Communities in statistically-validated networks

Chester Curme

In collaboration with:
Irena Vodenska
H. Eugene Stanley
Big picture
Big picture
Big picture

A =

\[
\begin{bmatrix}
1 & 0.60 & 0.01 & 0.95 & \\
0.60 & 1 & \\
0.01 & \\
0.95 & 1 & \\
\ldots & 0.75 & 1
\end{bmatrix}
\]
Big picture

\[ A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \]
Thesis overview

Focus on one experimental setting:

• Associate time series to nodes
Thesis overview

Focus on one experimental setting:

• Associate time series to nodes
• Interactions measured by lagged correlations
  – $A$ is asymmetric
Thesis overview

Focus on one experimental setting:

• Associate time series to nodes
• Interactions measured by lagged correlations
  – $A$ is asymmetric
• Low signal-to-noise
  – Statistical uncertainty is important
Thesis overview

1) Methodology and applications

2) Extend to seasonal time series and use to explain a phenomenon in financial markets
   - Curme, Tumminello, Mantegna, Stanley, Kenett (in preparation).

3) Relate community structures to statistical model performance
   - Curme, Vodenska, Stanley (submitted).

4) Using topic models to explain market movements
   - Curme, Zhuo, Moat, Preis (in preparation).
Thesis overview

1) **Methodology and applications**

2) **Extend to seasonal time series and use to explain a phenomenon in financial markets**
   - Curme, Tumminello, Mantegna, Stanley, Kenett (submitted).

3) **Relate community structures to statistical model performance**
   - Curme, Vodenska, Stanley (submitted).

4) **Using topic models to explain market movements**
   - Curme, Zhuo, Moat, Preis (submitted).
Data

Market returns
News sentiments

40 countries, 80 nodes
Statistical validation

\[ \rho_{ij} \]

Null

\[ r_{ij} \]
Statistically-validated network
Statistically-validated network

News $\rightarrow$ Markets: 12 links
Markets $\rightarrow$ News: 174 links
Community Structures

In undirected networks...
Community Structures

In directed networks...
Community Structures

In directed networks...
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]

Eigenvectors of \( AA^T \)

“Left singular vectors”
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]

"Left singular vectors"

\[ AA^T = \begin{bmatrix} \ldots & 1 & 1 & \ldots \\ \ldots & 0 & 1 & 1 & \ldots \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \end{bmatrix} \begin{bmatrix} \ldots & 1 & 1 & \ldots \\ \ldots & 0 & 1 & 1 & \ldots \end{bmatrix} \]
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]

- Eigenvectors of \( AA^T \)
  - "Left singular vectors"

\[ AA^T = \begin{bmatrix} i & j \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \ldots, & 2, & \ldots \\ \ldots, & 2, & \ldots \end{bmatrix} \]
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]

Eigenvectors of \( AA^T \)

Eigenvectors of \( A^T A \)

“Left singular vectors”

“Right singular vectors”
Finding bipartite substructure

Singular value decomposition:

\[ A = U \Sigma V^T \]

Eigenvectors of \( AA^T \)  \[ \sigma_1 \] \[ \sigma_2 \] \[ \ldots \]  
Eigenvectors of \( A^T A \)  

"Left singular vectors"  "Singular values"  "Right singular vectors"
Finding bipartite substructure

\[ U \begin{bmatrix} \sigma_1 & \cdots & \sigma_n \end{bmatrix} V^T \]
Finding bipartite substructure

<table>
<thead>
<tr>
<th>$U^1$</th>
<th>$V^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>New Zealand</td>
</tr>
<tr>
<td>Mexico</td>
<td>Philippines</td>
</tr>
<tr>
<td>Brazil</td>
<td>Australia</td>
</tr>
<tr>
<td>Chile</td>
<td>Japan</td>
</tr>
<tr>
<td>Argentina</td>
<td>Malaysia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$U^2$</th>
<th>$V^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>France</td>
</tr>
<tr>
<td>Mexico</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Brazil</td>
<td>Sweden</td>
</tr>
<tr>
<td>Chile</td>
<td>Finland</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>Belgium</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$U^3$</th>
<th>$V^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>China News</td>
</tr>
<tr>
<td>Australia</td>
<td>United States News</td>
</tr>
<tr>
<td>Philippines</td>
<td>United Kingdom News</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>Hong Kong News</td>
</tr>
<tr>
<td>Malaysia</td>
<td>Japan News</td>
</tr>
</tbody>
</table>
Making predictions
Making predictions

- Predict sign (+1 or -1) of time series one step ahead
Making predictions

- Predict sign (+1 or -1) of time series one step ahead
- Logistic regression
Making predictions

- Predict sign (+1 or -1) of time series one step ahead
- Logistic regression
- Inputs given by network
Evaluation

• Divide data into training and test sets.
• Construct network and train logistic regressions using training set.
• Evaluate accuracy on test set.
Evaluation

• Divide data into training and test sets.
• Construct network and train logistic regressions using training set.
• Evaluate accuracy on test set.

“ROC Curve”
Evaluation

• Divide data into training and test sets.
• Construct network and train logistic regressions using training set.
• Evaluate accuracy on test set.

“ROC Curve”

“AUC”
Network captures predictive relationships

Receiver operating characteristics

- True positive rate
- False positive rate

- New Zealand (area = 0.67)
- Philippines (area = 0.68)
- Japan (area = 0.71)
- Australia (area = 0.71)
- Malaysia (area = 0.61)
Network captures predictive relationships
Network captures predictive relationships
Restricting to network inputs boosts accuracy
Using communities to recommend inputs

Element of $V^1$
Using communities to recommend inputs

Element of $V^1$
Using communities to recommend inputs

Elements of $V^1$
Using communities to recommend inputs

Elements of $V^2$
Using communities to recommend inputs

Elements of $V^3$
Recommender system interpretation

Bolster against missing links
Recommender system interpretation

...and spurious links.
Summary

• SVN methodology reveals global network of interactions among market movements and financial news sentiment signals.
  – News responds to market movements.
Summary

• SVN methodology reveals global network of interactions among market movements and financial news sentiment signals.
  – News responds to market movements.

• Community structures show collective interactions among groups of countries.
Summary

• SVN methodology reveals global network of interactions among market movements and financial news sentiment signals.
  – News responds to market movements.
• Community structures show collective interactions among groups of countries.
• In this setting, community structures simultaneously form the basis of a “recommender system” for model inputs.
Thank you!

Questions?