Multilayer Aggregation with Statistical Validation:
Application to Investor Networks

Kęstutis Baltakys*, Juho Kanniainen, and Frank Emmert-Streib
*Laboratory of Industrial and Information Management, Tampere University of Technology, Finland
Corresponding author, kestutis.baltakys@tut.fi

1. Contribution

In this paper, we develop a new tractable procedure for **multilayer aggregation** based on statistical validation, which is applied to *investor networks* for illustration. The aggregation procedure can be used to integrate *security-wise* and *time-wise* information about investor trading networks, but it is not limited to finance. In fact, it can be used for different applications, such as gene, transportation, and social networks, were they inferred or observable.

Moreover, we propose two other improvements to investor network analysis: **transaction bootstrapping** and **investor categorization**. In the investor network inference, we use transaction bootstrapping for better statistical validation while investor categorization allows for constant size networks and having more observations for each node.

2. Aggregation with Statistical Validation

The network ensemble is aggregated into a weighted network $\{G_k\}_{k=1}^N \rightarrow G_w$, where the edge weights in the network G_w correspond to the number of particular edge occurrences in the ensemble, i.e.

$$n_{ij} = G_w(i,j) = \sum_{k=1}^{N} G_k(i,j), \quad (1)$$

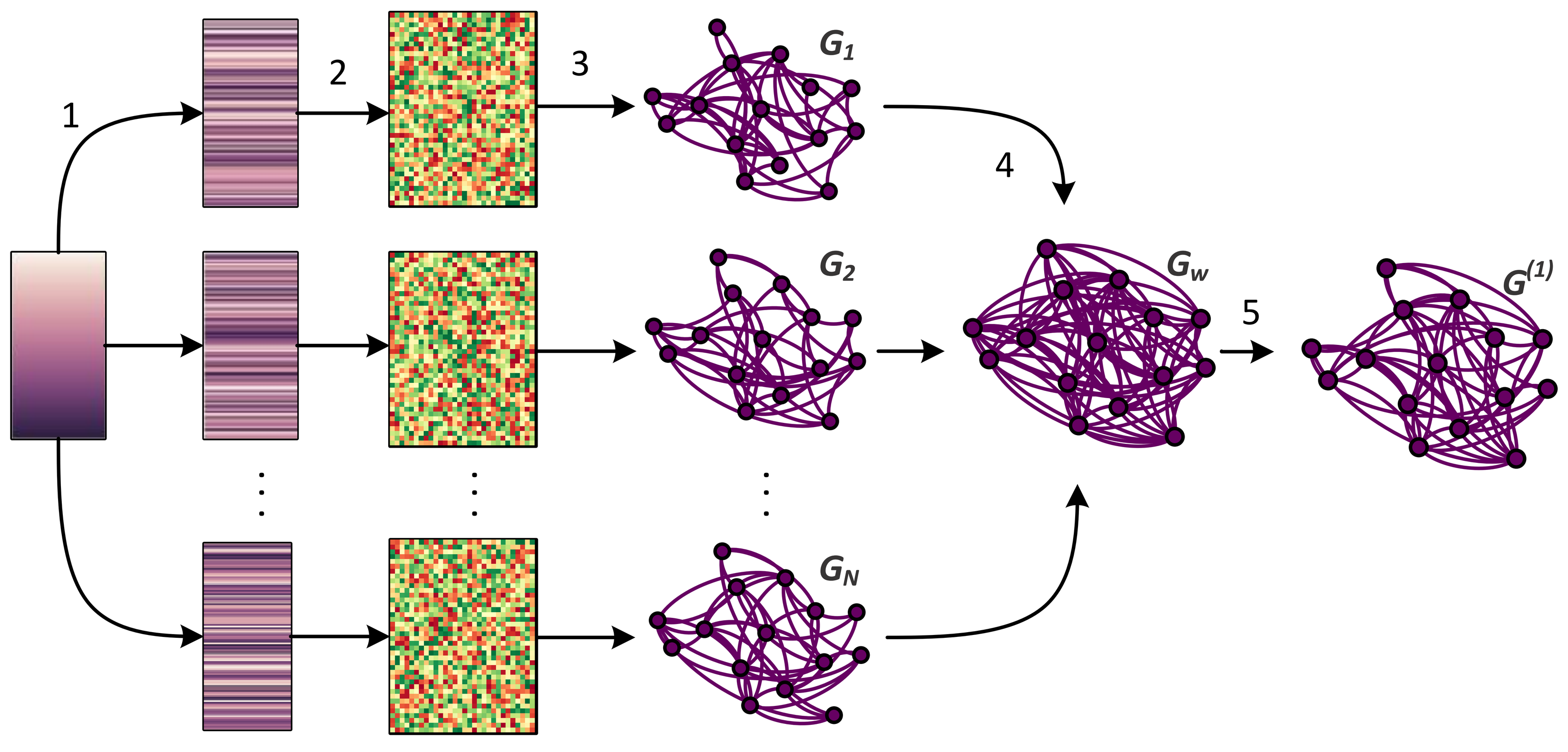
We conduct a statistical hypothesis test to remove the need for an arbitrary link threshold parameter:

