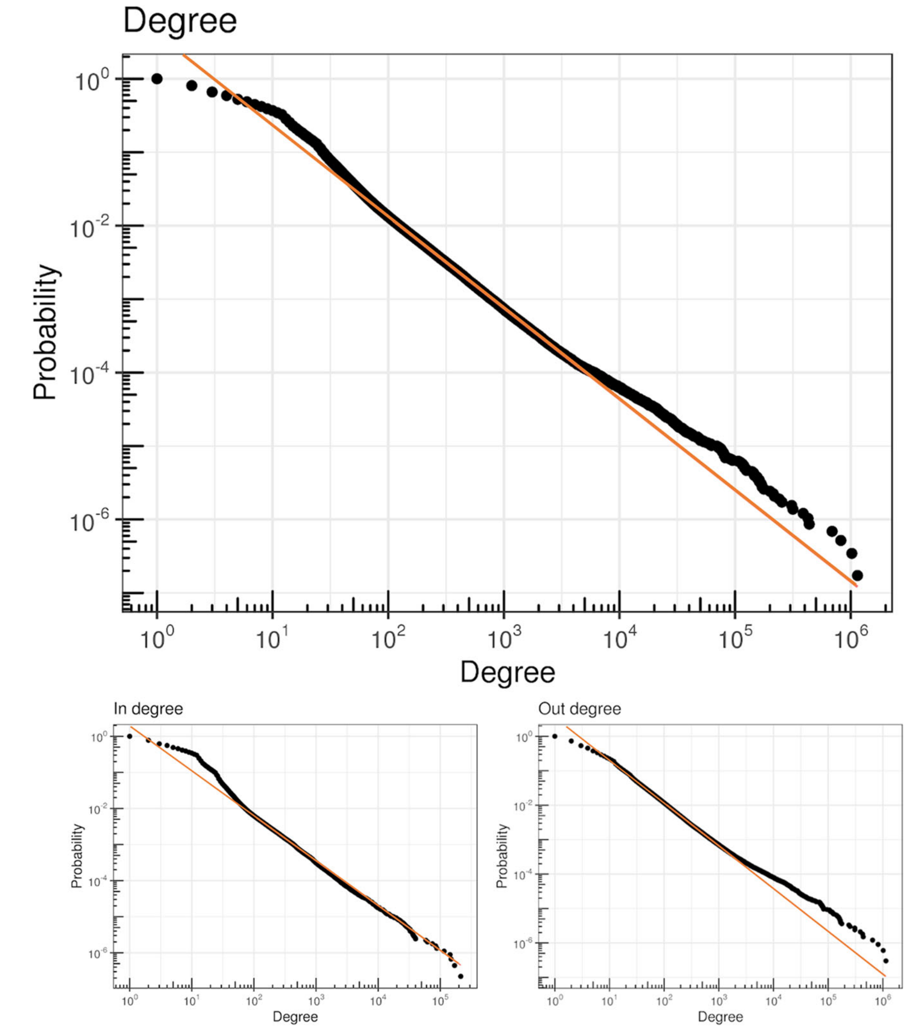


Fig. 5 Degree distribution Log-log plots of the degree distribution for the 2017 wire transfer network: top (total degree), bottom-left (in-degree) and bottom-right (out-degree). The slope of the fitting red line is between 2 and 3. Fitting line plotted even before *xmin* obtained from PoweRlaw library: *xmin* total = 111, *xmin* in-degree = 93, *xmin* out-degree = 29

between 2 and 3 in all three cases: $\gamma_{TOT} = 2.24$ for total degree, $\gamma_{IN} = 2.27$ for in-degree and $\gamma_{OUT} = 2.23$ for out-degree.

Figure 6 shows the degree distributions for the two types of actors of the network, natural persons and legal entities. As expected, natural persons are more likely to be among low degree nodes, but they become quickly rarer as the degree grows, as there is no natural person with degree higher than 2500, while nodes that represents bank accounts owned by legal entities can exhibit more than 10^6 connections. Please notice that here we plotted the degree distributions related to the 2017 network, because we want to highlight behavioural patterns within a 12-months time span. It is striking that a spike is shown for natural person's degree distribution around 12 degree nodes, suggesting monthly payments as also discussed later.

There can be several explanations for the high number of natural persons sending or receiving thousands of wire transfers and a small number of legal entities sending or receiving few wire transfers. For instance, a customer is labelled as natural person or legal entity only through a brief annotation when the contract is signed, but, after the signature, there is no auditing or control about the content of the annotation. Another possible reason could be that many business owners can have multiple accounts and can use them disregarding the main purpose of the account itself. As a matter of fact, further analysis would be necessary to explain the presence of small business and highly active natural persons in order to have a better comprehension. However, with their numerous incoming and outgoing links, big companies play a key role as the hubs of the network.

More than 82% of transfers in the network are repeated more than once between the same pair of nodes, since the payment may be often due to some kind of contract between the two owners, like buyer-seller or worker-employer relationships. This means that the key players of the network receive and send a very high number of wire transfers since the first days of observation: an analysis of the evolution in time of the Giant Connected Component (GCC) reveals how the whole network grows around the core of the 1% highest degree nodes since the 2017, 94% of nodes are connected into the GCC of the network (see Additional materials A).

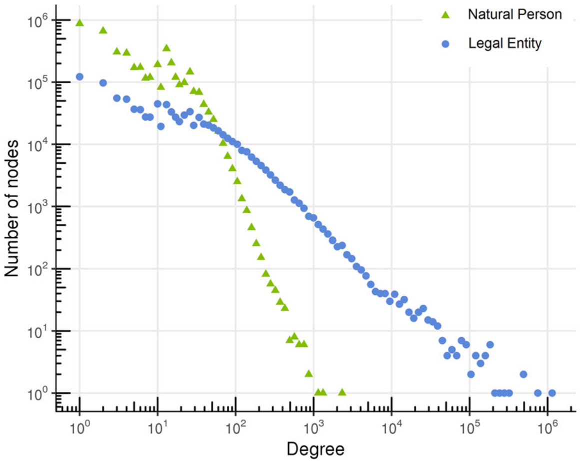


Fig. 6 Degree distribution by customer type: natural person (green) or legal entities (blue) in the 2017 wire transfer network

Figure 7 shows the growth of GCC as a function of average degree comparing Intesa Sanpaolo (ISP) network (red) with a series of synthetic Erdős-Renyi (ER) networks (Erdös and Rényi 1959) (green) that exhibit the same size (number of nodes and edges) of the first one. Specifically, since the edges of our network are labelled with the date in which the transfer has been done, we filtered progressively the edges obtaining a growing sequence of subgraphs from the whole network. For each of them we computed the mean degree and the percentage of the vertices that form the largest connected component, and compared the result with the same measures over the random ER graphs. It should be noticed that each point of the random reference model is the result of the averages of 10 different ER graphs. However, the different realisations return very stable values; for example, the GCC emerges (i.e., size of the largest component exceeds 50% of total nodes) when mean degree is between 1.383 and 1.387. We can see how the growth rate of the GCC is GCC is remarkably slower in our network. In particular, there is no phase transition between a disconnected connected network around $\langle k \rangle = 1$, as it happens in the random graph, that reproduces exactly what we expected from network theory (Erdös and Rényi 1959). Surprisingly, the largest component in ISP network is not the result of several smaller components melting into a bigger one over time, but instead is the result of a process in which the most important nodes of the network immediately bond into the same component, and the other nodes steadily join this core.

Among the highest degree nodes (first percentile), 94.5% are in the GCC since the first month of observation, that is January 2017. Their role in the connectivity of the network

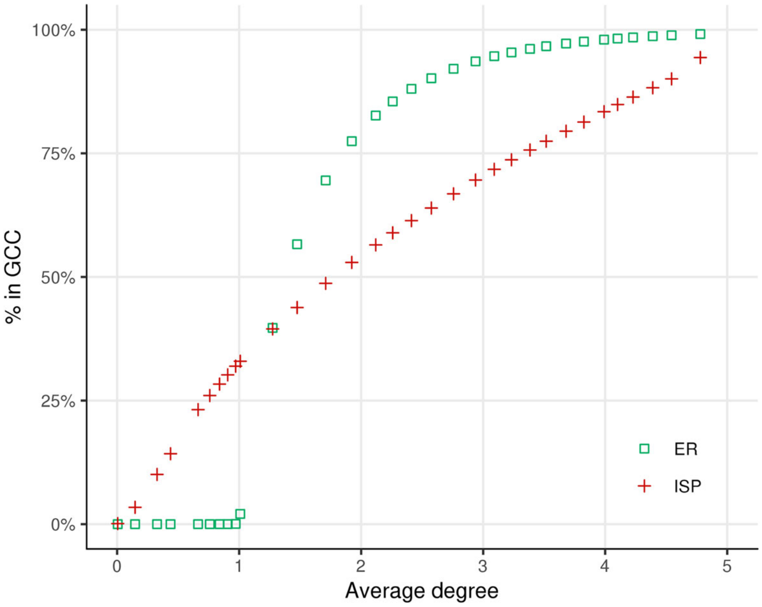


Fig. 7 Giant component compared to average degree: we compared the monthly evolution of the ISP network (red crosses) with a series of a Erdős-Renyi random graphs with same number of vertices and edges (green squares). Each point (x_i, y_i) represents a monthly observation of the network ISP (or its relative ER graph) that exhibits a mean degree $\langle k \rangle = x_i$ and a percentage y_i of nodes connected in its largest connected component. The ISP network (red) does not show phase transition, while Erdős-Renyi (green) graphs do just after $\langle k \rangle = 1$. It should be noticed that each point of the random reference model is the result of the averages of 10 different ER graphs. However, the different realisations return very stable values; for example, the GCC emerges (i.e., size of the largest component exceeds 50% of total nodes) when mean degree is between 1.383 and 1.387

is clearly highlighted also by the network's robustness, i.e., its ability to resist against random failures (usually defined as *node removal*). Although it is outside the scope of this paper, we mention here a preliminary result of a robustness attack we executed in our network that returns a signal that highest degree nodes are part of a backbone that makes the wire transfer network connected. The reader can find more details in the Additional materials A.

