Entropic Graph Embedding via Multivariate Degree Distributions

Cheng Ye, Richard C. Wilson and Edwin R. Hancock

Department of Computer Science
University of York
{ cy666, richard.wilson, edwin.hancock } @york.ac.uk

August 2014

Protein-Protein Interaction Networks

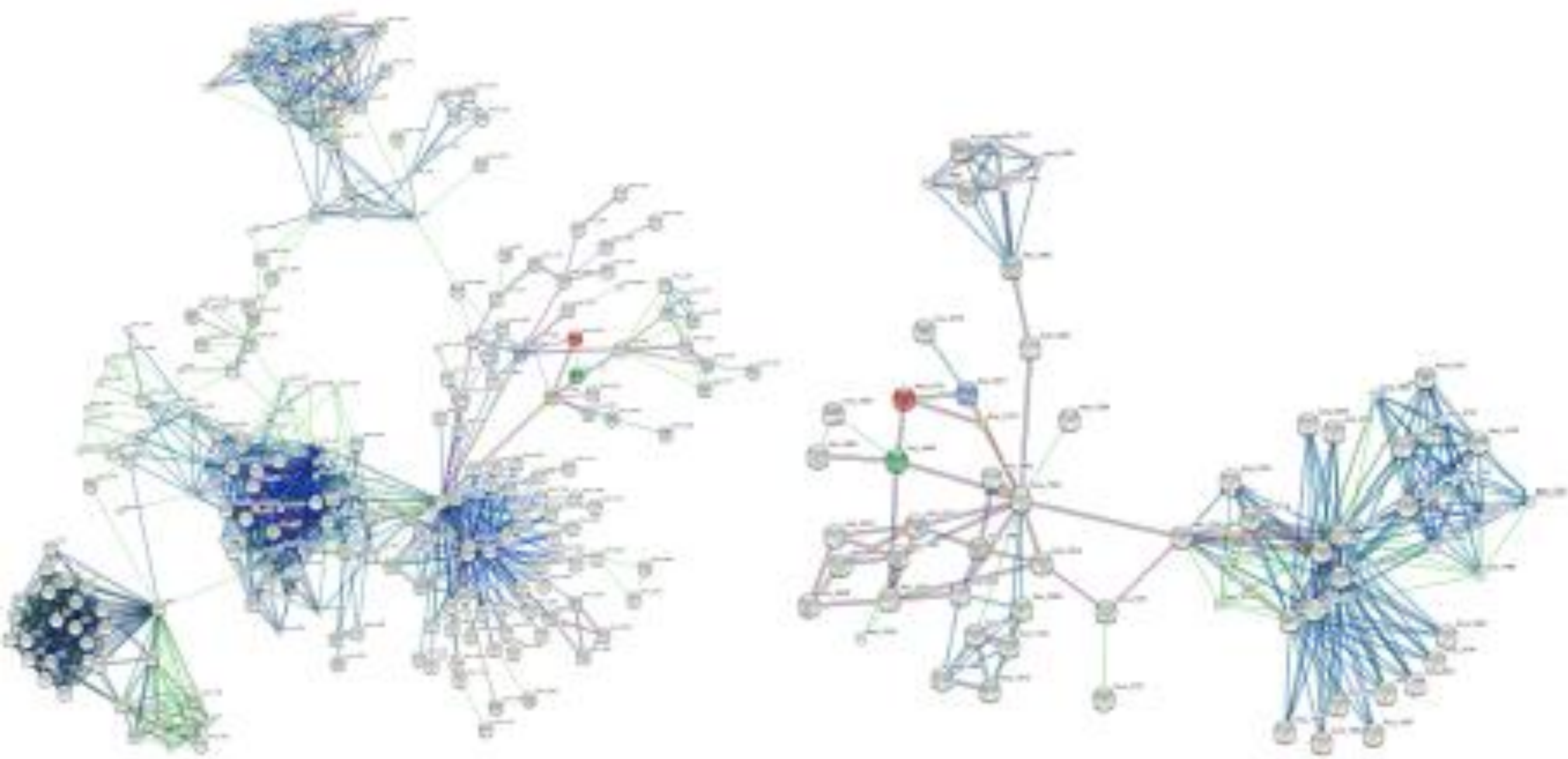


Figure: STRING Protein-Protein Interaction Networks

Characterising graph structure

- Topological: e.g. average degree, degree distribution, edge-density, diameter, cycle frequencies etc.

- Spectral or algebraic: use eigenvalues of adjacency matrix or Laplacian, or equivalently the co-efficients of characteristic polynomial.

- Complexity: use information theoretic or combinatorial measures of structural organisation or randomness (e.g. Shannon entropy).

Graph Entropy

- Entropic measures of complexity: Many possibilities - Shannon , Erdos-Renyi, Von-Neumann.

- Problems: Difficult to compute for graphs. Require either probability distribution over nodes of graph or combinatorial (micro-canonical state) characterisation. Remains open problem in literature.

- Uses: Complexity level analysis of graphs, learning structure via description length, construct information theoretic kernels.

Entropy

- Thermodynamics: measure of disorder in a system. Change in entropy with energy measure temperature of system $\Delta E = T\Delta H$.

- Statistical mechanics: Entropy is measure of uncertainty of microstates of a system $H=-k \Sigma_{\iota} p_i \ln p_i$ – Boltzmann.

- Quantum mechanics: Confusion of states $H=-kTr[\rho \ln \rho]$ in terms of density matrix ρ for states of operator O– Von Neumann.

- Information theory: Shannon information $H=- \Sigma_{\iota} p_i \ln p_i$ – in terms of probability of transmission of a message in an information channel.

Von Neumann entropy

- Passerini and Severini – normalised Laplacian $L=D^{-1/2}(D-A)\ D^{-1/2\ l}$ is density matrix for graph.

- Exploited to compute approximate VN entropy for undirected graphs by Han et al (PRL) 2013 and by Cheng et al (Phys Rev E 2014) for directed graphs.

- Used for graph kernel construction (Bai JMIV 2013) and learning generative models of graphs (Han SIMBAD 2011).

- Recently used as unary feature to analyse and classify complex network time series.

This work

- Entropy is function of degree statistics for nodes connected by edges. Distribution of entropy controlled by distribution of degree (in/out degree for directed graphs).

- Can distribution of entropy with degree (entropy weighted degree distribution) be used as a feature to distinguish graphs of different intrinsic structure?

- Investigate multivariate histograms of entropy over degree as graph feature-vectors.

Von Neumann entropy and node degree

Von-Neumann Entropy

- Passerini and Severini – normalised Laplacian is density matrix for graph Hamiltonian

$$\hat{L} = D^{-1/2}(D-A)D^{-1/2} = \hat{\Phi}\hat{\Lambda}\hat{\Phi}^T$$

- Associated Von Neumann entropy is .

$$H_{VN} = -\sum_{i=1}^{|V|} \frac{\hat{\lambda}_i}{2} \ln \frac{\hat{\lambda}_i}{2}$$

Approximation

- Quadratic entropy

$$H_{VN} = \sum_{i=1}^{|V|} \frac{\hat{\lambda}_i}{2} \left\{ 1 - \frac{\hat{\lambda}_i}{2} \right\} = \frac{1}{2} \sum_{i=1}^{|V|} \hat{\lambda}_i - \frac{1}{4} \sum_{i=1}^{|V|} \hat{\lambda}_i^2$$

- In terms of matrix traces

$$H_{VN} = \frac{1}{2} Tr[\hat{L}] - \frac{1}{4} Tr[\hat{L}^2]$$

Computing Traces

- Normalised Laplacian

$$Tr[\hat{L}] = |V|$$

- Normalised Laplacian squared

$$Tr[\hat{L}^2] = |V| + \sum_{(u,v)\in E} \frac{1}{4d_u d_v}$$

Simplified entropy

Collect terms together, von Neumann entropy reduces to

$$H_{VN} = \frac{1}{4}|V| - \sum_{(u,v)\in E} \frac{1}{4d_u d_v}$$

Properties

Based on degree statistics

Extremal values for cycle and star-graphs

Can be used to distinguish Erdos-Renyi, small worlds, and scale free networks.

