On methods to assess the significance of community structure in networks of financial time series.

Argimiro Arratia [1] & Martí Renedo

[1]Universitat Politècnica de Catalunya, SPAIN
Department of Computer Science & BGSMath
argimiro@cs.upc.edu `http://www.cs.upc.edu/~argimiro/`

ITISE -Granada, 19 Sept 2017

Outline

Main goals

- The main objective of this work is to validate the results of clustering methods on networks built from currency exchanges and compare them.
- Those networks are built by evaluating correlations between exchange rate time series and using them as weighted edges, where the exchanges themselves are the vertices.
- To prove there is a significant community structure, we compare the results to the clusterings found in random networks of the same degree sequence.

Our case study: A Forex Network

Our data: The currencies included in the study are 13 of the 15 most traded as of April 2013:

- US dollar (USD)
- euro (EUR)
- yen (JPY)
- pound sterling (GBP)
- Australian dollar (AUD)
- Swiss franc (CHF)
- Canadian dollar (CAD)
- Mexican peso (MXN)
- Chinese yuan (CNY)
- New Zealand dollar (NZD)
- Swedish krona (SEK)
- Hong Kong dollar (HKD)
- Singapore dollar (SGD)

The Forex Network

- nodes will consist of every pair of currencies exchange rate return series ($n = 78$) and (possibly) their inverses (resulting in $n = 156$ nodes);
- edges between exchange-nodes weighted accordingly to their similarity.
- Using the USD exchanges data, the exchange rate of every other pair of currencies is defined: $XXX/YYY = \frac{XXX/USD}{YYY/USD}$.

Pearson correlation network

Given $\rho(r^i, r^j)$ the Pearson Correlation between two returns r_i, r_j, the weighted similarity matrix A^ρ is given by:

$$A^\rho_{ij} = \frac{1}{2}(\rho(r^i, r^j) + 1) - \delta_{ij} \qquad (1)$$

which scales the Pearson correlation from $[-1, 1]$ to $[0, 1]$, while the Kronecker delta δ_{ij} removes self-edges.

Note that $A^\rho\left(\frac{XXX}{YYY}, \frac{ZZZ}{TTT}\right) = 1 - A^\rho\left(\frac{XXX}{YYY}, \frac{TTT}{ZZZ}\right)$,
We cannot determine *a priori* wether two exchange rates will be correlated directly or with one of them inverted. As a result, we have to include all inverses in the network (so, $n = 156$)

Distance correlation network

The empirical distance correlation $\mathcal{R}_n(X, Y)$, between two samples X and Y, is the square root of

$$\mathcal{R}_n^2(X, Y) = \begin{cases} \frac{\mathcal{V}_n^2(X,Y)}{\sqrt{\mathcal{V}_n^2(X)\mathcal{V}_n^2(Y)}}, & \mathcal{V}_n^2(X)\mathcal{V}_n^2(Y) > 0 \\ 0, & \mathcal{V}_n^2(X)\mathcal{V}_n^2(Y) = 0 \end{cases} \qquad (2)$$

where $\mathcal{V}_n(X, Y)$ is the distance covariance.

The distance correlation [1] is always positive and in the $[0, 1]$ interval, so we only need to remove self edges to the matrix of correlations to obtain the adjacency matrix of the graph.

Community detection (clustering)

The clustering of the networks into communities is done using the Potts method. It consists on minimising an objective function, the Potts Hamiltonian, which evaluates the strength[1] of a partition of the graph.

The Hamiltonian of the Potts system of the partition $\mathcal{P}$ of a weighted undirected graph with adjacency matrix A is given by

$$H(\mathcal{P}) = -\sum_{ij}[A_{ij} - \gamma P_{ij}]\delta(c_i, c_j)$$

where γ is a parameter which determines how likely vertices are to form communities.

[1]Considering a strong partition one that has strong links inside the communities and weak links between them.