A network with mixed assortative/disassortative behaviour

The whole wire transfers network is the result of the interplay of the financial activities between natural persons and legal entities; now we explore in detail such interaction, answering to research questions as (i) how the two groups are related to each other, (ii) what drives their connections, (iii) if natural persons tend to share money with other natural persons or whether they preferentially interact with businesses or institutions. Hence, in this subsection we focus on some properties (degree correlation, assortativity, clustering coefficient) that can provide a general picture of the interactions among different groups of nodes.

First of all, we computed the *degree correlation function* k_{nn}, that is the average degree of the neighbours of all degree-k nodes. Lets keep in mind that in our network the degree is also a proxy for nodes' type: as observed above and clearly displayed in Fig. 6, low degree nodes correspond more frequently to natural persons and high degree nodes are mainly legal entities. Figure 8 shows the $k_{nn}(k)$ of the network (red squares): low degree nodes, whose majority corresponds to natural persons, tend to connect with neighbours with a high average degree, and the other way around. This general disassortative tendency is strong: a randomisation of the network (green crosses in Fig. 8) with degree distribution preservation exhibits the same tendency but with a slower decay, meaning that the disassortativity is only in part due to the degree distribution itself, exhibiting the so called "structural disassortativity" (Barabási and et al 2016). Plus, in the Intesa Sanpaolo's network, there is an inflection right over nodes with a degree of a few tens, where accounts are progressively more likely to be owned by legal entities. The high disassortativity of the network could be explained by the fact that natural persons (very likely to correspond to low degree nodes) tend to receive salaries and pensions from big businesses and to pay bills and taxes to high degree institutions; this interpretation however does not account for the high number of wire transfers among natural persons.

Additionally, Fig. 9 shows the *average clustering coefficient* C_k for all the k-degree nodes of our network (red squares). Low degree nodes tend to be more clustered, while high degree nodes have a lower clustering coefficient. Once again, if compared with a degree distribution preserving randomisation (green crosses), Intesa Sanpaolo's network shows a stronger dependency of the clustering coefficient on the degree of the nodes, as a result of a process driven by the reasons of the payments: natural persons pay each other for personal reasons within local communities and their families for sharing housing expenses, mortgage payments, and so on (more on this in "Insights from text analysis" section). This drives triangle closure among natural persons, while (big) companies spread their payments over different and distant receivers, without closing triangles so often. It is noticeable that for medium size businesses the clustering coefficient is halfway between natural persons and big companies: they still behave as a business, but working in a local territory they tend to pay and be payed by persons who may know (and pay) each other.

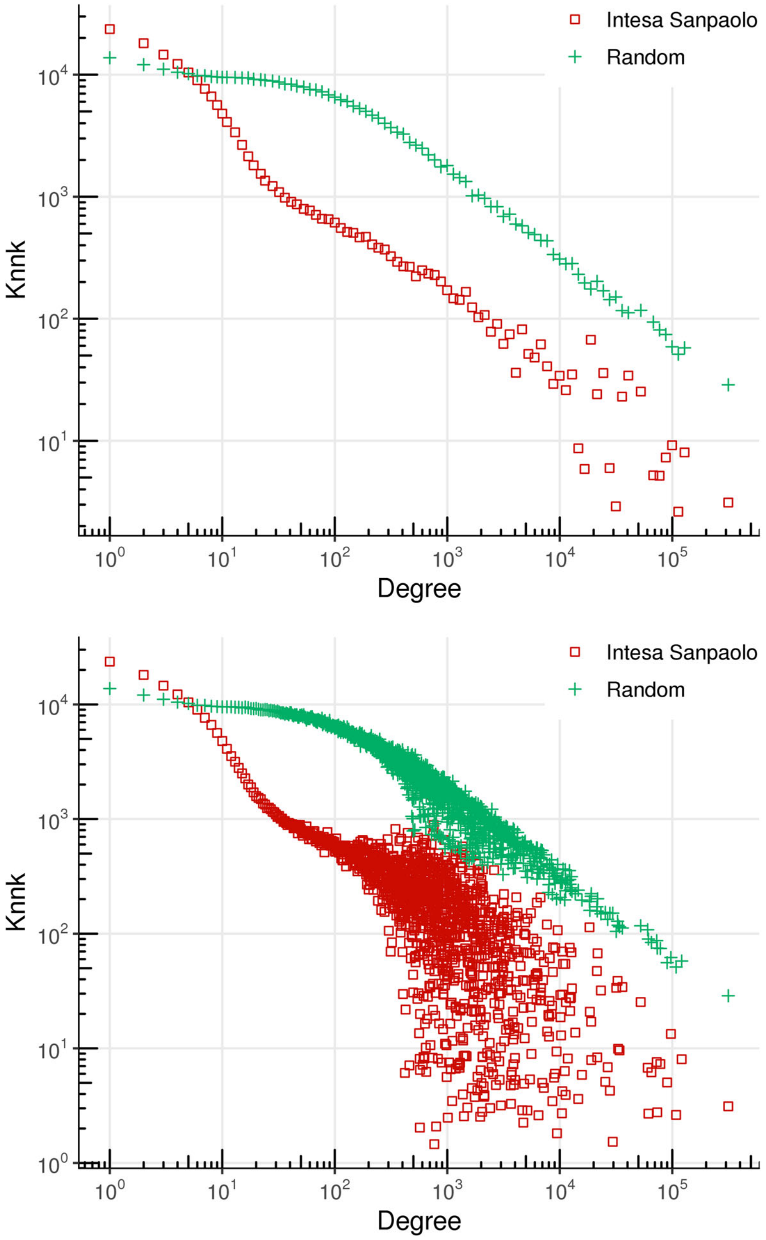


Fig. 8 $k_{nn}(k)$ for the wire transfer network (red squares) and its randomised version (green crosses), on a log-log scale with (top) and without logarithmic binning (bottom). In the ISP network, low degree nodes, tend to connect with neighbours with an average high degree, and the other way around. The random network shows a signal to "structural disassortativity", but the tendency in the ISP is much stronger. Plus, in the ISP's network, there is a clear inflection right over nodes with a degree of a few tens, where nodes are progressively more likely to be legal entities

Recalling that the degree is a proxy for discriminating roughly between natural persons and legal entities (Fig. 6), we could notice that $k_{nn}(k)$ seems to be more affected by such distinction: in fact, we have a curve's inflection right over 10^2-degree nodes. Conversely, the average clustering coefficient C_k bursts out beyond expectations for very low degrees ($k \sim 10$), and decreases smoothly immediately after. This could point out that the size of the node is the main reason for a higher or lower clustering coefficient, and not its type.

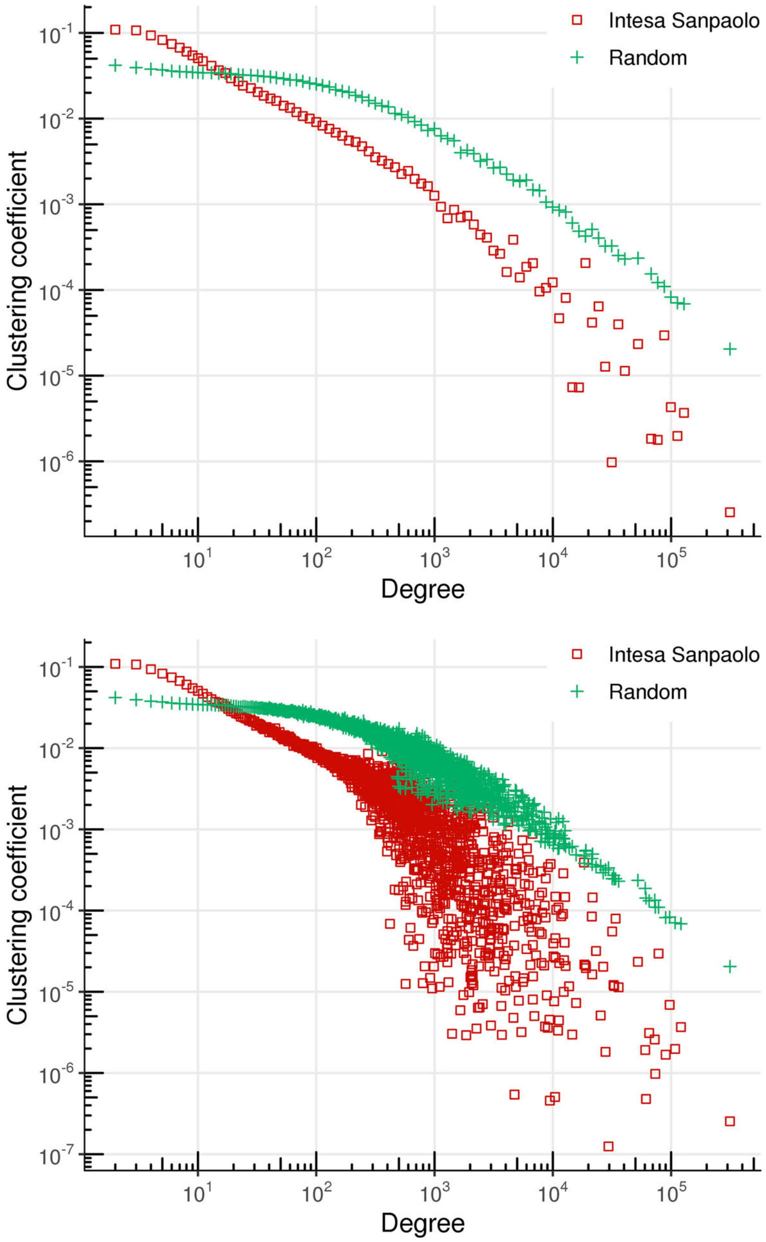


Fig. 9 Clustering coefficient as a function of the degree for the wire transfer network (red squares) and its randomised version (green crosses), on a log-log scale with (top) and without (bottom) logarithmic binning. We plot C_k, that is the average clustering coefficient of all the k-degree nodes. Intesa Sanpaolo's network shows a stronger dependency of the clustering coefficient on the degree of the nodes, as a result of a process driven by the reasons of the payments: natural persons (low degree, see Fig.6) pays each other in local communities and families, therefore many triangles are likely to be closed, while big companies (high degree) spreads their payments over different and distant receivers, without closing triangles so often