Extend to directed graphs

- Described in Cheng et al Phys. Rev E 2014.
- Commence from Chung's spectral specificiation of directed graph Laplacian.
- Repeat analysis steps to extend quadratic approximation of VN entropy to directed graphs.
- Find specific approximations for strongly and weakly directed graphs.

Directed Laplacian

$$P_{ij} = \begin{cases} \frac{A_{ij}}{d_i^{out}} & \text{if } (i,j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Transition matrix

$$\phi P = \rho \phi$$

left eigenvector

$$\frac{\phi(i)}{\phi(j)} \approx \frac{d_i^{in}}{d_j^{in}}.$$

components

Noramised Laplacian

$$\tilde{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^T \Phi^{1/2}}{2}.$$

$$\Phi = diag(\phi(1), \phi(2), ...)$$

$$
\begin{aligned}
Tr[\tilde{L}] &= Tr[I] - \frac{1}{2}Tr[P\Phi^{-1/2}\Phi^{1/2}] - \frac{1}{2}Tr[P^T\Phi^{1/2}\Phi^{-1/2}] \\
&= Tr[I] - \frac{1}{2}Tr[P] - \frac{1}{2}Tr[P^T].
\end{aligned}
$$

$$
Tr[\tilde{L}] = Tr[I] = |V|,
$$

Trace calculations

$$
\begin{aligned}
Tr[\tilde{L}^2] &= Tr[I^2 - (\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}) + \\
&\quad \frac{1}{4}(\Phi^{1/2}P\Phi^{-1/2}\Phi^{1/2}P\Phi^{-1/2} + \Phi^{1/2}P\Phi^{-1/2}\Phi^{-1/2}P^T\Phi^{1/2} + \\
&\quad \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{-1/2}P^T\Phi^{1/2})] \\
&= Tr[I^2] - Tr[P] - Tr[P^T] + \frac{1}{4}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi] + Tr[P^T\Phi P\Phi^{-1}] + Tr[P^{T^2}]) \\
&= |V| + \frac{1}{2}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi]), \qquad (12)
\end{aligned}
$$

$$
Tr[P^2] = \sum_{i=1}^{|V|}\sum_{j=1}^{|V|} P_{ij}P_{ji} = \sum_{(i,j)\in E_2} \frac{1}{d_i^{out}d_j^{out}}.
$$

$$
Tr[P\Phi^{-1}P^T\Phi] = \sum_{i=1}^{|V|}\sum_{j=1}^{|V|} P_{ij}^2\frac{\phi(i)}{\phi(j)} = \sum_{(i,j)\in E} \frac{\phi(i)}{\phi(j)d_i^{out^2}}.
$$

Von Neumann Entropy

- Partition edge set into undirectional edges (one way only) E_1 and bidirectional edges (both ways) E_2.

- Edge set is union $E=E_1 U E_2$

Directed Graphs

Von Neumann entropy depndes on in-degree and out-degree of vertices connected by edges

$$H = 1 - \frac{1}{V} - \frac{1}{2V^2}\left\{ \sum_{(u,v)\in E} \frac{d_u^{in}}{d_v^{in}(d_u^{out})^2} - \sum_{(i,j)\in E_2} \frac{1}{d_u^{out} d_v^{out}} \right\}$$

Development comes from Laplacian of a directed graph (Chung).

$$H = 1 - \frac{1}{V} - \frac{1}{2V^2}\left\{ \sum_{(u,v)\in E} \left(\frac{d_u^{in}}{d_u^{out}} \right) \frac{1}{d_u^{out} d_v^{in}} - \sum_{(i,j)\in E_2} \frac{1}{d_u^{out} d_v^{out}} \right\}$$

Strongly Directed Graphs

Most of edges are unidirectional, few bidirectional edges ($|E_1| >> |E_2|$)

$$H = 1 - \frac{1}{V} - \frac{1}{2V^2} \sum_{(i,j) \in E} \frac{1}{d_u^{out} d_v^{in}}$$

Weakly Directed Graphs

Most of edges are bidirectional, few unidirectional edges ($|E_1| << |E_2|$)

$$H = 1 - \frac{1}{V} - \frac{1}{2V^2} \sum_{(u,v)\in E} \frac{d_u^{in} / d_u^{out} + d_v^{in} / d_v^{out}}{d_u^{out} d_v^{in}}$$

Development comes from Laplacian of a directed graph (Chung).

Links to assortivity

- Weakly directed: proportional to in/out degree ratio of nodes; inversely proportional to product of out degree of start node and in degree of end node (degree flow).

- Strongly directed: inversely proportional to product of out degree of start node and in degree of end node.

Entropy Feature Vectors

Multivariate histograms of entropy with degree

Entropy Increments

When cardinality of bidirectional edge set is very small, i.e., the graph is strongly directed (SD), the entropy formula can be simplified a step further

$$H_{VN}^{SD} = \frac{1}{2|V|} \sum_{(u,v)\in E} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out^2}} \right\}$$

Normalized local entropic measure for each undirectional edge in the graph

$$I_{uv} = \frac{d_u^{in}}{2|E||V| d_v^{in} d_u^{out^2}}$$

For bidirectional edges , we add an additional contribution to the above measure

$$I'_{uv} = \frac{1}{2|E_b||V| d_u^{out} d_v^{out}}$$

Feature Vector from Entropy Distribution

Graph characterization: based on the statistical information conveyed by edge entropy distribution

Representation: 4D histogram over the in and out-degrees of the two vertices connected by an edge.

Potential problem - bin-contents can become sparse in a high dimensional histogram. Compute cumulative distribution function (CDF) over predefined quantiles.

Let $P(X = d_i^{in})$ be the in-degree probability distribution of the graph, the corresponding CDF is .

$$F_X(d_i^{in}) = P(X \leq d_i^{in})$$

Quantising the multivariate entropy distribution

The m-quantiles of the in-degree distribution are

$$Q_j = \operatorname*{argmin}_{d_i^{in}} \left\{ F_{Q_j}(d_i^{in}) - \frac{j}{m} \right\}$$

Assign each vertex degree quantile labels ranging from 1 to m, allowing us to construct a 4D histogram whose size in each dimension is fixed to m.

Storing information: m x m x m x m array (histogram) M

elements: represent the histogram bin-contents,

indices: represent the degree quantile labels of the vertices.

Elementwise accumulation: formally given as

$$M_{ijkl} = \sum_{\substack{q_u^{out}=i,\, q_u^{in}=j \\ q_v^{out}=k,\, q_v^{in}=l \\ (u,v)\in E}} \left\{ \frac{d_u^{in}}{2|E||V|d_v^{in} d_u^{out^2}} \right\}$$

Bidirectional Edges

Bidirectional edges: additionally accumulate

$$M'_{ijkl} = \sum_{\substack{q_u^{out}=i, q_u^{in}=j \\ q_v^{out}=k, q_v^{in}=l \\ (u,v)\in E_b}} \left\{ \frac{1}{2|E_b||V|d_u^{out} d_v^{out}} \right\}$$

Feature Vectors

Feature vector: concatenate the elements of M to give long-vector

$$v = (M_{1111}, M_{1112}, \cdots, M_{111m}, M_{1121}, M_{1122}, \cdots, M_{mmmm})^T$$

of length m^4.

