Properties of B2B invoice graphs and detection of structures

Joannès Guichon[1] Nazim Fatès[1] Sylvain Contassot-Vivier[1] Massimo Amato[2]

[1] Université de Lorraine, CNRS, INRIA, LORIA, F-54000, Nancy, France

[2] Bocconi University, Milan, Italy

joannes.guichon@inria.fr

Problem description

In Economy, a major issue is the potential lack of liquidity for settling the debts generated by payment delays among companies. Since this lack may trigger cascading failures, we analyze the interconnection of debts. Settling debts means lowering the systemic risks. The structure of such debt networks is not currently well known.

We analyze the structure of B2B invoice graphs. In particular, we address the possibility to identify relevant communities in such networks.

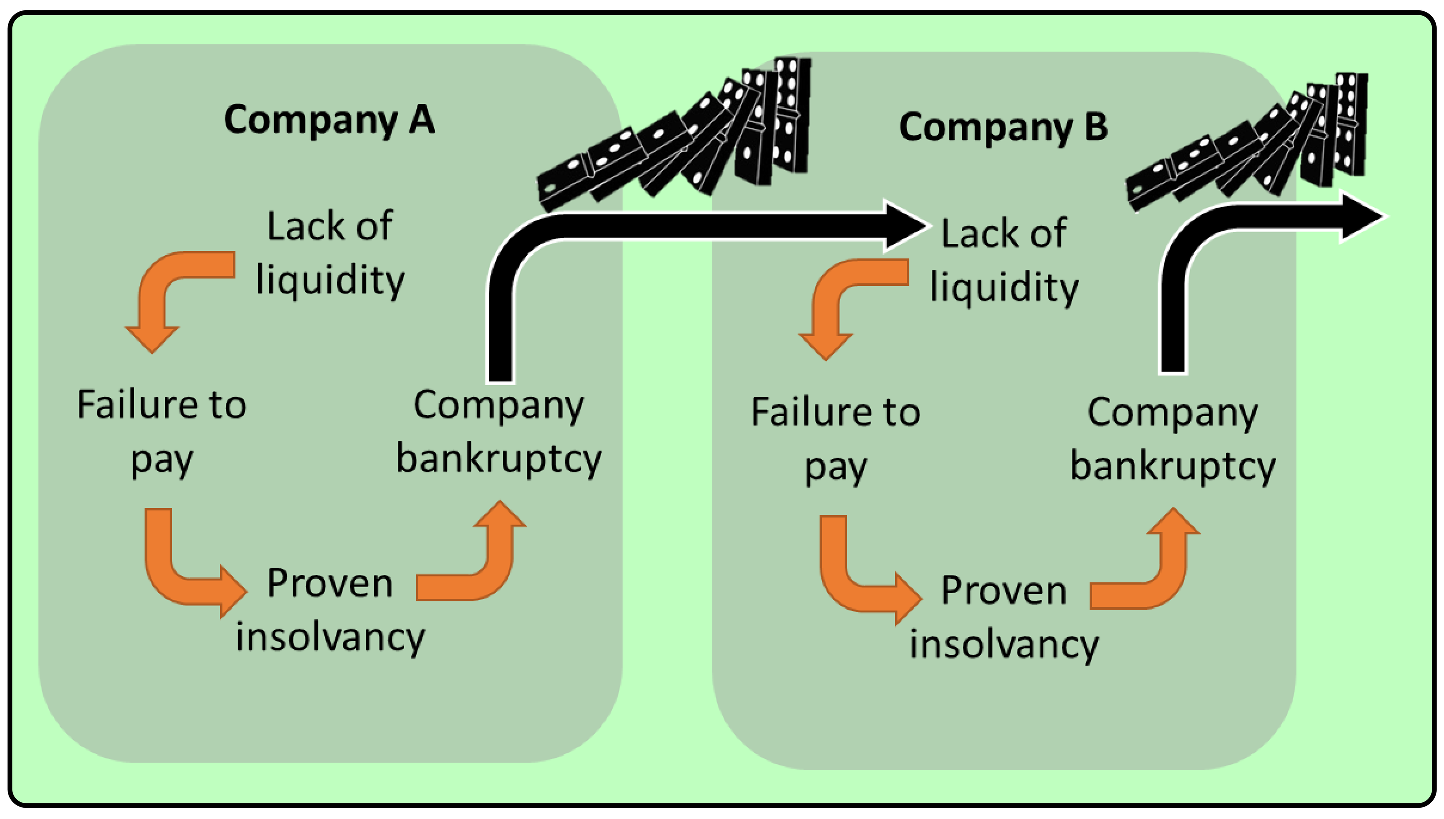