This dependency of the clustering coefficient on the degree of the nodes is a characteristic of hierarchical networks (Ravasz and Barabási 2003), where high degree nodes spread their links over many smaller nodes. In this network, if a hierarchy exists, the general hierarchical structure is ultimately broken at the lowest levels, where families and local clusters exchange money within themselves: they also pay or are payed by big businesses, closing the cycle and injecting money flows at the top of the hierarchy. However, more on directions of money flows is discussed in "Inter- and intra-components flows" section.

As observed above, this network tends to be generally disassortative, especially among big companies, which mainly interact with natural persons or legal entities corresponding to lower degree nodes. Furthermore, hubs appear to be less clustered; natural persons are instead more clustered in small groups, and they equally share money with companies and with natural persons too, making some communities of this network more assortative. We tried to capture this mixed behaviour of the network in the degree correlation heat maps of Fig. 10, by plotting the total number of wire transfers from a generic d_x-degree node to a generic d_y-degree node.

In the degree correlation heat map there are four areas of interest:

(a) In the bottom-left corner, there is a denser area around nodes with a degree $k \in [10, 100]$ that suggests a more assortative behaviour w.r.t. the rest of the network.

(b) The second area includes a dense horizontal band involving receivers whose degree is approx. $\in [10, 100]$, that get money from all ranges of senders: many of the related payments are probably salaries from companies to natural persons.

(c) The third zone is symmetric to the second one, and it shows that low degree nodes with degree $\in [10, 100]$ are paying other nodes with variable sizes.

(d) Finally, the top-right part of the heat map, involving business and institutions only because of the higher degrees of the nodes involved, is noisy and sparser, and no clear signal arises from it.

Assuming that a regular account owner would pay a rent or a mortgage per month, receive a salary or a pension per month, pay at least one average bill per month, plus other expenditures, most of the activities involving natural persons are expected to be represented by nodes with a degree in the range [10, 100]. This hypothesis is also supported by the degree distribution by customer type (Fig. 6) that clearly shows that in such a range we have more natural persons than legal entities.

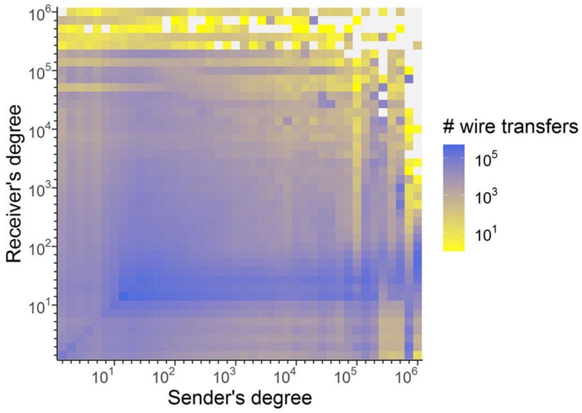
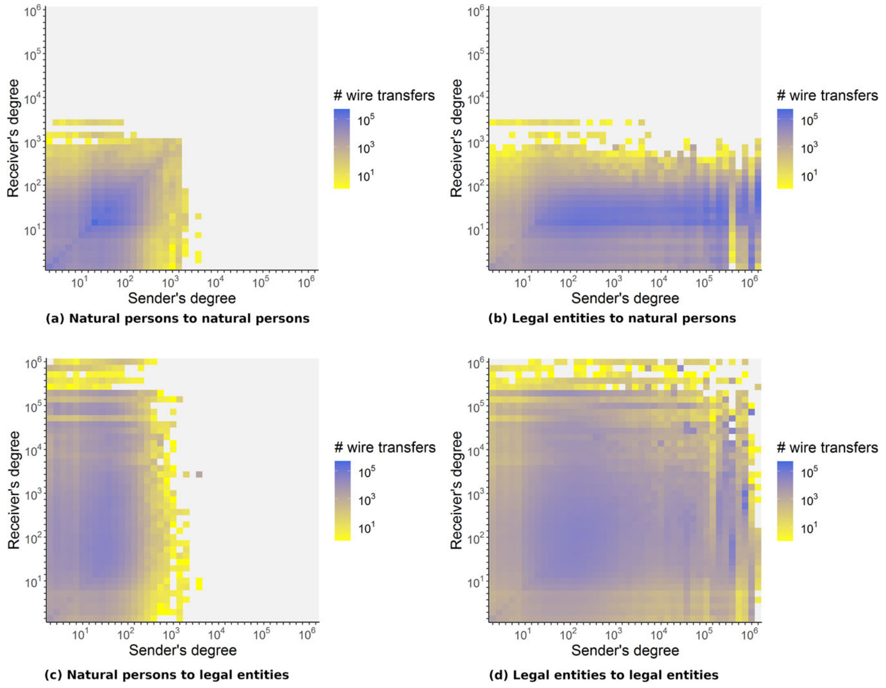


Fig. 10 General degree correlation heat map: The heat map plots, for each point $(d_x; d_y)$, the number of directed links connecting nodes with degree d_x to nodes with degree d_y. Here we refer to the 2017 wire transfers network

In order to make the above identified patterns easier to read and to have some deeper understanding of the natural persons/legal entities different behaviours, we plotted in Fig. 11 other four heat maps according to the declared type of the customers. In Fig. 11a we depict the interactions between natural persons only: it is clear that wire transfers between natural persons involve only lower degree nodes and that there is a stronger signal of assortativity opposed to the general disassortive trend of the whole network. Fig. 11b shows the legal to natural persons payment dynamics, suggesting the hypothesis of a general employer-employee pattern because of the darker horizontal band that emerged also in the general heat map. Fig. 11c shows the payments from natural to legal persons: this plot highlights that natural persons send payments to businesses with variable size, independently on their node's degree. We noticed that many of these payments are set on a regular basis. Fig. 11d, finally, shows a noisy and unclear signal from businesses to businesses, whose payments are spread almost uniformly in the full spectrum; quite predictably, the area related to payments from hubs to hubs is much sparser.

It should also be observed that the heat map shown in Fig. 10 can be reconstructed summing up all the values in the four plots displayed in Fig. 11. Therefore, stacking all the plots except (d), we have cleaner signals of the payments dynamics involving natural persons. Nevertheless, degree correlation does not help much to unveil legal to legal entities patterns. These dynamics will be explored in more details in "Inter- and intra-components flows" section, while further analyses on degree correlation and assortativity can be found in Additional materials C, where we also ran the analysis proposed in (Maslov and Sneppen 2002) to get a fuller picture of the underlying structure.



(a) Natural persons to natural persons

(b) Legal entities to natural persons

(c) Natural persons to legal entities

(d) Legal entities to legal entities

Fig. 11 Degree correlation heat maps: these heat maps have been created filtering the declared type (natural person or legal entity) of the payees and the payers. Again, *x* axes represents sender nodes' degree, and *y* axes is the receiver nodes' degree

Insights from text analysis

In order to dig deeper in the dynamics underlying the patterns we discussed above, we analysed also the *purpose of the payment* field, that is edited by the payer when the wire transfer is settled. As an example, we extracted this information from the wire transfers settled during January 2016, and we found out that:

(a) the most common topic used for natural persons paying natural persons is related to rent and housing expenses;
(b) for legal entities paying natural persons is related to salary and pension;
(c) for natural persons paying legal entities is related to payment, fee and settlement;
(d) for payments from legal entities to legal entities, the most recurrent words are "settlement" and "invoice", but those labels are too wide to give further insights on the reasons behind the payments.

More details of this analysis will be provided in Additional materials B. In order to mitigate the noise we found while exploring the legal to legal entities interactions, we need to adopt some different analytical tools to better understand what happens when legal entities are involved on both sides of the wire transfers, using also business sectors information. This is one of the main points addressed in the next section.

Inter- and intra-components flows

Among the other results we described so far, we found a signal of a hierarchical structure in the wire transfer network, that suggests the emergence of flows of money that stream down from higher degree nodes to lower degree nodes. As a partial validation of this hypothesis, we have that the majority of edges of the networks (68.9%) are established from higher degree nodes to lower degree ones (i.e., for these edges $e_{ij} \in E \implies k_i > k_j$), while the transfers that make the money flow upstream are rarer (29.4%).