Strongly directed graphs: entropy formula does not depend on d_v^{out} - dimensionality of matrix M is reduced to 3.

PCA: perform PCA on feature vectors (entropy component analysis)

➢ Experiments and Discussion

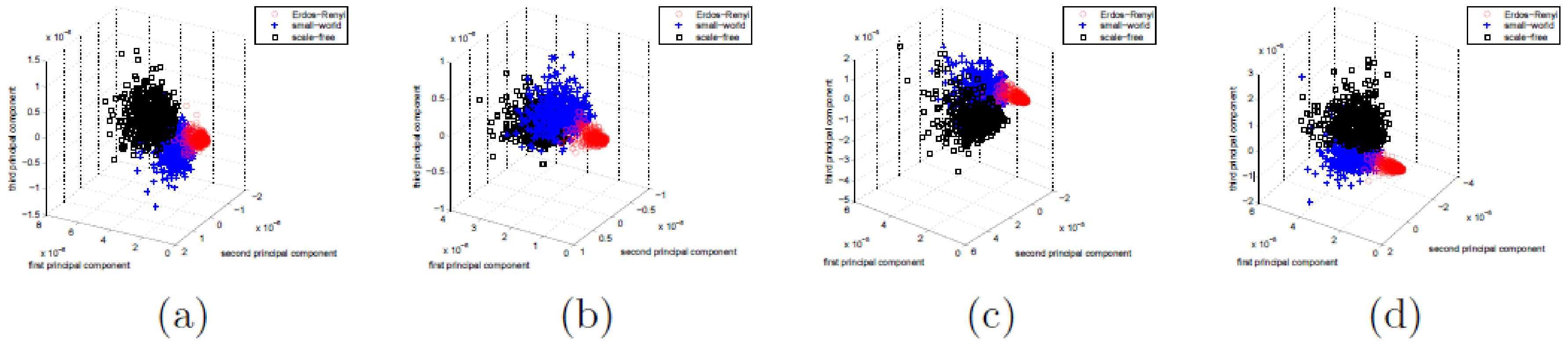
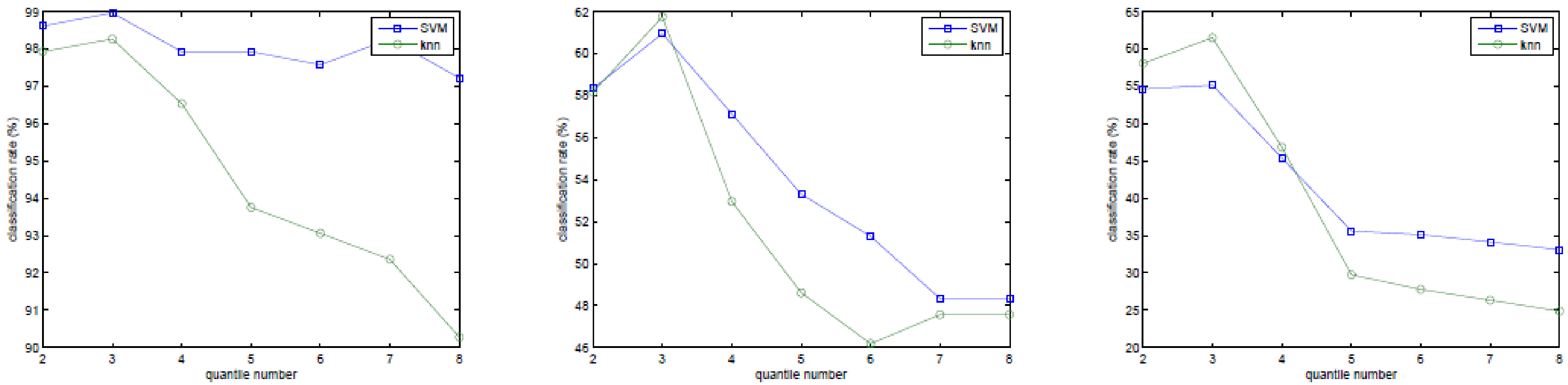


Fig. 1. Clustering performance for random graphs using PCA: a) FF feature vectors extracted from normal directed graphs; b) SD feature vectors extracted from normal directed graphs; c) FF feature vectors extracted from SD graphs; d) SD feature vectors extracted from SD graphs. Red: Erdős-Rényi graphs; blue: "small-world" graphs; black: "scale-free" graphs.

Three classes of random graphs well separated, but

- ``small-world'' graphs and ``scale-free'' graphs show some overlap.
- Suggests the full feature vectors are efficient in distinguishing any normal directed graphs
- Reduced vectors effective only for strongly directed graphs.



(a) (b) (c)

Fig. 2. Average classification rates for both SVM and kNN classifiers with different quantile numbers on datasets: a) 4-object data; b) 8-object data and c) Protein Data. Square: SVM classifier; circle: kNN classifier.

Observtions:

- classification performance is particularly good on 4-object data
- on 8-object data and 6-class protein database, the accuracy is still acceptable.
- all vertices have the same out-degree 3, classification rates peak when m=3 since feature vectors preserve in and out-degree statistics.

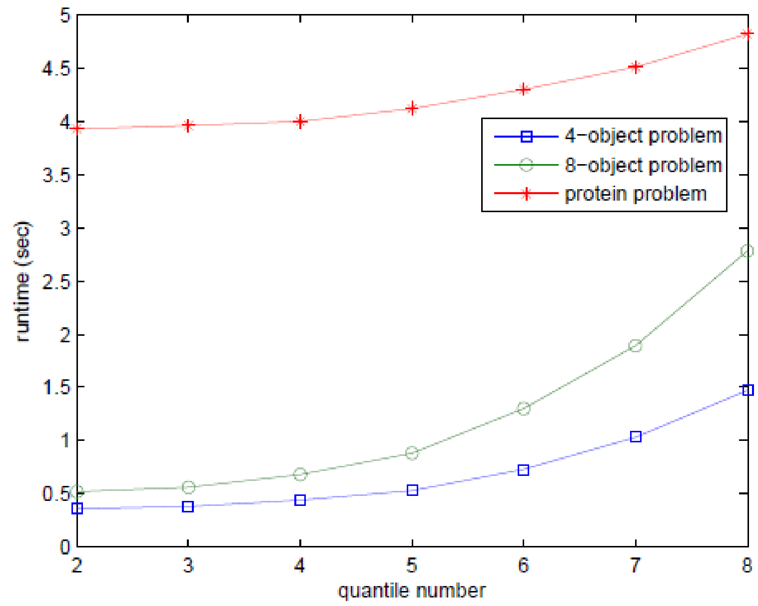


Fig. 3. Average experimental runtime with various quantile numbers for different classification problems. Square: 4-object problem; circle: 8-object problem; star: protein problem.

It is clear that our directed graph characterization is computationally tractable as the runtime does not increase rapidly even when the size of the feature vector becomes particularly large.