Scoring functions

- Scoring functions will be used to validate the results of the clustering methods.
- These functions, originally defined for unweighted graphs [2], have been extended to the weighted case.
- There are functions based on internal connectivity, external connectivity, and a combination of both.

	weighted score
Internal density	$f(S) = \frac{\tilde{m}_s}{n_S(n_S-1)/2}$
Edges Inside	$f(S) = \tilde{m}_S$
Average Degree	$f(S) = \frac{2\tilde{m}_S}{n_S}$
Expansion	$f(S) = \frac{\tilde{c}_s}{n_s}$
Cut Ratio	$f(S) = \frac{\tilde{c}_s}{n_s(n-n_s)}$
Conductance	$f(S) = \frac{\tilde{c}_s}{2\tilde{m}_s+\tilde{c}_s}$
Normalized Cut	$f(S) = \frac{\tilde{c}_s}{2\tilde{m}_s+\tilde{c}_s}$
Maximum ODF	$f(S) = \max_{u \in S} \frac{\sum_{v \notin S} w_{uv}}{\tilde{d}(u)}$
Average ODF	$f(S) = \frac{1}{n_s} \sum_{u \in S} \frac{\sum_{v \notin S} w_{uv}}{\tilde{d}(u)}$

Clustering coefficient

There are many possible extensions of the clustering coefficient for weighted networks, but some don't work well with complete graphs. Here we use the following definition [3]:

Definition

- For $t \in [0,1]$ let A_t be the adjacency matrix with elements $a_{ij}^t = 1$ if $w_{ij} \geq t$ and 0 otherwise.
- Let C_t the clustering coefficient of the graph defined by A_t.
- The resulting weighted clustering coefficient is defined as

$$\tilde{C} = \int_0^1 C_t \; dt \tag{3}$$

Random graph generation

- We want to see if the observed community structure in the FX networks is actually significant or if it can be found in similar but randomized networks.
- The switching model [4] shuffles the edges of the graph while preserving its degree sequence. We adapt it to weighted graphs:
  - Given vertices A,B,C and D, transfer a certain weight $\bar{w}$ from w_{AC} to w_{AD}, and from w_{BD} to w_{BC}. We will select only sets of vertices such that $w_{AC} > w_{AD}$ and $w_{BD} > w_{BC}$ (*i.e.* transfering weight from "heavy" edges to "weak" edges).
  - To keep the variance constant, we select

$$\bar{w} = \frac{w_{AC} + w_{BD} - w_{AD} - w_{BC}}{2}. \tag{4}$$

Comparing clusterings: Variation of information

The variation of information [5] is a distance in the space of partitions of a set which is based on information theory.

We use VI to determine how similar two clusterings of our networks are

Definition

The variation of information of partitions $\mathcal{P}$ and $\mathcal{P}'$ information is given by:

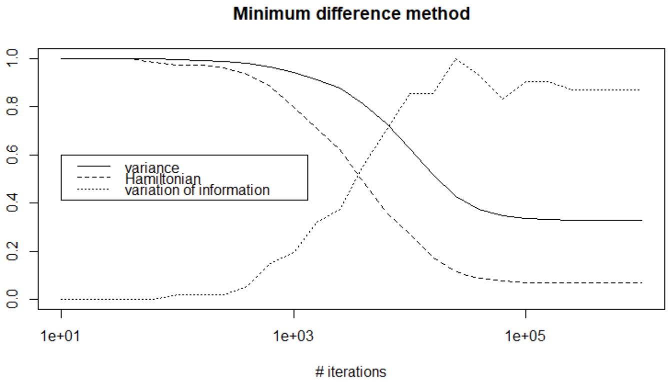
$$VI(\mathcal{P},\mathcal{P}') = \mathcal{H}(\mathcal{P}) + \mathcal{H}(\mathcal{P}') - 2I(\mathcal{P},\mathcal{P}'), \tag{5}$$

where $\mathcal{H}$ is the entropy and $\mathcal{I}$ is the mutual information.

Number of iterations for shuffling algorithm

To determine how many iterations of the algorithm are enough to sufficiently "shuffle" the network, look at VI of resulting clustering respect to the initial one (Figure 1).
As the algorithm transfers weight between the edges, the VI increases, until it stabilises roughly after 10^4 iterations. Then, it is enough to run algorithm for 10^5 iterations to generate each random graph while still being very fast to compute.

Figure: Normalised variance, Potts Hamiltonian and variation of information after applying the proposed algorithm with the minimum difference method. Horizontal axis is on logarithmic scale.