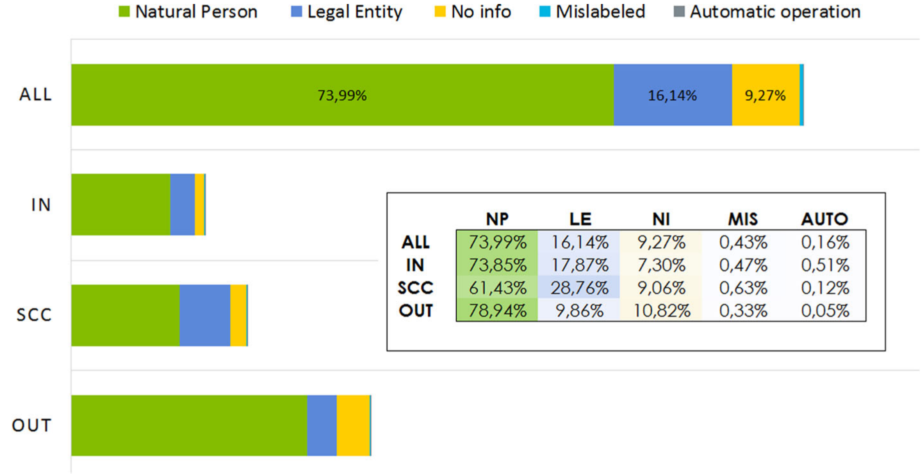
However, it is hard to account this global behaviour to the natural persons vs legal entities dynamic only, as lower degree nodes can be, to some extent, legal entities or mislabelled legal entities. In fact, despite of the disassortativity of the network, $\sim$ 18% of wire transfers are among legal entities only, accounting for $\sim$ 42% of the circulating money (see Fig. 2). This part of the network is opaque to basic network analysis, as it is not clear how businesses interact with each other: a deeper analysis is required.

It is necessary to zoom in the network's structure through a bow-tie analysis, by isolating parts of the wire transfers ecosystems and looking for a clearer characterisation of the largest connected component. Then, we exploit qualitative information as the business sector of a legal entity, when such data is available in the Customer Registry (see "Dataset description" section).

After executing a bow-tie analysis, we found the following components, whose sizes are reported in percentages w.r.t. the total number of nodes of the network: SCC (24.03%), IN (18.33%), OUT (40.84%), Tubes and Tendrils (9.11%), and Disconnected (7.68%).

In Fig. 12 we outline the sizes of the three main components (SCC, IN, and OUT) in comparison with the whole network, with internal node type distributions to emphasise the proportion between natural persons and legal entities.

OUT is the largest connected component by far: flows of money stops here, and they are not re-injected again in the Intesa Sanpaolo wire transfer network, as they are

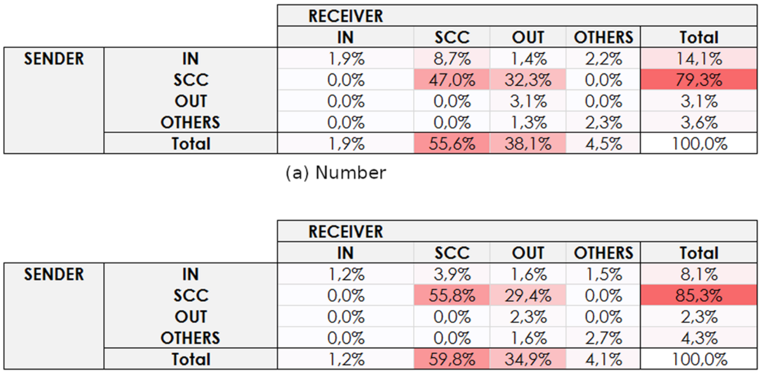


	NP	LE	NI	MIS	AUTO
ALL	73,99%	16,14%	9,27%	0,43%	0,16%
IN	73,85%	17,87%	7,30%	0,47%	0,51%
SCC	61,43%	28,76%	9,06%	0,63%	0,12%
OUT	78,94%	9,86%	10,82%	0,33%	0,05%

Fig. 12 Node type distribution in different bow-tie components: the length of the bars represents the size of the components respect to the whole network (ALL). Please note that a small percentage of accounts are labelled "No Info" because owners did not authorise analysis of their data; moreover, a very small portion of accounts have been mislabelled. Actual values are reported in the internal table, too

maybe spent as cash or via credit card; nodes in OUT can be seen as the ultimate targets of the majority of the streams in our graph. OUT consists to a large extent of natural persons (∼ 80%), and legal entities are under represented. Nevertheless, the dynamics within this component seem marginal in comparison with the whole network, because of a low level of engagement between its nodes: most of the payments involving nodes in OUT come from the central SCC in terms of numbers and amount (see Fig. 13). 56% of the wire transfers in OUT are settled among natural persons (see Additional materials D for all the percentages), meaning that the main activity for this kind of nodes is to receive money for rental purposes, housing expenses or among relatives (as discussed in "Insights from text analysis" section).

IN is the third largest component. As already observed for OUT, interactions within IN are still pretty limited both in number and in amount (see Fig. 13), since the majority of



		RECEIVER				
		IN	SCC	OUT	OTHERS	Total
SENDER	IN	1,9%	8,7%	1,4%	2,2%	14,1%
	SCC	0,0%	47,0%	32,3%	0,0%	79,3%
	OUT	0,0%	0,0%	3,1%	0,0%	3,1%
	OTHERS	0,0%	0,0%	1,3%	2,3%	3,6%
	Total	1,9%	55,6%	38,1%	4,5%	100,0%

(a) Number

		RECEIVER				
		IN	SCC	OUT	OTHERS	Total
SENDER	IN	1,2%	3,9%	1,6%	1,5%	8,1%
	SCC	0,0%	55,8%	29,4%	0,0%	85,3%
	OUT	0,0%	0,0%	2,3%	0,0%	2,3%
	OTHERS	0,0%	0,0%	1,6%	2,7%	4,3%
	Total	1,2%	59,8%	34,9%	4,1%	100,0%

(b) Amount

Fig. 13 Distribution of wire transfers by number (top) and amount (bottom): between bow-tie components. For each target principal components (columns) we split the contribution from all the source components (rows). The majority of wire transfers are between nodes in SCC (both by amount and number).

payments move forward directly to the SCC core, with a small percentage of transactions that are settled to nodes in OUT or "dispersed" in tubes and tendrils.

Finally, despite its relatively limited size in terms of nodes (24% of the whole network), *SCC* is definitely the core of our ecosystem: almost 80% of the wire transfers are settled from accounts in SCC, and they are responsible for more then 85% of the amount transferred (see Fig. 13). Quite intuitively, it is possible to guess that money spirals out inside this component (55.8% of overall amount is exchanged among nodes in SCC), and that wire transfers are entangled together creating a "hairball problem". Hence, it is hard to observe some kind of order emerging spontaneously from SCC, and we need a deeper analysis to better understand what happens between natural persons and legal entities, as well as between nodes with different degrees.

First of all, we observe that more than 43% of businesses (legal entities) are represented by nodes in the SCC core. Among those private companies, we find here the biggest ones; in fact, the top 500 hubs of the network are in SCC. In Fig. 14 we show the ratio of natural persons and legal entities at the receiving end of a wire transfer in SCC, when the sender is a natural person (top) or a legal entity (bottom). We can observe the same general trend, even if with very different magnitudes: lower degree nodes, regardless their type, tend to send money to legal entities, while higher degree nodes tend to send money to natural persons.

Now, we can try to combine some results we described in "Network analysis" section and what we understood so far from the bow-tie analysis. Many paths flow from nodes in IN, circulate in the SCC core, and ends in OUT. However, the majority of payments are settled inside SCC; therefore, it would be a huge mistake to locate principal wealth sources in IN. Instead, in SCC we observe again a complex natural person/legal entity interplay, where actors with similar degrees rarely interact with each other because of the disassortative trend we already discussed (Fig. 8). Moreover, we have that hubs are more likely to settle payments to natural persons, meaning that their main role is to pay salaries and retirement pensions (see "Insights from text analysis" section). Apparently, lower degree nodes behave differently, paying legal entities (mainly taxes, service fees), and partly natural persons (rentals, house expenses). In this latter case, lower degree nodes exhibits a more assortative behaviour (Fig. 11a), connecting the core of the network to tightly connected smaller communities. Recall also that the majority of nodes are natural persons, so the role of lower degree nodes is fundamental to maintain the connectivity of the network, especially SCC with the rest of the nodes (mainly in OUT).

On the other hand, the role of interactions among legal entities remains unclear, even if the bow-tie analysis shows that such transactions fuel the engine of the ecosystem. In the following paragraph we use information from business sectors to better understand those dynamics.

Grouping money flows by business sectors

To study the role of legal entities and to let emerge a signal of supply chains, from production of a commodity to the delivery of a service to final customers, we grouped money flows within and across the three main network components by business sectors. The Customer Registry provides information about a bank account's business sector. Although this is a coarse-grained information, it allows us to start distinguishing between the diverse activities characterising legal entities. This information is provided as ATECO