Context and modeling

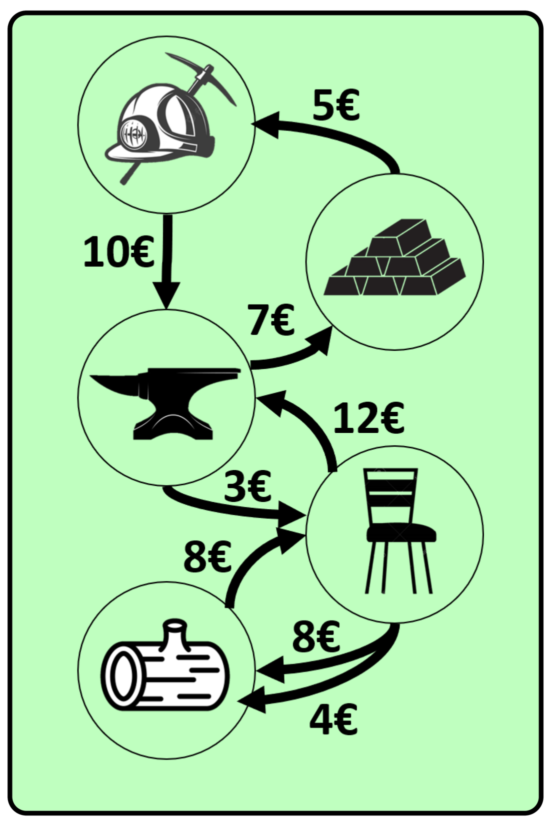
Delays in payments have a huge financial effect on companies, at the very least by increasing their financial exposure, at worst by inducing a **snowball effect of company failures**.

Analysis of a large dataset from Infocert[1] that contains 27 million invoices between companies (B2B activity) over a year.

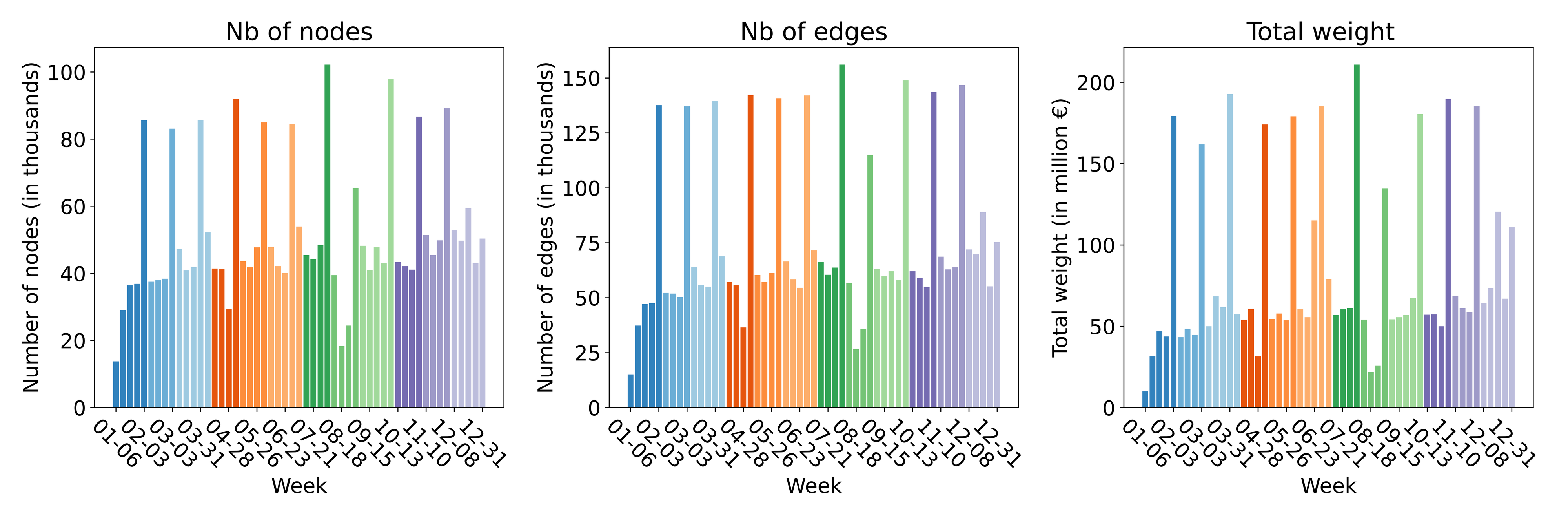
The aggregation of invoices forms **time-based directed and weighted multi-graphs**. Companies are the vertices and invoices are the edges.

Invoices are created and paid in a dynamic way on a daily basis. However, the debt graphs we consider are defined over a large enough period of time to obtain a sufficiently connected graph. This then allows us to reduce the economic pressure on vertices by suppressing edges.