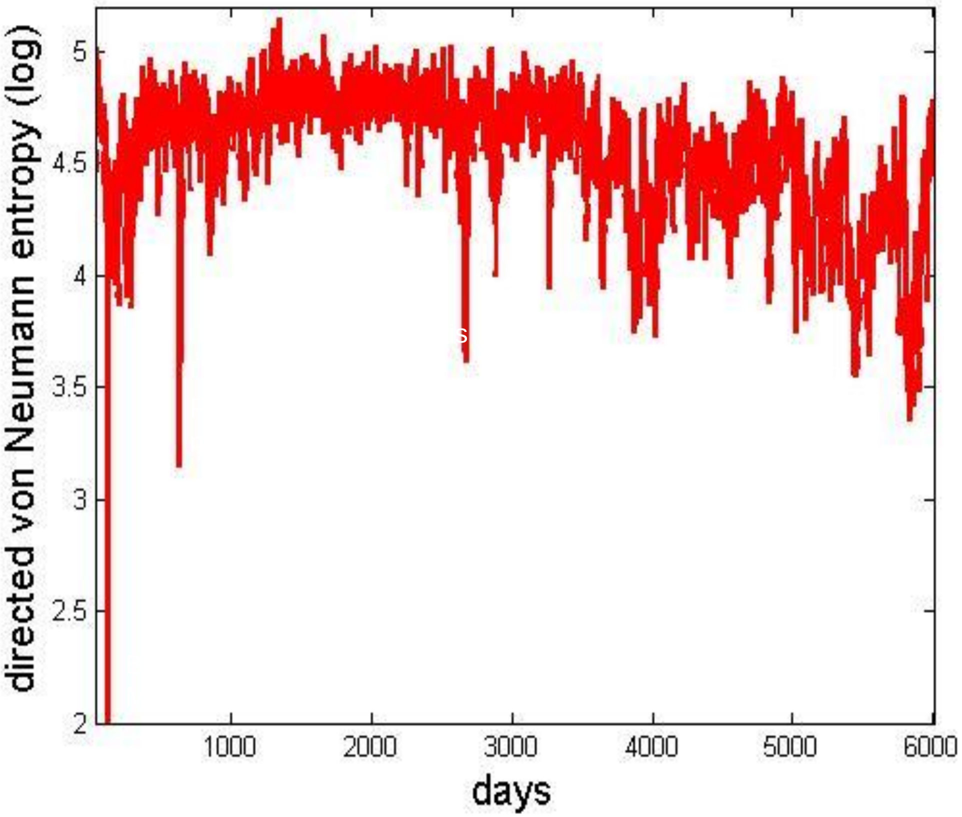
Financial Market Data

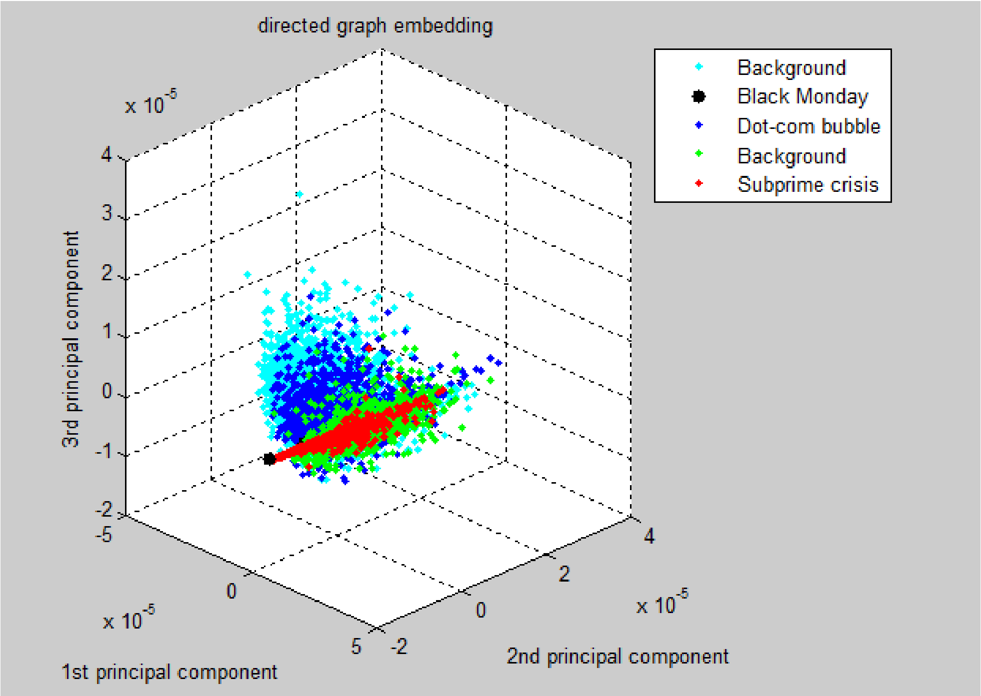
- Look at time series correlation for set of leading stocks.

- Create undirected or directed links on basis of time series correlation.

- **Directed von Neumann entropy as stock market network evolves.**

- **Troughs represent financial crises, while the deepest one corresponds to Black Monday, 1987.**

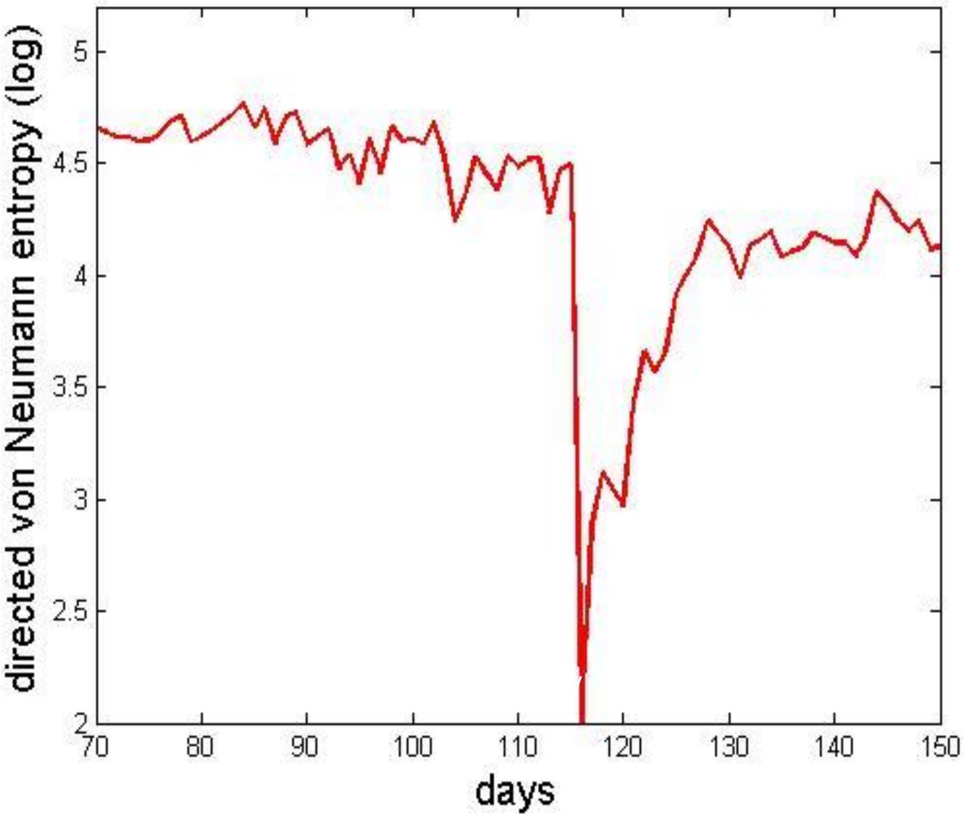




PCA applied to entropy feature vectors: distinct epochs of market evolution occupy different regions of the subspace and can be separated. Black Monday is a clear outlier. There appears to be some underlying manifold structure.