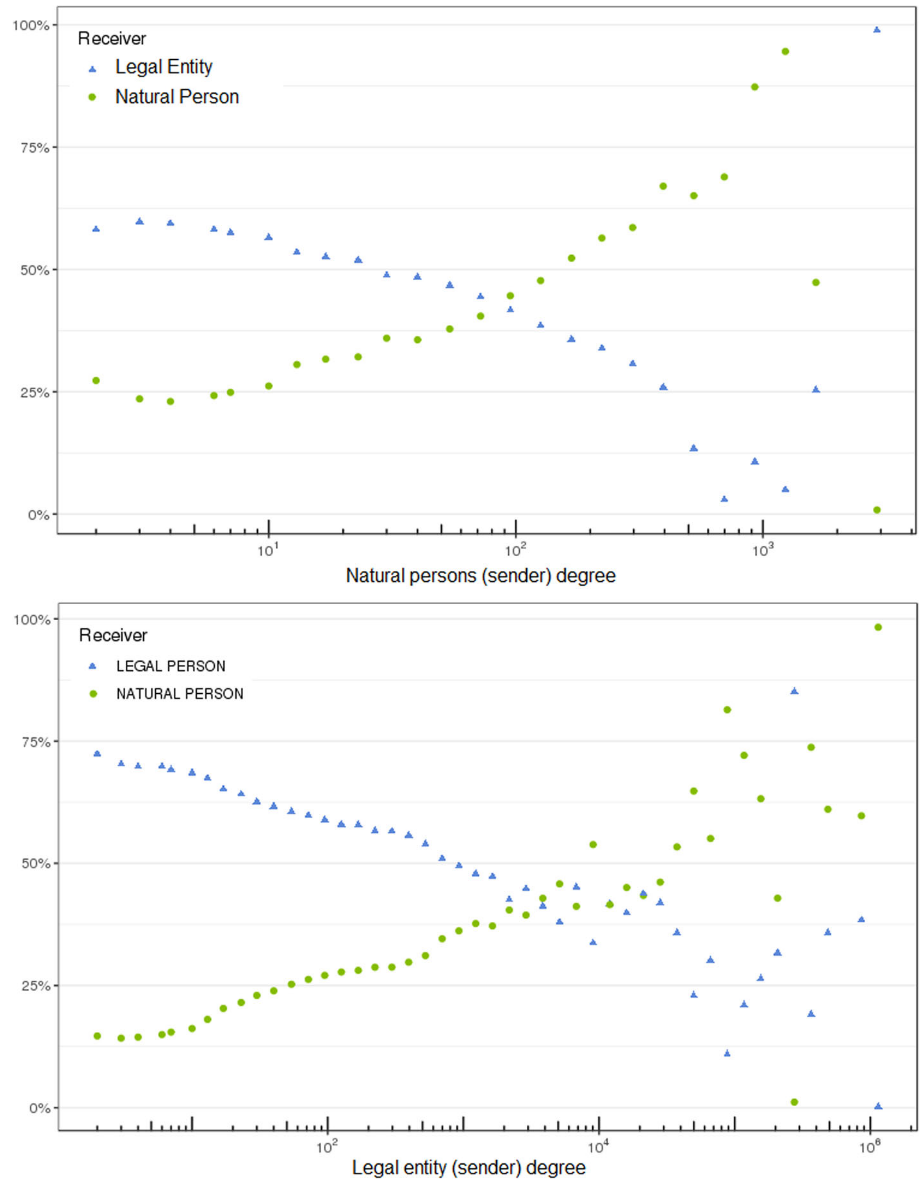


Fig. 14 Receiver type per sender type/degree in the SCC core: ratio of natural and legal receivers for natural (top) or legal (bottom) sender, according to their degree. Lower degree nodes tend to send money to legal entities, regardless their type, while higher degree nodes tend to send money to natural persons

codes, as mentioned in "Dataset description" section. Since ATECO classification is the national adaptation of the European nomenclature Nomenclature statistique des activités économiques dans la Communauté européenne (NACE), we preferred to refer to the international one, which we then grouped as reported in Table 1. This grouping reduces the number of the possible labels a given node may have, producing non-overlapping business areas.

The arc diagrams in Figs. 15 and 16 show the wire transfers total amount flowing within and between the three principal components of the graph between natural persons and legal entities. The latters are grouped by their NACE code in both figures.

Table 1 Conversion table between NACE codes and groups used in this work

Codes	Group
A	Agriculture and Farming
B	Mining
C	Industry
D, E	Services providers
F, L	Real Estate
G	Commerce
H	Transportation
I, N	Tourism
J, M, S	Professional Services
K	Finance
O	Public Administration
P	Education
Q	Health and Social Services
R, T, U	Other

Directions of payments are clock-wise: arcs on the right of the diagram are meant to be read from top to bottom, while arcs on the left are interpreted the other way around. Nodes are sized after their number of occurrences in the component, while edges are thicker proportionally to wire transfers total amount. Self-loops have been removed from the diagram.

First, it is possible to see how the general activity within nodes in IN and OUT is generally lower than within the SCC core, where the top players of this network are. Second, amounts flowing from a category can be disproportionate to the number of nodes in that sector: for example, focusing on SCC, it is striking that despite their numbers, natural persons send back to legal entities in the network only a small fractions of money they receive. Moreover, legal entities in SCC operating in "finance" send a huge amount to natural persons despite their size. A similar pattern, with a different order of magnitude, could be easily observed between "commerce" and "industry". Notice also that natural persons are the receivers of the widest streams in SCC. The only exception is represented by "commerce" companies that is more likely to settle payments to "industry" than to natural persons. This behaviour is not dependent on the large number of natural persons itself, as the number of the nodes into a category is not correlated to the amount of wire transfers settled, as already noted above. Plus, natural persons' role as main collectors does not emerge clearly in IN nor in OUT, where natural persons are predominant as well. Removing self-loops from this analysis has no effect on the visualisation: the only significant ones are among natural persons of the three components.

Figure 16 shows wire transfers amounts between each of the components. Again, there is a wide flow streaming down from "finance" in SCC to natural persons in OUT. Another important money cascade is the one from "professional services" in SCC to the same category in OUT. Even if not visible from the arc diagram, the reader should be aware that the majority of wire transfers from SCC to OUT are settled toward natural persons (76.5%). However, if we consider amounts of transactions instead of their number, natural persons appear several times less than legal entities: the percentage decreases dramatically

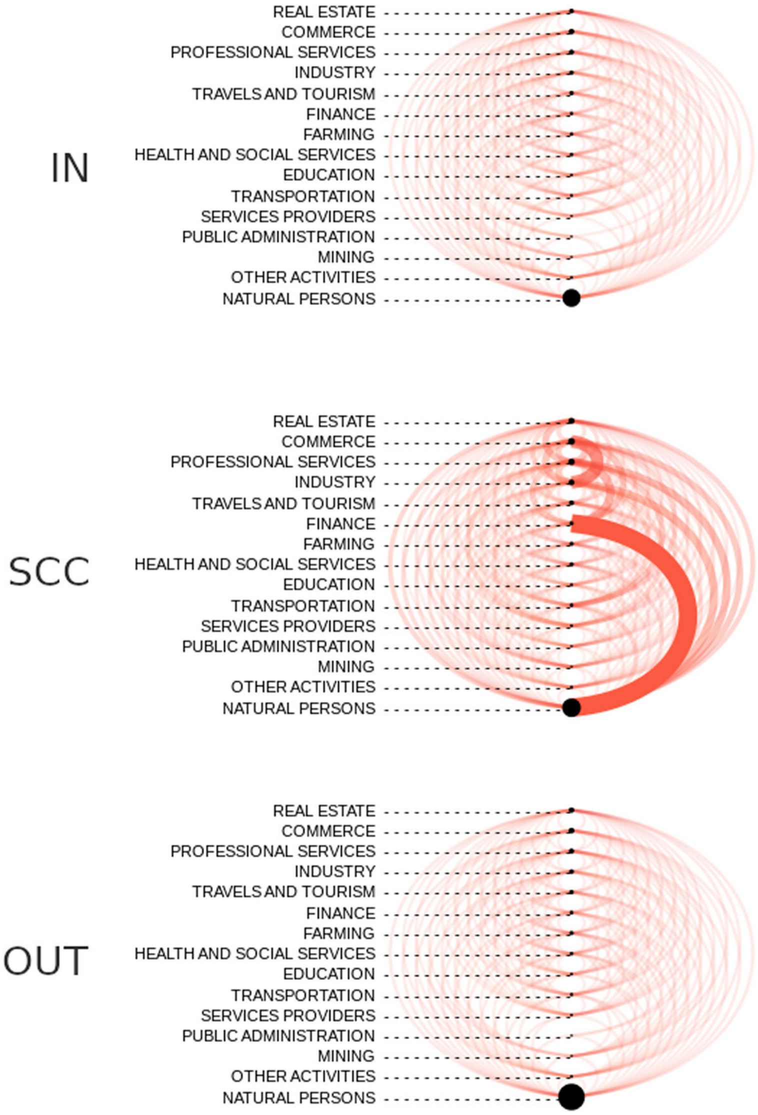


Fig. 15 Wire transfer within components: Total amount of wire transfers between nodes in the three components (IN, SCC and OUT). Legal persons are classified by their business sectors. The size of the points representing the sectors is proportional to the number of the nodes in each component referring to a given category. Directions of payments are clock-wise

to 17.3%. See figures in tabular form in the Additional materials D for actual percentages between all the categories in SCC and OUT.