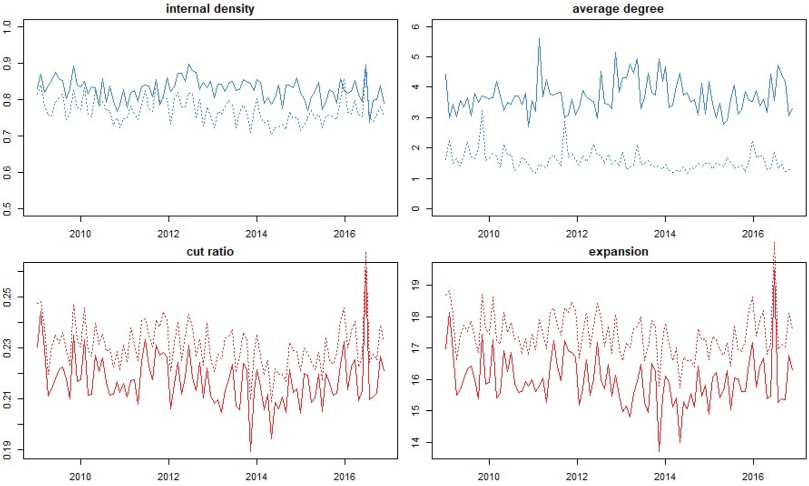
Scoring function values I

distance correlation

Scoring function values for the distance correlation networks. Solid and dashed lines correspond to the observed and randomized networks. Scores in which higher values represent stronger clusters are represented in blue, while those in which lower is better are represented in red.

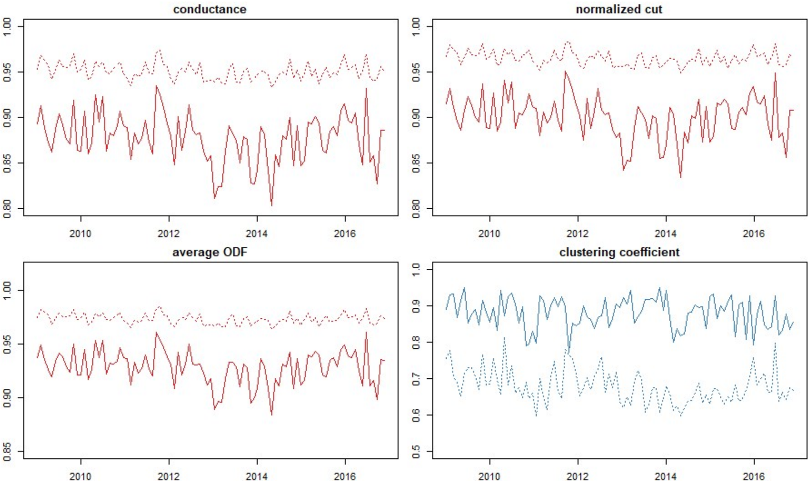
Scoring function values II

distance correlation



Scoring function values III

distance correlation



	distance correlation			Pearson correlation		
	original	randomized	variation	original	randomized	variation
internal.density	0.83	0.81	1.90 %	0.85	0.91	-6.33%
edges.inside	24.02	3.00	701.49%	89.21	3.70	2313.02%
av.degree	4.44	1.62	174.37%	8.87	2.07	329.50%
expansion	16.97	18.70	-9.24%	35.46	37.96	-6.57 %
cut.ratio.	0.23	0.25	-6.87%	0.24	0.25	-3.74%
conductance	0.89	0.95	-6.24%	0.87	0.95	-8.53%
norm.cut	0.91	0.97	-5.37%	0.89	0.96	-7.32%
max.ODF	0.94	0.97	-3.84%	0.92	0.97	-5.41%
average.ODF	0.94	0.97	-3.84%	0.92	0.97	-5.40%
clustering.coef	0.89	0.75	17.82%	0.91	0.83	9.72 %
hamiltonian	-79.49	-18.24	335.79%	-259.35	-57.89	348.02 %

Table: Means of the scoring functions over the 2009-2016 period for the randomized and observed networks, as well as the percentage of increase of one respect to the other.

Conclusions

- Our results show that FX networks have a significant community structure not present in similar random networks.
- The differences between the results of the Pearson and distance correlation methods are small in most cases, but the distance correlation has some advantages (more consistent hamiltonian, better clustering coefficient).
- The methods proposed here could be used for evaluating the results of clustering algorithms on weighted networks in general.

References I

N. K. Bakirov G. J. Székely, M. L. Rizzo.
Measuring and testing dependency by correlation of distances.
The Annals of Statistics, 35(6):2769–2794, 2007.

Jaewon Yang and Jure Leskovec.
Defining and evaluating network communities based on ground-truth.
Knowledge and Information Systems, 42(1):181–213, 2015.

Michael P. McAssey and Fetsje Bijma.
A clustering coefficient for complete weighted networks.
Network Science, 3(2):183–195, 2015.

References II

R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon.
On the uniform generation of random graphs with prescribed degree sequences.
Arxiv preprint cond-mat/0312028, 2003.

Marina Meilă.
Comparing clusterings - an information based distance.
Journal of Multivariate Analysis, 98(5):873 – 895, 2007.