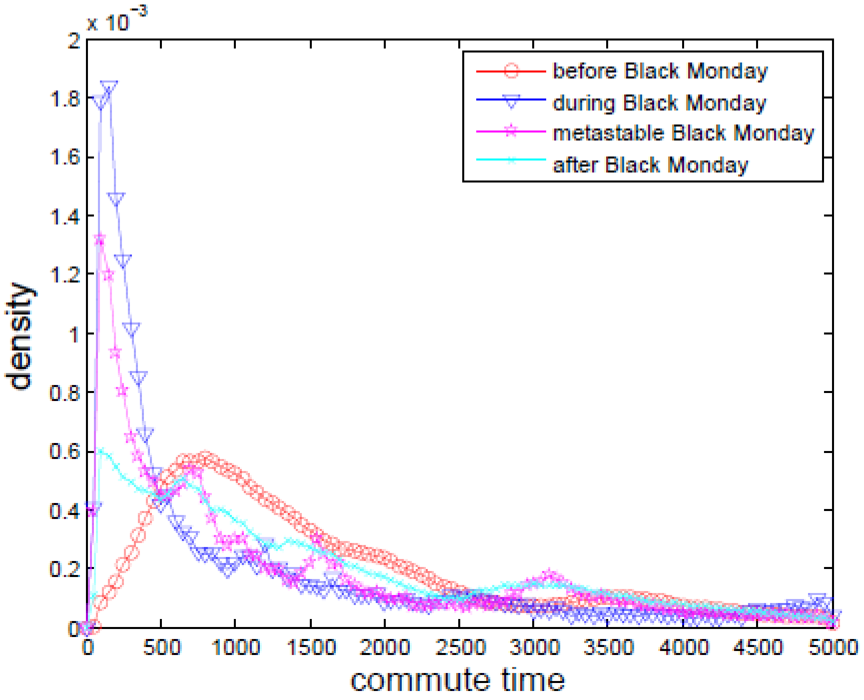
- **Directed von Neumann entropy change during Black Monday, 1987.**

- **Entropy witnesses a sharp drop on Black Monday and recovers in a few trading days' time.**



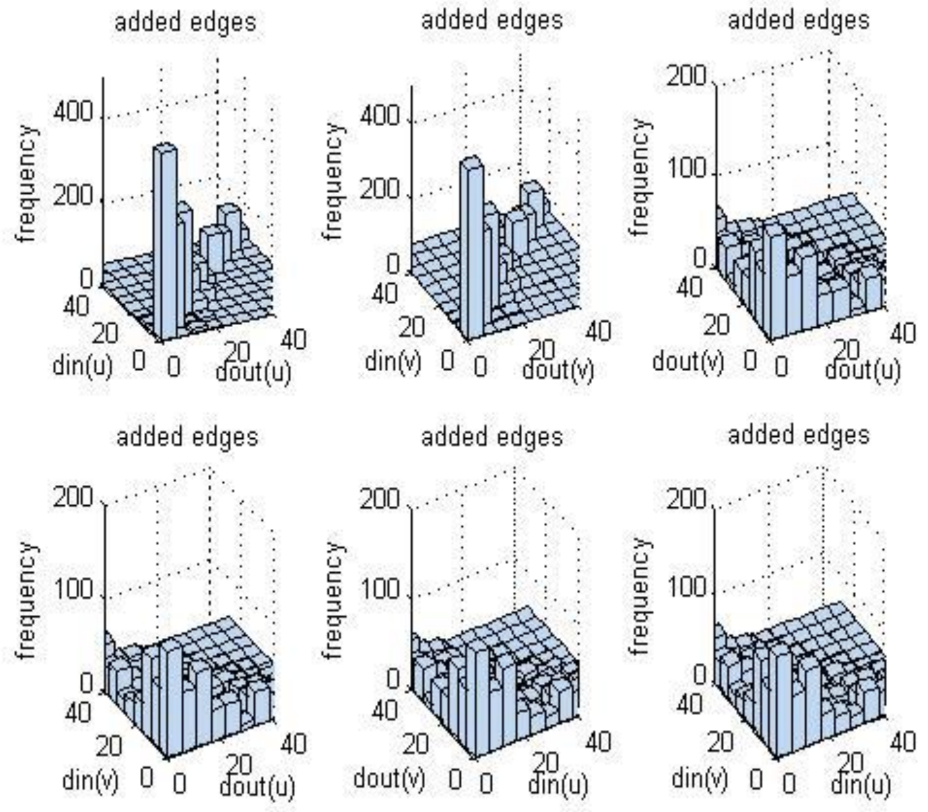
What happened

- Entropy drops during crisis and remains lower than pre-crisis level (via metastable intermediate state).
- Phase transition.
- Low degree connections replaced by high degree connections.
- Network pathlengths reduced



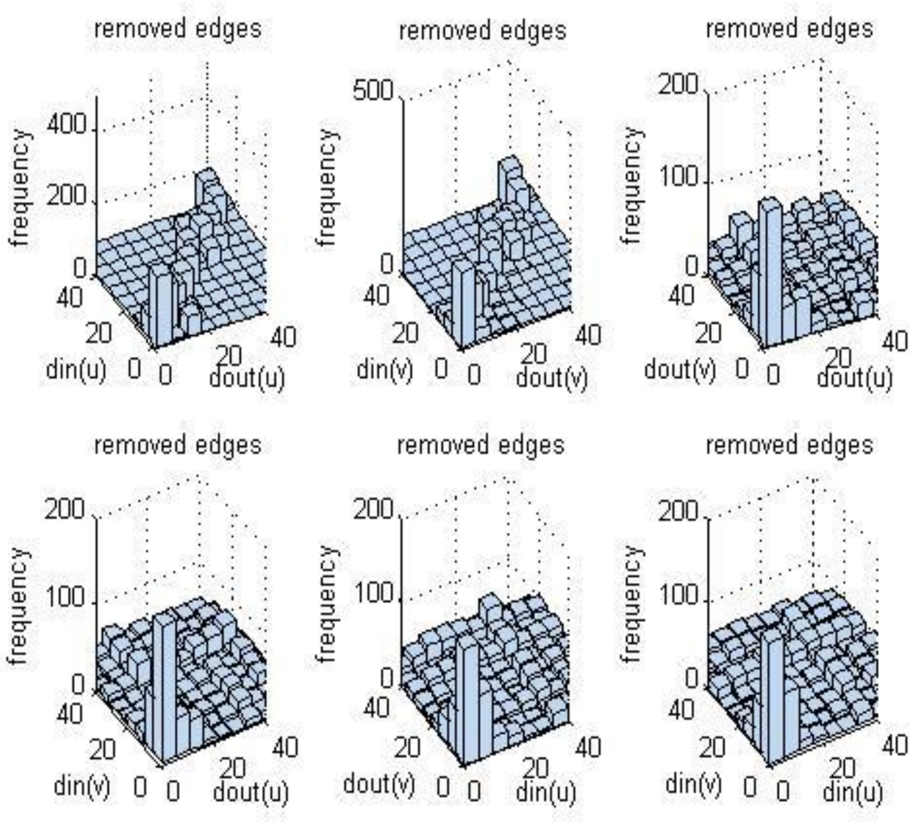
- **Histograms of added edges in degree space during financial crisis (Black Monday).**

- **Most edges have higher probability to appear between two nodes with low degrees.**