Inequalities and specialisations at a regional basis

So far, we grouped different money streams by business sectors to better highlight what happens within and across the three main bow-tie components. Now, in order to link dynamics between nodes in the network to official economic data so that local patterns can be identified, we analyse activities emerging from wire transfers on a geographical basis. We show that (i) the amount and the variety of payments made by both natural persons and legal entities change substantially on a regional basis, reflecting well known

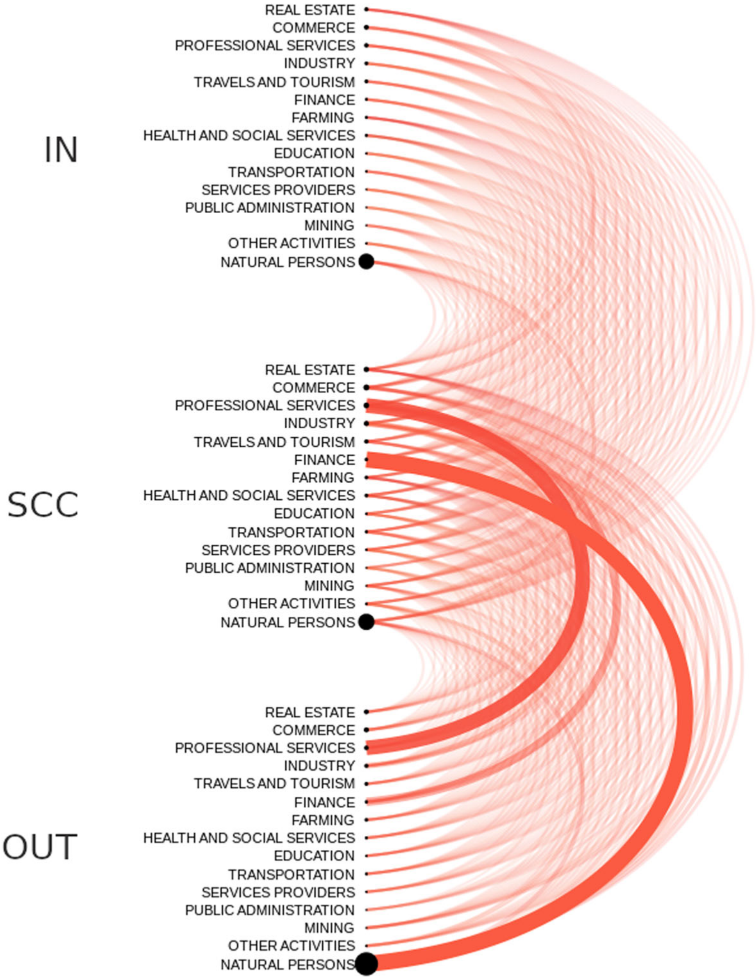
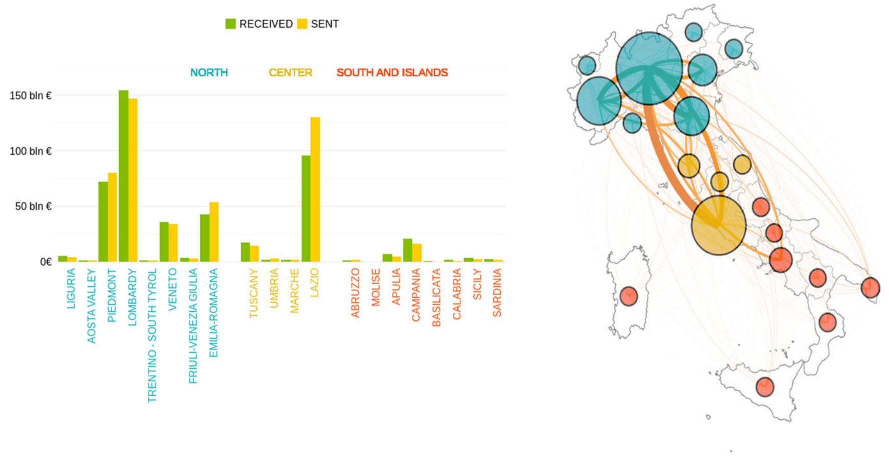


Fig. 16 Wire transfers between components: Total amount of wire transfers between natural persons and legal entities belonging to different components. Legal entities are classified separately by their business sectors. Directions should be read from top to bottom

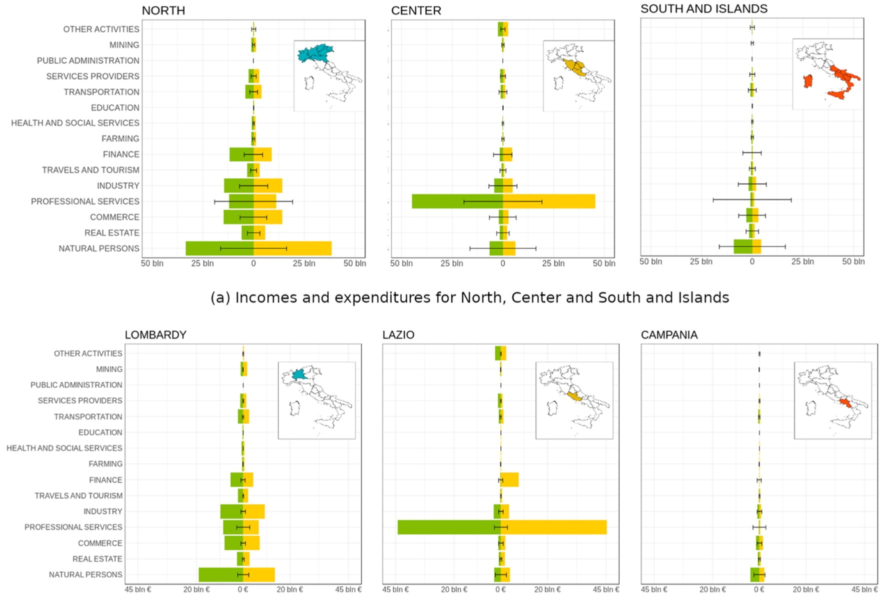
inequalities in the country (see Fig. 17a and b); (ii) combining information about the business sector and the geographical location of bank accounts can effectively give a support to spot possible supply chains (see Fig. 18a and b).

As reported in "Research case's description" section, like many other European countries, also Italy is characterised by large inequalities in the distribution of wealth and income. We also observed that the GDP per capita of the whole South Italy in 2017 is 45% lower than the Centre-North GDP per capita: these numbers are consistent with our findings in the ISP dataset: the total amount of wire transfers received by natural persons in Southern region is 44.6% lower than in Centre-North. We also recall that we found a high positive correlation between regional GDP distribution and both numbers and amounts of wire transfers grouped by regions (see the correlations scatter plot matrix in Fig. 3 back in "Dataset description" section). Therefore, hereafter we did not normalise amounts by regional population or ISP market share, because the distribution in itself is quite representative of the GDP; for example, Campania and Sicily are respectively the third and the



(a) by region (b) between regions

Fig. 17 Wire transfers total amount by (left) and between (right) regions: Northern bank accounts are on average richer than Southern bank accounts. Lombardy and Lazio emerge as the main economic centres of the country (confirmed by the official GDP per capita distribution, as in Fig. 3). We show also exchanges between regions: the width of each edge is proportional to the total amount transferred from one region to another, with directions to be read clockwise; each node's size represents out-strength of the related region, and the geographical area it belongs to. Richer Northern regions are the most interconnected and wider flows are exchanged between Lombardy, Piedmont, Veneto, Emilia-Romagna and Lazio (the only region from the Centre belonging to the "richest club")



(a) Incomes and expenditures for North, Center and South and Islands

(b) Incomes and expenditures for Lombardy, Lazio and Campania

Fig. 18 Incomes and expenditures distributions divided by NACE codes and grouped by geographical area (top) and sample regions (bottom). Incoming and outgoing amounts are coloured are coloured respectively in green and in yellow. Black ticks show the averages for the given business sector. Northern regions outperform in the majority of sectors, especially in industry, commerce, and finance. Regions in the Centre are particularly focused on professional services. Conversely, the South and Islands under perform in every NACE sectors. Also natural persons' behaviour is quite unequal, confirming the trend. Switching to regions, Lombardy is confirmed as the leading region in all the sectors, but professional services that are mainly centred in Lazio. Although Campania is the most active in the South, it still performs under the regional average in all sectors, but commerce. It is also striking that Campania receives more money from the other regions than what it sends

fourth most populated Italian regions, but this is not reflected in their GDP, nor in their wire transfers activity.

Hence, in this section, we focus on the total amount of incoming and outgoing wire transfers grouped by the 20 Italian regions, divided in three macro-areas: North, Centre and South and Islands. Such information is summarised in Fig. 17a. Additionally, Fig. 17b shows the aggregate information in form of a graph, to better highlight interactions between pairs of regions. The width of each edge is proportional to total amount transferred from one region to another, the colour and the size of each node represents respectively the out-strength and the in-strength of the related region. The full table of amounts shared between regions is available in Additional materials E.

Generally speaking, Southern regions send/receive less money to/from the rest of the country. This can be due to several factors, like lower GDP per capita and salaries, but also some other factors harder to catch: for example, as stated in the ISTAT report, underground economy is significantly higher in South Italy (Report ISTAT 2017). Lombardy, Piedmont, Veneto, Emilia-Romagna and Lazio are clear outliers in this distribution: regions in the "richest club" are tightly interconnected with each other and wider streams of money flow in, out, and within these regions. The only region from South that has a not neglectable exchange with the richest club is Campania.

If we want to look for specialisations within an area or a region we have to split the above described information by NACE business sectors. In Fig. 18a we show the distribution of different incomes and expenditures for each business sector, grouped by North, Centre, and South and Islands. Incoming and outgoing amounts are coloured respectively in green and in yellow. Values are compared to the averages (black ticks) for the given business sector. Northern regions outperforms in the majority of sectors, especially in industry, commerce, and finance. Regions in the Centre are particularly focused on professional services. Conversely, the South and islands under perform in every NACE sectors. Also natural persons' behaviour is quite unequal, confirming the trend.

In Fig. 18b we have a finer-grained visualisation of business sectors total incomes and expenditures distributions for three specific regions: Lombardy (North), Lazio (Centre), and Campania (South). Lombardy is confirmed as the leading region in all the sectors, but professional services that are mainly centred in Lazio. Although Campania is the most active in the South, it still performs under the regional average in all sectors, but commerce. It is also striking that Campania receives more money from the other regions than what it sends, that looks like a quite distinctive behaviour in the South. This could also be connected to the regional inequality and the assortative behaviour we observed among natural persons: we know that natural persons (i.e., low degree nodes) are more likely to exchange money among themselves, within inner social circles and families. We also understood that the companies that fuel the nation engine are more likely located in Norther regions and in Lazio. As a consequence, there is a chance that we have a signal that southern people working for some company or organisation located in the North or in Lazio, are in charge for family's expenses, and house related fees. Although this hypothesis relies on common sense, we did not validate it properly, as we plan to do as future work. For a small multiple visualisation of all the 20 regions, see figures enclosed in the Additional materials F.