Weekly variations



Comparative study of temporal decompositions

Temporal analysis is done daily, weekly and monthly:

- January and August are months with less activity
- Concentration of exchanges on the last day of each month, putting high workload on month-end weeks.

Statistical study on monthly graphs

- Presence of a giant strongly connected component every month: around 35% of the edges and 8% of the vertices
- Mean degree: 3.23
- Clustering cœfficient under 0.02, meaning many nodes lowly linked
- Mean diameter: 29 and average path length: 9.45

Degree distribution over months

Comparison of degree distribution between our multi-graph and classical graphs.

The distribution is close to the power-law $P(k) = k^{-\gamma}$ of scale-free graphs.

If we extract communities from such graphs, we are likely to keep the scale-free aspect.

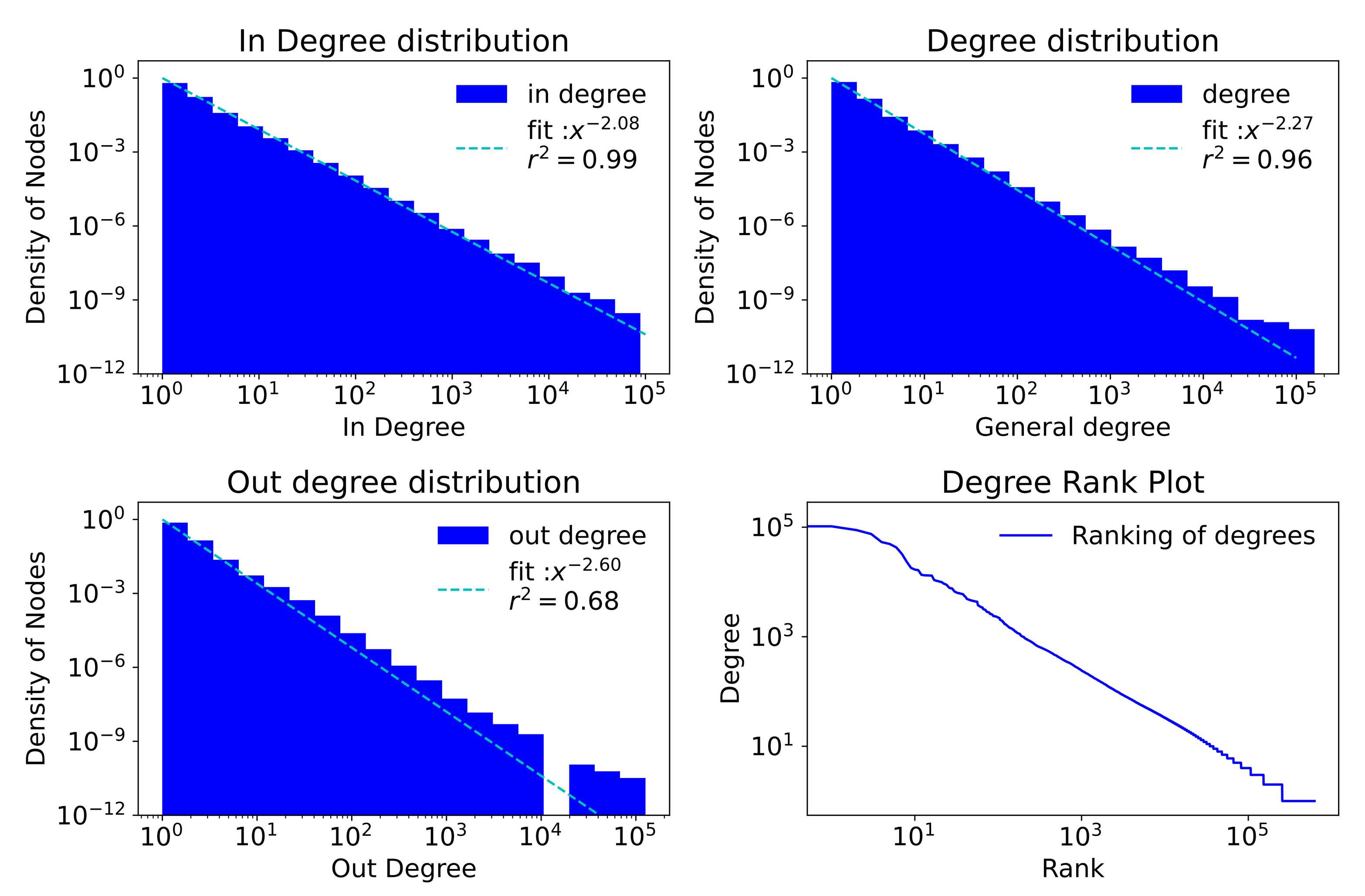
Unexpectedly, using the Maximum Likelihood Estimator, we observe for each month two different laws for *in-* and *out*-degree.