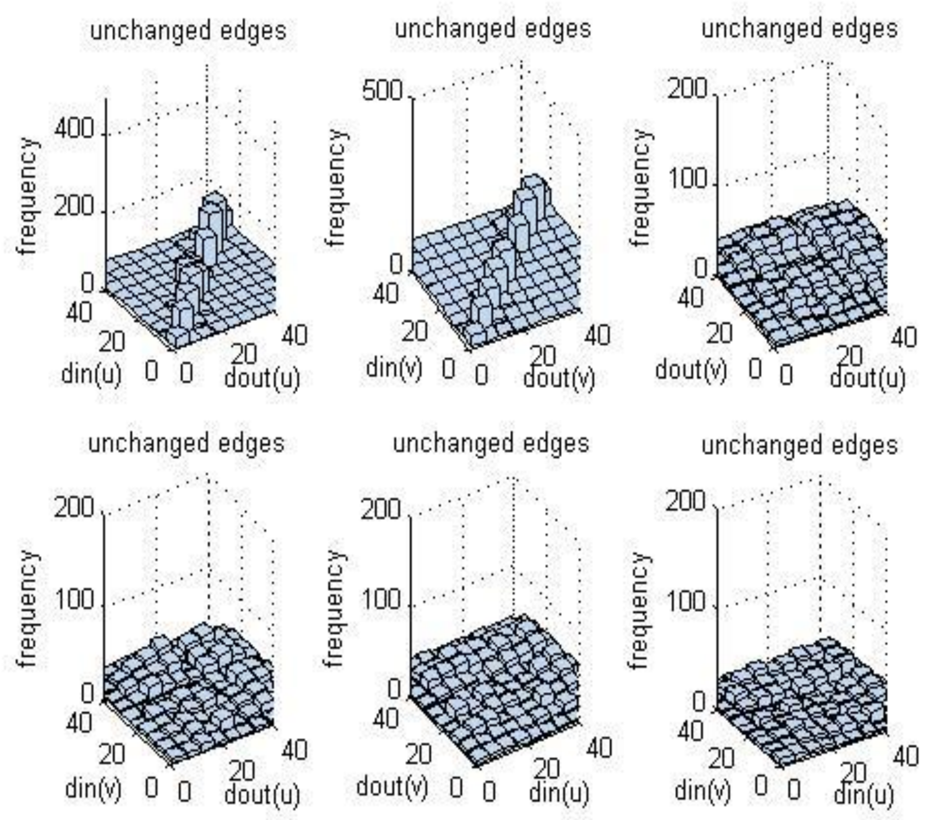


- **Histograms of removed edges in degree space during financial crisis (Black Monday).**

- **Edges that connect two nodes with low degrees have higher probability to disappear.**



- **Histograms of unchanged edges in degree space during financial crisis (Black Monday).**
- **Edges that connect two nodes with higher degrees are more likely to remain unchanged.**



…but

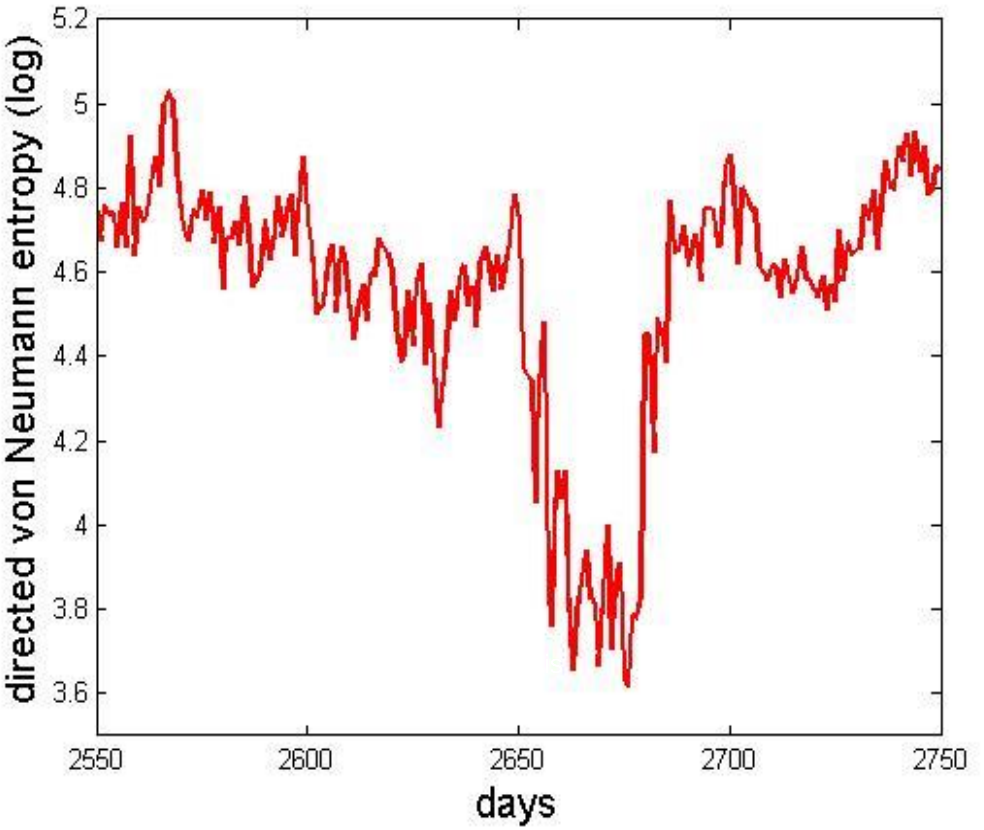
- Different crises have a different structure.
- Entropy (plus additional thermodynamic concepts such as temperature and volume) allow these to be analysed and classified.
- Taxonomy of phase transitions.

Conclusions

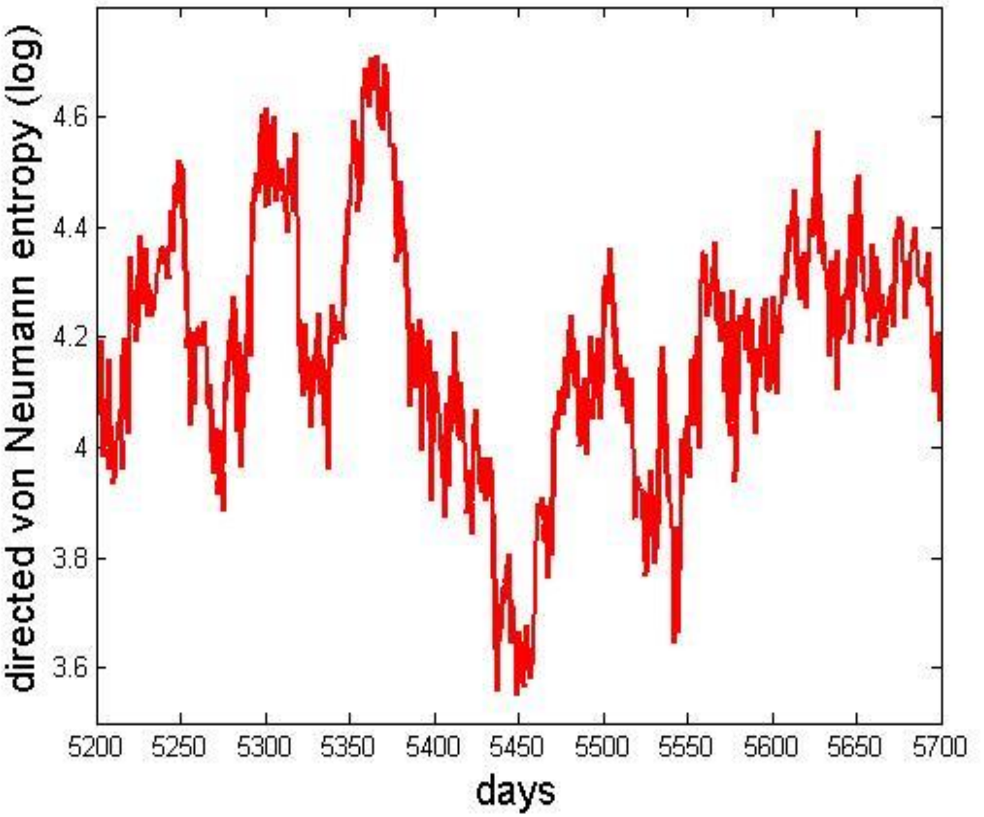
- Method is form of entropy component analysis and is easily kernelised.
- Effective and efficient tool for complex network analysis.
- Future: explore entropy kernel component analysis of time series data.