Regional inequalities emerge clearly from our data, and as recalled above they are good predictors of the actual GDP per capita distribution. Also, we observed how different regions may manifest specialisations in one or more business sectors. To wrap up some of the conclusions we have drawn so far, just recall that in the previous section we observed that the bow-tie analysis is able to capture money streams that start flowing partially from IN, are reinforced and keep circulating in SCC, and then drain in the OUT outfall. Grouping such streams by business sectors, we have high level depiction of supply chains that emerge from the bow-tie components; in fact, we have commerce actors that buy services from industry, and finance organisations that pay mainly natural persons in the SCC core as well in OUT. Quite interestingly, professional services companies in SCC streams into the finance sector in SCC and in OUT, but also widely into other professional services operating in OUT: it is like that the professional services sector is made of companies of different sizes that pay each other in cascade, probably belonging to a inherently interdependent supply chain. We already observed a signal of a hierarchical structure in our wire transfer network. Such a signal seems to hold especially among the professional services supply chain; in fact, 85% of wire transfers from professional services in SCC to professional services in OUT are settled from a higher degree company to a lower degree actor, with a substantial part of that money ending up in the accounts owned by natural persons: 65% of wire transfers of professional services in OUT are set toward natural persons. Finally, it is evident that the professional services core is localised in Lazio (especially in Rome), whereas finance and industry most important actors are located in the North, especially in Lombardy. Such interdependence between regions, business sectors, companies belonging to different supply chains and finally natural persons is hardly to be explored by means of local data analysis only; conversely, networks provide a terrific tool to make sense of the complexity of wire transfers between individuals under a comprehensive framework that connects global phenomena to local patterns.

Final remarks and conclusions

The main goal of our work was to explore the connections between structural network inequalities and bank's customer spending behaviours, within an entire national ecosystem made of natural persons and legal entities, different business sectors, and supply chains that span distinct geographical regions. We had the opportunity of analysing the whole set of wire transfers data recorded in 24 months by Intesa Sanpaolo, a leading banking group in the Eurozone and also the most important in Italy in terms of market share. Hence, this study provides insights on how to capture, understand, and predict emerging phenomena in a nation-wide complex economic system such the Italian one.

Also in this domain, networks have proven to be exceptionally good at linking global complex phenomena to local patterns. In the introduction we used the analogy with the "globe": as the spherical model of the Earth, networks allow for both latitudinal and longitudinal explorations of our wire transfers dataset, showing that structural inequalities can be a proxy for differences in customers' spending behaviours according their node's degree, a tool for unfolding patterns in business sectors across network's components and geographic regions, and also a model for predicting other economic values, such as the regional GDP per capita heterogeneous distribution.

In the rest of this section, we recall the highlights of our empirical findings, we outline some of the known weaknesses of this research, and then some future work than can be done on the ISP wire transfer dataset.

Empirical analyses highlights

As in a globe, you can have an overview of the ecosystem finding emerging phenomena, and then take a closer look to better understand the underlying interactions. In this perspective, one of the main interesting characteristics of network theory is that graph based models are able to predict many different characteristics and general patterns. For example, it is not much surprising that we found a clear heavy tailed behaviour in the network's degree distribution: hubs are big companies or organisations and the majority of low degree nodes are natural persons. Nevertheless, when some facts do not match expectations, we can look for much more interesting clues:

(i) The GCC does not emerge by phase transition when the average degree $\langle k \rangle > 1$; on the contrary, it evolves continuously around the same core of the top 1% highest degree nodes, that are also responsible for the robustness of the entire network.

(ii) The network is clearly disassortative, but small degree nodes buck the trend; in fact, in the same picture we have the interplay between a strong-hierarchical behaviour, with higher degree nodes (big companies or organisations) that pay salaries to natural persons or services to lower degree businesses, and also a flatter scenario of smaller and tightly clustered communities, composed mainly by natural persons, that share money among them, likely to contribute to family expenses; in fact, the reasons for such payments are mainly due to generic housing expenses, including rent, condominium fees, mortgage payments.

(iii) The mixed behaviour recalled above seems to lead to a hierarchical structure in which there would be a general direction of money flowing down from higher to lower degree nodes, but for a $\sim$ 30% of transactions that move upstream.

(iv) Even if the majority of the nodes of the network correspond to natural persons, the larger amount of money are exchanged with and between legal entities, that also are more likely to have higher degree values.

These preliminary findings, motivated us to perform a closer inspection at the interactions among legal entities grouped by business sectors and by geographic regions. This turned out to be a useful tool for a deeper understanding of the interplay between different supply chains; in fact, we figured out that the hierarchical structure of our network that makes money to flow down from higher degree nodes to lower degree players, spiral within the SCC core, and then drain into a wide OUT lake, where many small companies and natural persons are ($\sim$ 41% of the nodes). We found a signal that companies operating in "commerce" and "industry", located mainly in the North (especially in Lombardy) are able to fuel the ecosystem's economic engine, together with actors offering "professional services", that from the headquarters in Lazio, show a more vertical extension between different components. Such structural inequalities and specialisations by regions are also good predictors of an official economic measure, such as the GDP per capita.

Weaknesses of the research

All the analyses presented in this paper are based on the customers and the wire transfers data provided by ISP. This dataset is very rich and informative, and allowed us to get several empirical insights on a big and complex macro-economic system, such as a relevant part of the Italian population and industry. At the same time, for the sake of clarity, we must acknowledge some limitations of the data itself, to let future research to deal with such weaknesses of the approach.

First, this data is limited to Intesa Sanpaolo customers only. We excluded from the analysis all the wire transfers incoming/outgoing from/to other banks' customers, as we could not retrieve any information about the other endpoint of such payments; including other banks' accounts as nodes in our graph would have distorted our comprehension of the network itself, because we would have introduced millions of nodes whose payments history is, for the vast majority, hidden to us.

Second, the dataset encompasses wire transfers only. Other channels of payment, like credit cards or cash, are excluded from our observation. This has the effect of making us focus on a determined money transfers, like rents or salaries, while neglecting or underestimating other kinds of payments, e.g., those due to shopping; it is likely that this bias may have let us misinterpret the importance of some business sectors, like commerce. Plus, this caveat affects, in principle, also our bow-tie analysis: for instance, nodes found in OUT may be in fact customers that prefer to spend their money via credit card transactions.

Third, in a few occasions we noticed that a node's official labelling of natural person or legal entity did not match with the spending behaviour of the node. As discussed in "Dataset description" section, this label is applied to bank accounts when they are opened, and it may happen that, even years later, some customers start using their accounts in a different way, without changing their legal definition accordingly. It is the case, for instance, of customers that start a small business as free-lancers and they run it with the same account they used until then for personal savings only. This discrepancy between the legal definition of an account and its real use led in a few cases to some outliers. However, the actual quantity of this mislabelled persons is an open issue.

Finally, not all customers gave their consent to analyse their personal information: in application of the GDPR normative, accounts' owners have the right to their consent. This led to a not neglectable quota of accounts we labelled as "no info", as we cannot check whether they were, for example, natural persons or legal entities. However, it should be noticed that information on wire transfers coming to/from those customers, as this kind of records is a trace of a service provided by the bank, and as a consequence it belongs to the bank and can be analysed.

Future work

First of all, it would be quite interesting to acquire a finer-grained knowledge on supply chains, exploring sub-sectors and also municipalities, with the objective of predicting the rise and the fall of locally clustered economic districts, if datasets with wider time intervals will be made available. Additionally, the access to other forms of payments, such as credit cards, and the fusion with similar datasets from other banks, can return a very precise and exploitable picture of a national and also international financial system, solving one of the main weaknesses of the research that we discussed above.

Then, we think that a deeper robustness analysis of the GCC should be executed, extending the results mentioned at the end of "Hubs and connectivity" section: different node removal strategies can be compared in order to identify the nodes and the edges that actually connect the wire transfer network; moreover, the vulnerability of the network must be interpreted in the context of the Italian financial system or its geography.

Another problem that we plan as future work is to understand the outliers' behaviours: there are high degree natural persons (with thousands of wire transfers sent or received) and also very low degree legal persons. If this is due to mislabelling in data or if there are other reasons behind their existence, may help to unveil some characteristic of the national financial system difficult to be investigated otherwise.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1007/s41109-020-00314-x.

Additional file 1: Additional materials a: gCC and robustness. This appendix contains further investigations on the GCC evolution and also some preliminary results on a robustness attack we performed in our wire transfer network. These experiments are mentioned in "Hubs and connectivity" section. We also put here a what-if analysis on a projection of the network with natural persons only.

Additional file 2: Additional materials b: insights from text analysis. This appendix describes the text analysis that has been performed on the purpose of the payment field of each wire transfer transaction. The outcomes of such experiment are briefly presented in "Insights from text analysis" section.

Additional file 3: Additional materials c: assortativity. This appendix contains further investigations about the results presented in "A network with mixed assortative/disassortative behaviour" section, such as a knnk variant computed as the median of the degrees of the neighbors of each node of degree k, and a density heatmap as the one in Fig. 10 but compared with a null model, and a density heatmap based on the total amounts shared depending on the degree of receivers and senders.

Additional file 4: Additional materials d: Bow-tie Main components. This appendix contains in tabular forms all the values used to produce plots and diagrams shown in "Inter- and intra-components flows" section.

Additional file 5: Additional materials e: wire transfers total amount between regions. This appendix contains one figure in tabular form showing total amount sent/received between regions grouped by North, Centre, South and Islands. These values have been used to produce plots in "Inequalities and specialisations at a regional basis" section.