$H_0^{n_{ij}}$: **The number of networks n_{ij} in the ensemble with an edge between i and j is less than $n_0(\alpha)$, where α is the significance level.**

We define p as the probability of two investor groups being randomly connected. We estimate the probability p, for two groups to be connected by chance in an N network ensemble, as the fraction of the actual number of edges in the ensemble $\sum_{i>j,k}\{G_k(i,j)\}$ to the number of all possible links in the ensemble $N \times (K(K-1)/2)$, where K is the number of investor groups. Then n_{ij} follows a binomial distribution, $B(p,N)$ and

$$p_{ij} = P(n \geq n_{ij}) = \sum_{n=n_{ij}}^{N} \binom{N}{n} p^n (1-p)^{N-n} \quad (2)$$

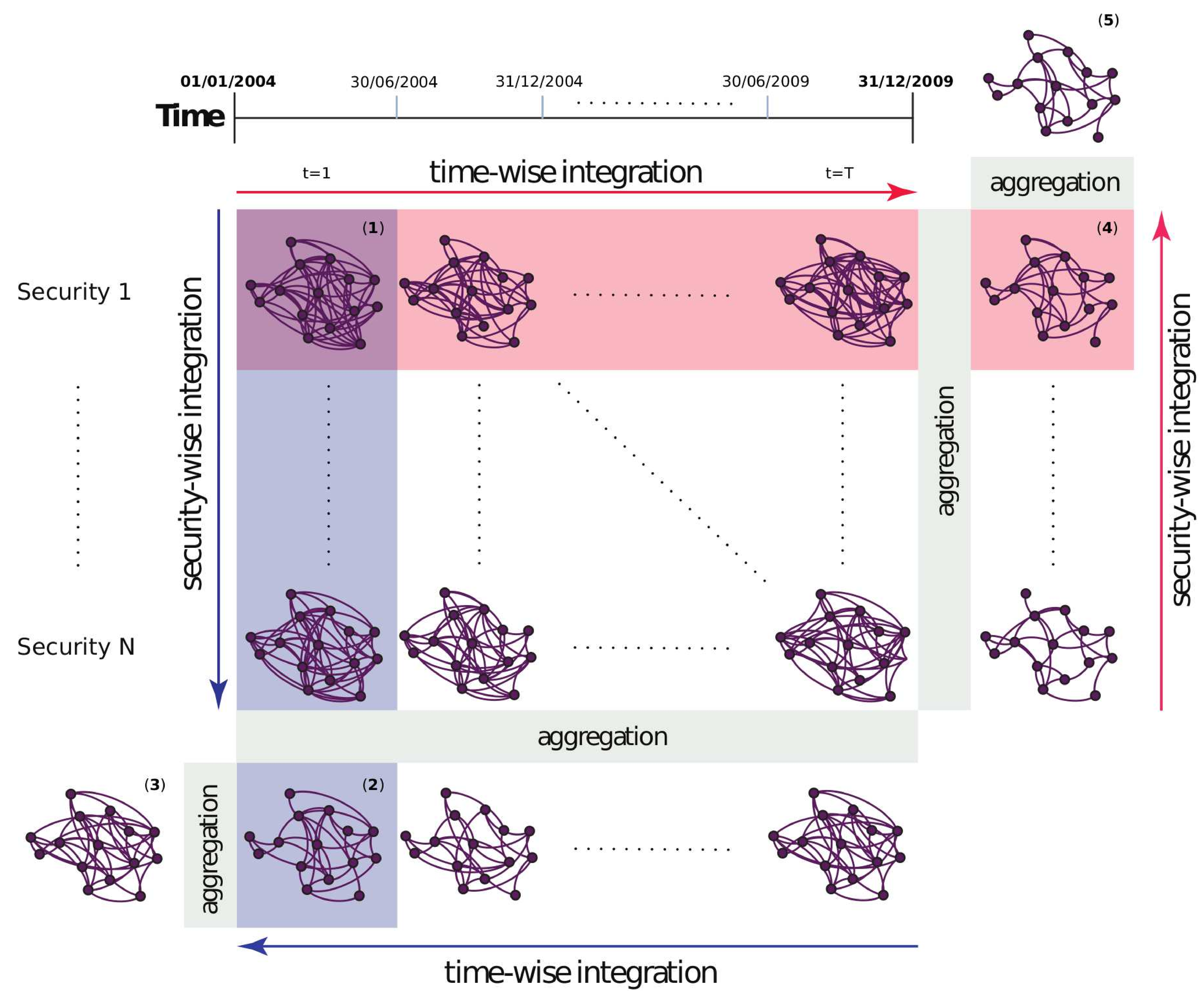
3. Transaction Bootstrapping



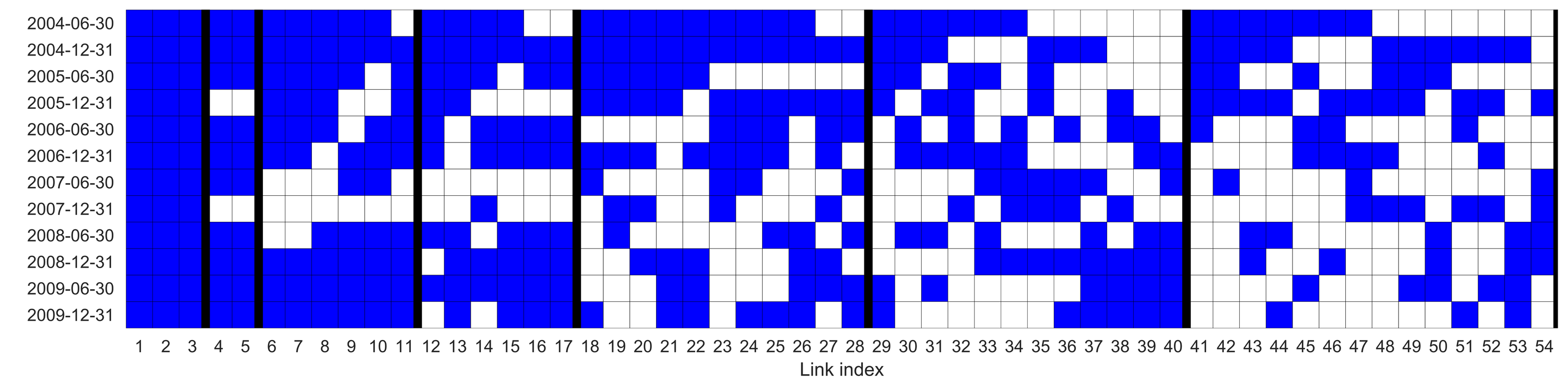
A recipe for robust investor network inference for a specific security:

1. Re-sample original transactions
2. Calculate daily net traded volumes for each category
3. Apply a network inference method to net volume time series
4. Aggregate the resulting network ensemble (eq. 1)
5. Validate network links at a chosen significance level (eq. 2)

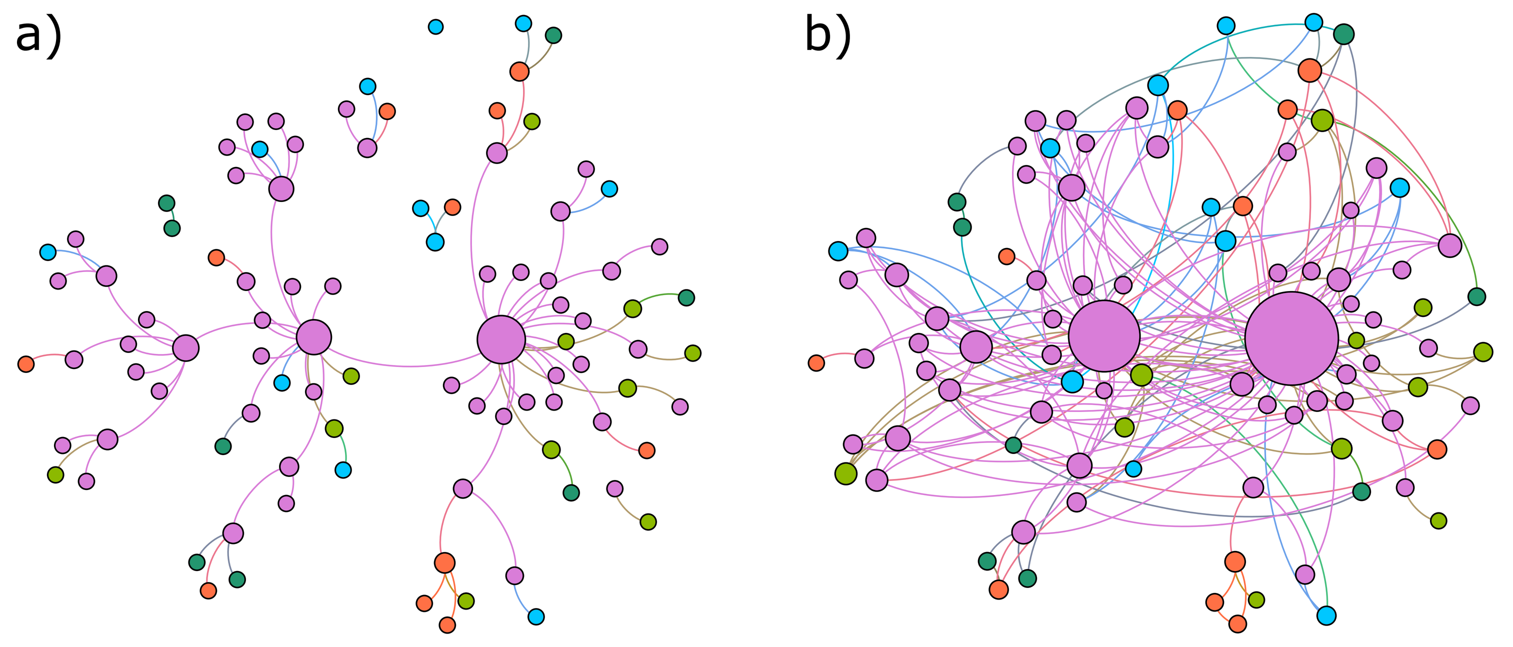
4. Multilayer Aggregation Scheme



5. Link Re-occurrence in Time-wise Aggregation

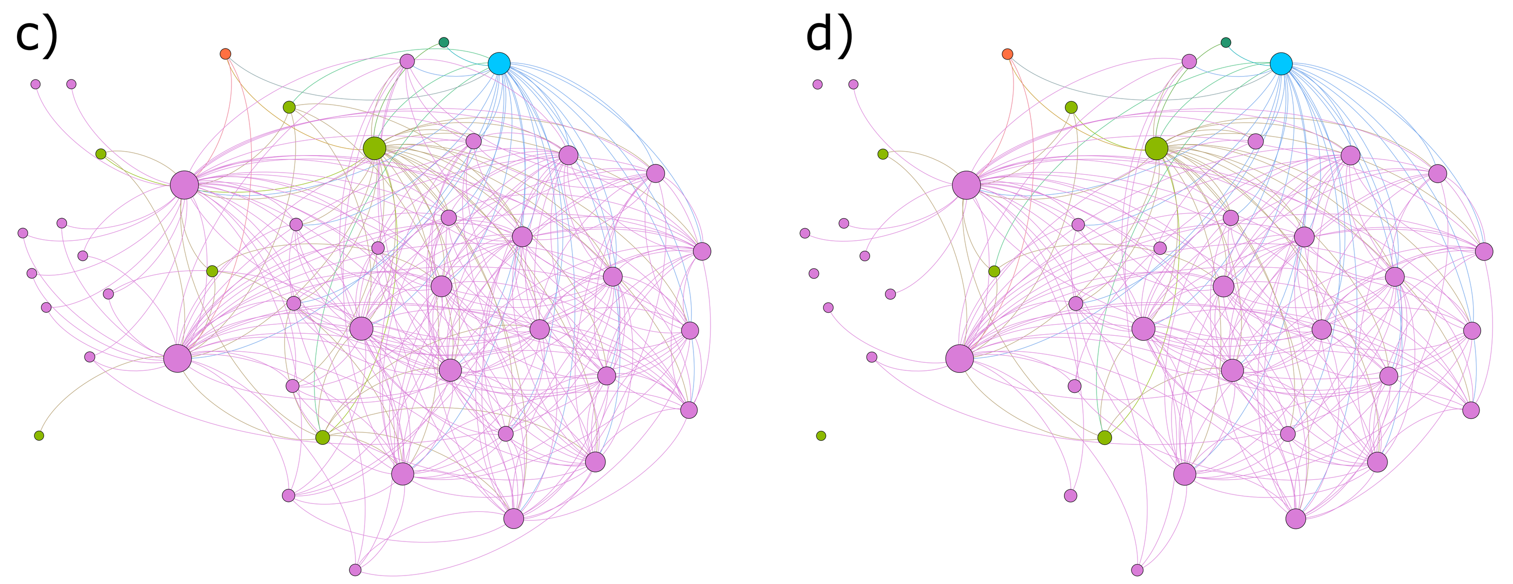


6. Networks



Nokia *investor networks* inferred using **C3NET** algorithm:

a) Original 6 year data
b) Original 6 year data, with transaction bootstrapping



100 security *investor networks* inferred using **C3NET** algortihm with transaction bootstrapping over 12 six-month non-overlapping periods:

c) Aggregation order: security-wise → time-wise:
$\{G_{st}\}_{(S\times T)} \Longrightarrow \forall t\{G_{st}\}_{s=1}^{S} \xrightarrow{\text{n. agg.}} \{G_t\} \Longrightarrow \{G_t\}_{t=1}^{T} \xrightarrow{\text{n. agg.}} \widetilde{G}$
d) Aggregation order: time-wise → security-wise:
$\{G_{st}\}_{(S\times T)} \Longrightarrow \forall s\{G_{st}\}_{t=1}^{T} \xrightarrow{\text{n. agg.}} \{G_s\} \Longrightarrow \{G_s\}_{s=1}^{S} \xrightarrow{\text{n. agg.}} \widehat{G}$

Networks $\widetilde{G}$ and $\widehat{G}$ are equivalent to networks (3) and (5) from the multilayer aggregation scheme.

● - Households, ● - Non-profit organizations, ● - Other companies,
● - Financial and insurance companies, ● - Government institutions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675044 (BigDataFinance).