- **Directed von Neumann entropy change during 1997 Asian financial crisis.**

- **Entropy decreases significantly and does not recover until after a relatively long time.**

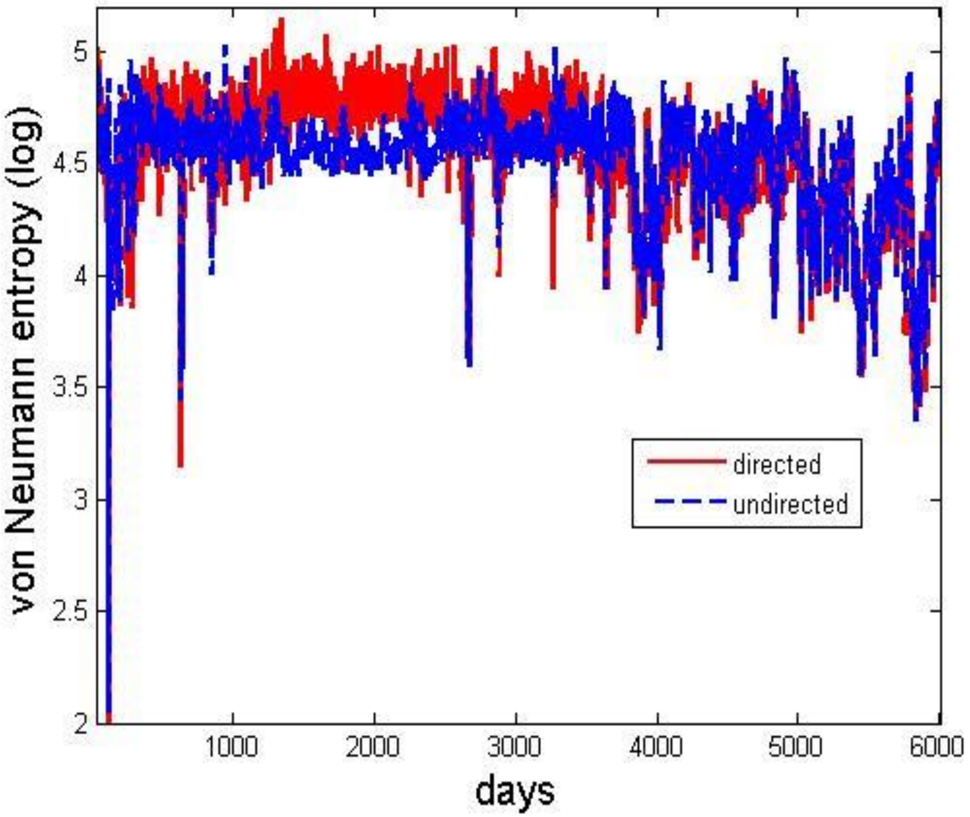


- **Directed von Neumann entropy change during Bankruptcy of Lehman Brothers.**
- **Entropy fluctuates significantly and cannot recover in a long time.**

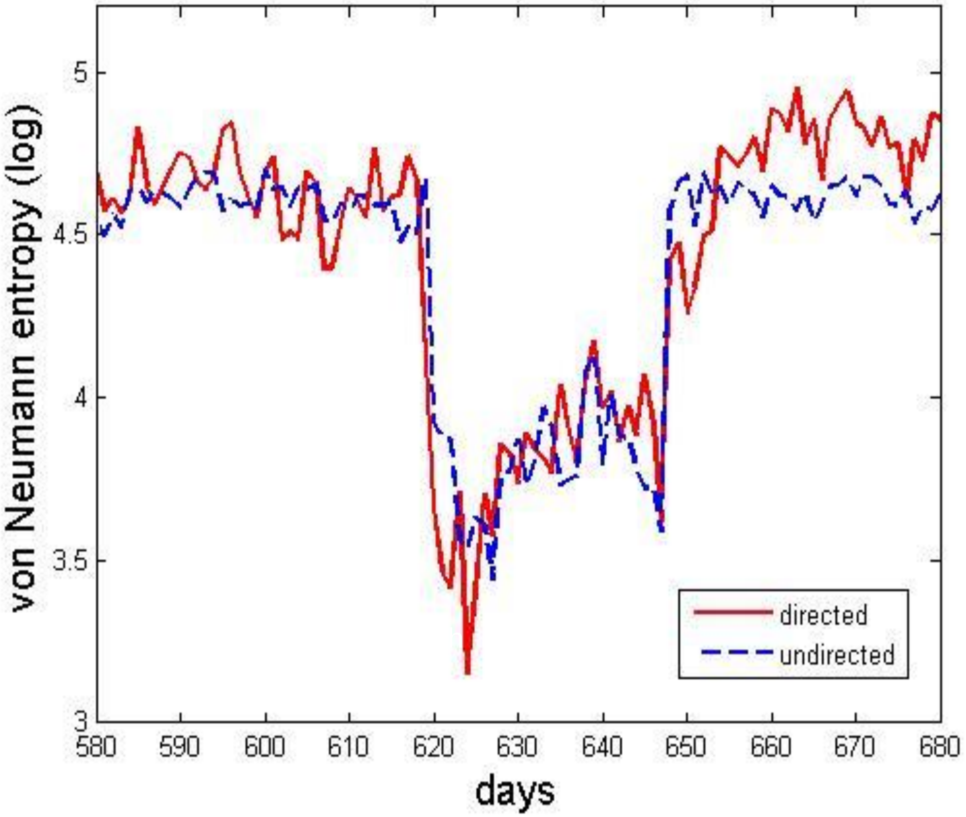


- Comparison of directed and undirected von Neumann entropies as stock market network evolves.

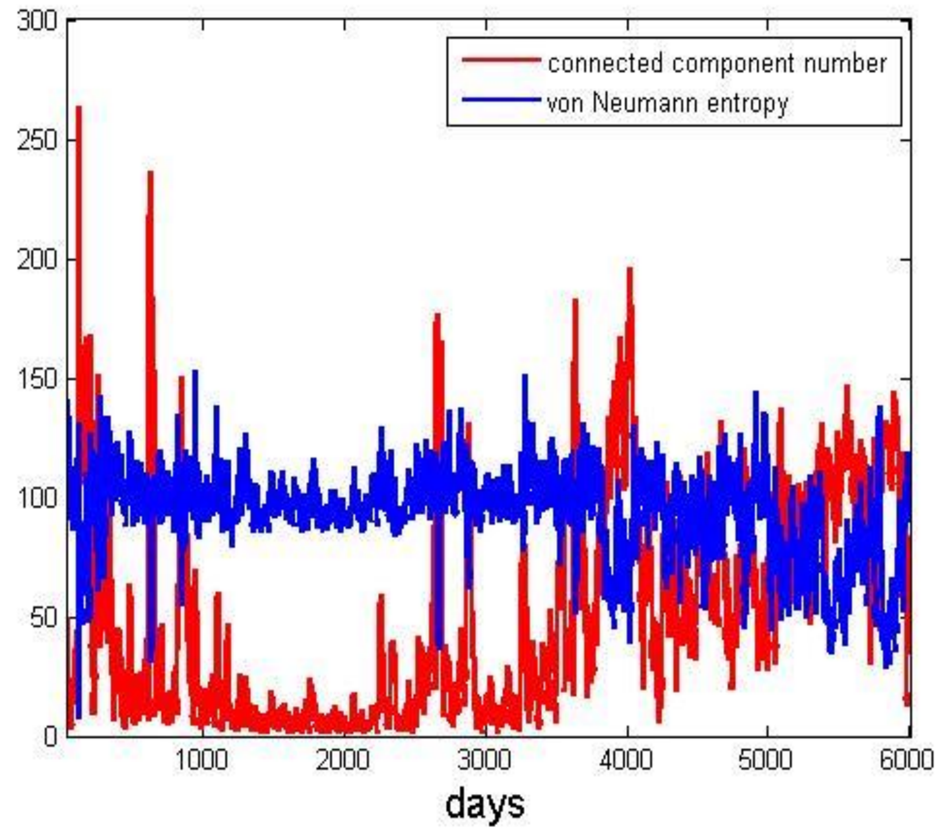
- Directed entropy is more sensible to slight changes in network structure than undirected entropy.