Additional file 6: Additional materials f: distribution of incomes and expenditures for all the italian regions by sectors. This appendix contains three figures showing incomes and expenditures distributions divided by NACE codes for all the Italian regions, completing information provided for only three regions in "Inequalities and specialisations at a regional basis" section, Fig. 18.

Acknowledgements

The authors would like to acknowledge Roberta Sinatra, Alessandro Flammini and Filippo Menczer for feedback and insightful conversations on preliminary analyses we performed on this dataset. The authors would like also to thank Luigi Ruggerone, Arianna Miola and Valeria Ricci (Intesa Sanpaolo Innovation Center) for organisational support of the project and stimulating discussions, Marco Lamieri (ISP Risk Management) and Ilaria Sangalli (ISP Research Department) for helpful comments in a previous version of this manuscript. Last but not least, we thank the anonymous reviewers for their careful reading of our manuscript and their profound observations and stimulating recommendations, not to mention the additional analyses they suggested to perform that improved the quality of this paper substantially: they would have deserved to appear as co-authors, if we only knew their identities.

Authors' contributions

Conceptualisation: Giancarlo Ruffo. *Data curation:* Alfonso Semeraro, Silvia Ronchiadin. *Formal analysis:* Alfonso Semeraro, Marcella Tambuscio, Silvia Ronchiadin, Giancarlo Ruffo. *Methodology:* Giancarlo Ruffo, Marcella Tambuscio, Alfonso Semeraro. *Validation:* Silvia Ronchiadin, Laura Li Puma. *Visualisation:* Alfonso Semeraro, Silvia Ronchiadin. *Writing - original draft:* Alfonso Semeraro, Giancarlo Ruffo, Silvia Ronchiadin, Marcella Tambuscio. *Writing - review & editing:* Alfonso Semeraro, Marcella Tambuscio, Silvia Ronchiadin, Laura Li Puma, Giancarlo Ruffo. The author(s) read and approved the final manuscript.

Funding

The authors from Università degli Studi di Torino acknowledge financial support from Intesa Sanpaolo Innovation Center under a cooperative agreement between the two institutions.

Availability of data and materials

The official statistics are available through Istituto Nazionale di Statistica (https://www.istat.it).The data that support the findings of this study are available from Intesa Sanpaolo but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Intesa Sanpaolo.

Competing interests

The authors declare that they have no competing interests. The authors of this manuscript have read the journal's policy and have the following competing interests: Silvia Ronchiadin and Laura Li Puma are employed at Intesa Sanpaolo Innovation Center; however, this industrial affiliation does not alter the authors' adherence to ANS policies on sharing data and materials.

Author details
[1]Dipartimento di Informatica, Università degli Studi di Torino, Turin, Italy. [2]Austrian Center for Digital Humanities, Austrian Academy of Sciences, Vienna, Austria. [3]Intesa Sanpaolo Innovation Center, Turin, Italy.

Received: 1 March 2020 Accepted: 31 August 2020
Published online: 13 October 2020

References

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A (2015) Systemic risk and stability in financial networks. Am Econ Rev 105(2):564–608
Amini H, Cont R, Minca A (2016) Resilience to contagion in financial networks. Mathematical finance 26(2):329–365
Bank of Italy (2019) Payment System Statistics. https://www.bancaditalia.it/pubblicazioni/sistema-pagamenti/index.html. Accessed 27 Jan 2020
Barabási A-L, Bonabeau E (2003) Scale-free networks. Sci Am 288(5):60–69
Barabási A-L, et al (2016) Network Science. Cambridge University Press, Cambridge
Battiston S, Caldarelli G, May RM, Roukny T, Stiglitz JE (2016) The price of complexity in financial networks. Proc Natl Acad Sci 113(36):10031–10036
Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G (2012) Debtrank: Too central to fail? financial networks, the fed and systemic risk. Sci Rep 2:541
Beiró MG, Bravo L, Caro D, Cattuto C, Ferres L, Graells-Garrido E (2018) Shopping mall attraction and social mixing at a city scale. EPJ Data Sci 7(28):1–21
Boss M, Elsinger H, Summer M, Thurner 4 S (2004) Network topology of the interbank market. Quant Finan 4(6):677–684
Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A, Wiener J (2000) Graph structure in the web. Comput Netw 33(1-6):309–320
Caccioli F, Barucca P, Kobayashi T (2018) Network models of financial systemic risk: A review. J Comput Soc Sci 1(1):81–114
Caldarelli G, Chessa A (2016) Data Science and Complex Networks: Real Case Studies with Python. Oxford Scholarship Online, Oxford
Caldarelli G, Chessa A, Pammolli F, Gabrielli A, Puliga M (2013) Reconstructing a credit network. Nat Phys 9(3):125–126
Ciani E, Torrini T (2019) The geography of Italian income inequality: recent trends and the role of employment. Occas Papers (Questioni di Economia e Finanza) 492:173–208
Colladon AF, Remondi E (2017) Using social network analysis to prevent money laundering. Expert Syst Appl 67:49–58
Didimo W, Liotta G, Montecchiani F, Palladino P (2011) An advanced network visualization system for financial crime detection. In: 2011 IEEE Pacific Visualization Symposium. IEEE, Hong Kong. pp 203–210
Dong X, Suhara Y, Bozkaya B, Singh VK, Lepri B, Pentland A (2018) Social bridges in urban purchase behavior. ACM Trans Intell Syst Technol (TIST) 9(3):33
Dreżewski R, Sepielak J, Filipkowski W (2015) The application of social network analysis algorithms in a system supporting money laundering detection. Inf Sci 295:18–32
Elliott M, Golub B, Jackson MO (2014) Financial networks and contagion. Am Econ Rev 104(10):3115–53
Erdös P, Rényi A (1959) On random graphs i. Publ Math Debr 6:290
Gross Domestic Product (GDP) at Current Market Prices by NUTS 2 Regions (2019). http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_10r_2gdp&lang=en. Accessed 27 Jan 2020
Haldane AG, May RM (2011) Systemic risk in banking ecosystems. Nature 469(7330):351
Inaoka H, Ninomiya T, Taniguchi K, Shimizu T, Takayasu H, et al (2004) Fractal network derived from banking transaction–an analysis of network structures formed by financial institutions. Bank Jpn Work Pap 4:1–32
Intesa Sanpaolo (2020) Italian Leader with a European Scale. https://group.intesasanpaolo.com/content/dam/portalgroup/repository-documenti/investor-relations/Contenuti/RISORSE/Documenti%20Pdf/en_gruppo/Brochure_istituz_uk.pdf. Accessed 27 Jan 2020
Leo Y, Fleury E, Alvarez-Hamelin JI, Sarraute C, Karsai M (2016) Socioeconomic correlations and stratification in social-communication networks. J R Soc Interface 13(125):20160598
Lublóy A (2006) Topology of the hungarian large-value transfer system. Tech Rep 57:7–38
Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296(5569):910–913
May RM, Arinaminpathy N (2009) Systemic risk: the dynamics of model banking systems. J R Soc Interface 7(46):823–838
Newman ME (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci 98(2):404–409
Newman M (2010) Networks: an Introduction. Oxford University Press, Oxford
Pozzi F, Di Matteo T, Aste T (2013) Spread of risk across financial markets: better to invest in the peripheries. Sci Rep 3:1665
Pugliese E, Cimini G, Patelli A, Zaccaria A, Pietronero L, Gabrielli A (2019) Unfolding the innovation system for the development of countries: co-evolution of science, technology and production. Sci Rep 9:16440
Ravasz E, Barabási A-L (2003) Hierarchical organization in complex networks. Phys Rev E 67:026112
Regulation (EU) 2016/679 of the European Parliament and of the Council (2016) Official Journal of the European Union, L 119/1. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN. Accessed 27 Jan 2020
Report ISTAT (2017) Anno 2017 Conti Economici Territoriali. https://www.istat.it/it/files//2018/12/Report_Conti-regionali_2017.pdf. Accessed 27 Jan 2020
San Pedro J, Proserpio D, Oliver N (2015) Mobiscore: towards universal credit scoring from mobile phone data. In: International Conference on User Modeling, Adaptation, and Personalization. Springer, Cham. pp 195–207
Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A, White DR (2009) Economic networks: The new challenges. Science 325(5939):422–5
Singh VK, Bozkaya B, Pentland A (2015) Money walks: implicit mobility behavior and financial well-being. PLoS ONE 10(8):0136628
Singh VK, Freeman L, Lepri B, Pentland AS (2013) Predicting spending behavior using socio-mobile features. In: Social Computing (SocialCom), 2013 International Conference On. IEEE, Alexandria. pp 174–179

Sobolevsky S, Bojic I, Belyi A, Sitko I, Hawelka B, Arias JM, Ratti C (2015) Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity. In: 2015 IEEE International Congress on Big Data. IEEE, New York. pp 600–607

Sobolevsky S, Massaro E, Bojic I, Arias JM, Ratti C (2017) Predicting regional economic indices using big data of individual bank card transactions. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE, Boston. pp 1313–1318

Soramäki K, Bech ML, Arnold J, Glass RJ, Beyeler WE (2007) The topology of interbank payment flows. Phys A Stat Mech Appl 379(1):317–333

Statistiche Demografiche ISTAT (2017) Data Base on Line. http://demo.istat.it/pop2017/index.html. Accessed 27 Jan 2020

Wagner A (2003) How the global structure of protein interaction networks evolves. Proc R Soc London Ser B Biol Sci 270(1514):457–466

World Economic Outlook Database (2018) International Monetary Fund, IMF.org. https://tinyurl.com/y2wmcnes. Accessed 27 Jan 2020

Zhang Y, Pennacchiotti M (2013) Predicting purchase behaviors from social media. In: Proceedings of the 22nd International Conference on World Wide Web. ACM, Rio de Janeiro. pp 1521–1532

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.