Mean values are:

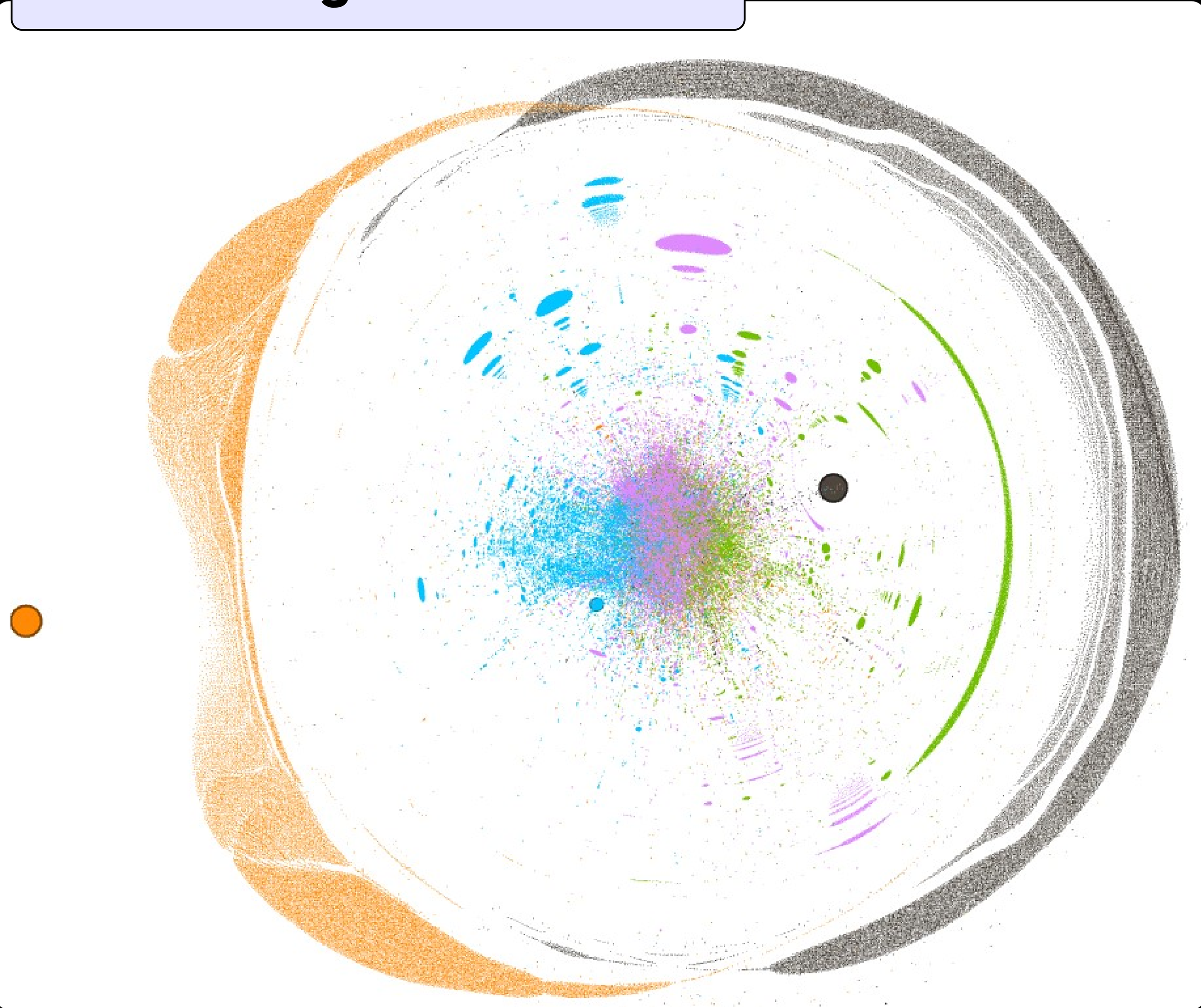
- **In-degree**: γ of 2.57 with deviation of 0.04
- **Out-degree**: γ of 2.05 with deviation of 0.02

It is worth noticing that gamma value depends on the considered time interval.

Degree distribution in January



Community visualization



Community detection

Division of the graph by using **modularity** → grouping nodes that are linked together:

A few large communities of 5%+ nodes on which to apply debt reduction algorithms.

Division of the graph by using the **k-core method** → nodes that are linked with a minimum degree of k:

The result is a core group we could use as a basis for debt reduction.

What's next?

Up to which extent do B2B graphs conform to scale-free graphs ?

What complementary information can be gained by analysing the distribution of invoices weights ?

Which characteristics define the giant strongly connected component observed in the network?

How can it be advantageously used in debt-netting algorithms ?

How can communities be further refined to better suit edge-centric applications like netting?

1- We wish to thank Infocert (Italian electronic invoices operator) for the valuable provided data.