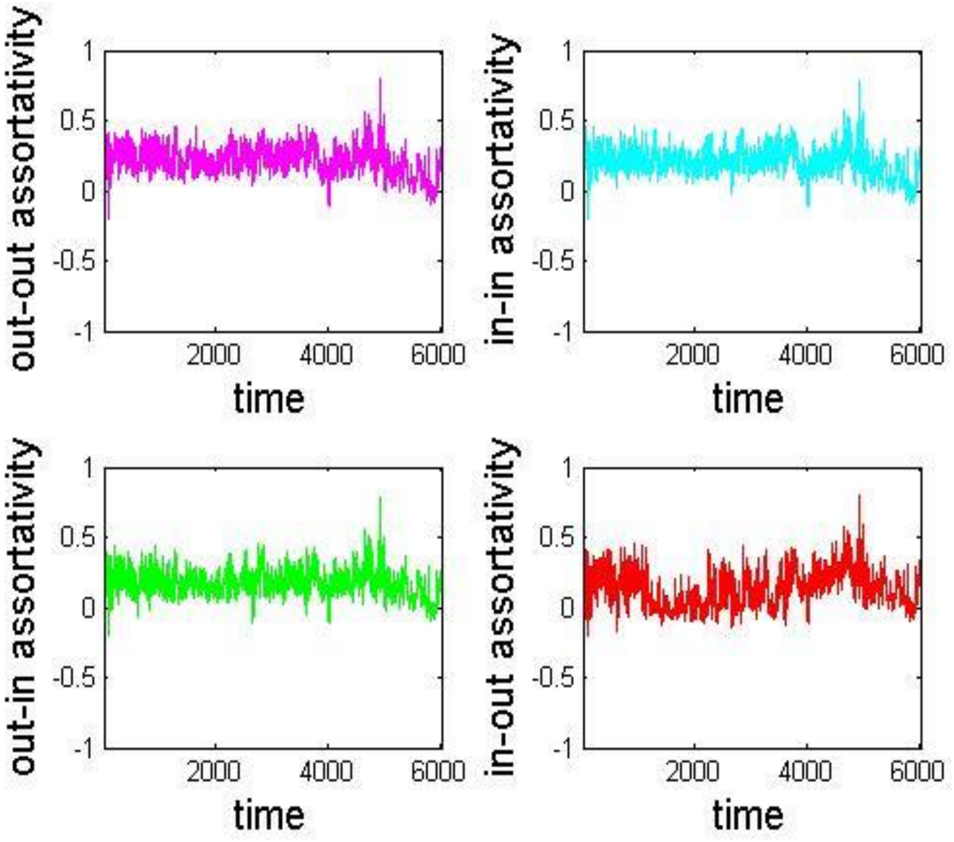
- **Directed and undirected von Neumann entropies change during Friday the 13th mini-crash.**
- **Directed entropy has a better performance in catching minor changes in the network structure than its undirected**



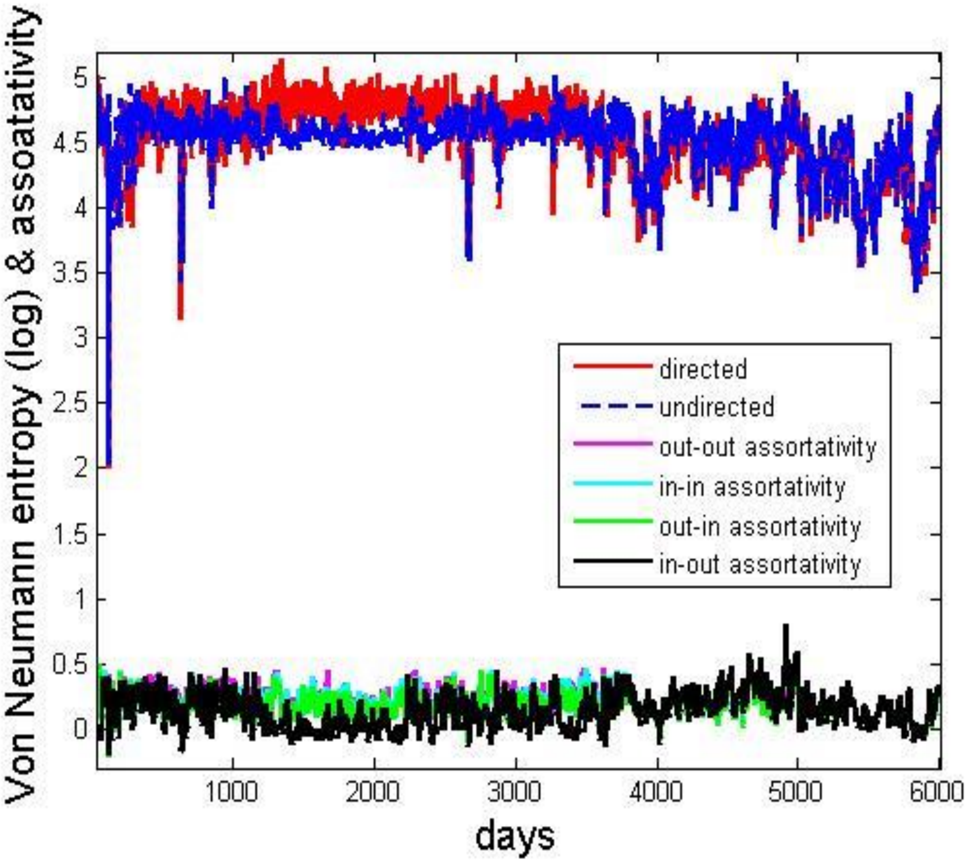
- Connected component number as stock market network evolves.
- Network becomes fragmented during financial crises.
- Von Neumann entropy is able to catch such structure changes.



- Node degree assortativity coefficients as stock market network evolves.
- Top-left: out-degree & out-degree; top-right: in-degree & in-degree; bottom-left: out-degree & in-degree; bottom-right: in-degree & out-degree.
- The smallest value appears on Black Monday, 1987.



- Directed/undirected von Neumann entropies and node degree assortativity coefficients as stock market network evolves.
- The indegree-outdegree assortativity behaves slightly different from the other three.
- Both directed and undirected entropies are able to detect the changes.



- Histograms of added/removed/unchanged edges on different von Neumann entropy components during financial crisis (Black Monday).
- Added/removed edges have the highest probability to happen either when the first component is equal to one or when the other components have small values.
- The probability of an edge remain unchanged becomes highest when the first component is equal to one and does not depend on other two components.