Introduction to network science & visualisation¹

Kimmo Soramäki,
Financial Network Analytics

¹ This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.
Introduction to Network Science & Visualization I

Dr. Kimmo Soramäki
Founder & CEO, FNA

www.fna.fi
Agenda

Network Science

- Introduction
- Key concepts

Exposure Networks

- OTC Derivatives
- CCP Interconnectedness

Correlation Networks

- Housing Bubble and Crisis
- US Presidential Election
Network Science and Graphs Analytics

Is already powering the best known AI applications

Knowledge Graph  Social Graph  Product Graph  Economic Graph  Knowledge Graph  Payment Graph
Network Science and Graphs Analytics

“Goldman Sachs takes a DIY approach to graph analytics”
For enhanced compliance and fraud detection

“PayPal relies on graph techniques to perform sophisticated fraud detection”
Saving them more than $700 million and enabling them to perform predictive fraud analysis, according to the IDC
(www.globalbankingandfinance.com, Jan 2016)

"Network diagnostics .. may displace atomised metrics such as VaR”
Regulators are increasing using network science for financial stability analysis.
(Andy Haldane, Bank of England Executive Director)
Further Resources on Network Analytics and Systemic Risk

*Risk Journal* founded by Kimmo Soramäki | [link](#)

*Risk Book* by Kimmo Soramäki and Samantha Cook, FNA’s Chief Scientist | [link](#)

Two-day Training course in London, New York and Singapore, instructed by Kimmo Soramäki | [link](#)
New Way of Looking at Data

- How is data connected with other data?
- How do these connections matter?
- How do complex systems move in time?

For the first time we are able to measure and model this!
"Systems with rich interactions between the components of the system"

e.g. financial markets, payment systems, road systems, friendship networks, ... almost every socio-economic system.
Main Modes of Analysis

- Top Down Analysis
- Bottom Up Analysis
- Features of Data
- Agent Based Models
Top Down Analysis

Typical use cases:

- Systemic risk analysis
- System monitoring
- System design
- System stress testing
- Clustering/Classification
- Early warning
- Anomaly detection
Typical use cases:

- Criminal investigation
- Terrorist networks
- Money laundering
- KYC & KYCC
- Fundamental investment analysis
- Supply chain analysis
Typical use cases:

- AI/ML
- Fraud algorithms
- Recommendation engines
- Algorithmic investment

FNA Research: [Comparison of Graph Computing Platform Performance](#)
Typical use cases:

- Central Counterparty Clearing
- Payment Systems
- FX Settlement
- Financial Markets
- Housing Markets
Journey

1. Advanced Analytics
   - Understand interconnectedness, data visualization
   - Identify Risks Concentrations

2. Monitoring
   - Detect risks and anomalies in real-time
   - Provide Early Warning

3. Simulation
   - What-if analysis, failure simulations & remediation scenarios
   - Predict Outcomes
Descriptive Analytics
Diagnostic Analytics
Predictive Analytics
Prescriptive Analytics

What happened?
Why did it happen?
What will happen?
What should happen?

Hindsight
Insight
Foresight

Value
Difficulty
Types of Networks
Network Concepts

Constituents

- Networks (graphs)
- Nodes (vertices)
- Links (ties, edges or arcs)

Links can be

- Directed (arcs) vs undirected (edges, ties)
- Weighed vs unweighted

Graph + properties = Network
Some Graph Types

- Trivial Graph
- Empty Graph
- Undirected Graph
- Complete graph, K4
- Complete graph, K7
- Directed Graph
Simple Graphs / Non-random Graphs

Star

Lattice

Ring
Simple Graphs / Non-random Graphs

Bipartite

Tree
Random Graphs

Random
(Erdos-Renyi)

Scale-free
(Barabasi-Albert)
Projection Networks

Left Projection

Right Projection
Centrality measures importance of nodes (or links) in a network. Depends on process that takes place in the network!

**Trajectory**
- Geodesic paths (shortest paths)
- Any path (visit no node twice)
- Trails (visit no link twice)
- Walks (free movement)

**Transmission**
- Parallel duplication
- Serial duplication
- Transfer

DHL Package = Transfer via shortest path  
Money = Transfer via random walks  
Virus = Serial duplication via paths  

etc.
Common Centrality Metrics

PageRank

Betweenness

Closeness

PageRank within Cluster
Community Detection

Often networks are large and complex and we want to simplify, categorize and label nodes into meaningful groups.

Community detection is an algorithmic way of doing this, and there are numerous methods available.

- Unsupervised learning, how do we know result is correct.
  What is correct?
- Which algorithm to choose?
- Some algorithms detect well large, but not small communities
- Is it a community or a cluster of several?
- What about overlapping communities?

Still more an art than a science. Try what works?

FNA Research: Overview and Comparison of Community Detection Algorithms
FNA
Filtering
www.fna.fi
Filtering

Often networks are large and complex and we want to filter out noise. Filtering techniques give solutions.
Top down analysis
Exposure Networks

www.fna.fi
First Financial Networks

Fedwire Interbank Payment Network (Fall 2001) was one of the first network views into any financial system.

Of a total of around 8000 banks, the 66 banks shown comprise 75% of total value. Of these, 25 banks completely connected

The research was subsequently used e.g. in congressional hearings to showcase the type of information that should be collected by financial institutions after the financial crisis.

Who pays whom?

Our data contains:

<table>
<thead>
<tr>
<th>date</th>
<th>buyer</th>
<th>seller</th>
<th>amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 July 2018</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>26 July 2018</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>26 July 2018</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>26 July 2018</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Who pays whom?

Payment from 1 to 3
Who pays whom?

Payments from 3 to 4 and 5 to 2
Who pays whom?

Payment from 3 to 4
Who pays whom?

Payments from 5 to 6 and 6 to 3
Who pays whom?

More payments ....

Thicker, darker links represent higher link weights, i.e., more payments.
Background
As part of global regulatory reforms, the Hong Kong Monetary Authority (HKMA) started in 2013 to collect derivatives data through the Hong Kong Trade Repository.

Objective
Bring more transparency to derivatives markets using the data collected by trade repositories.

Insights
Initial framework for analysing this new data source to assess the financial stability of the market and potential risks. This includes development of maps for the chain of exposures between institutions.

HKMA: A first analysis of derivatives data in the Hong Kong Trade Repository
Interconnectedness in the Global System of CCPs
Scope of Analysis

Comparison with BIS "Analysis of Central Clearing Interdependencies" (2017)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CCPs</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td>Jurisdictions</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Clearing Members</td>
<td>n/a</td>
<td>811</td>
</tr>
<tr>
<td>Parents Organizations</td>
<td>307</td>
<td>563</td>
</tr>
<tr>
<td>Roles</td>
<td>5 (member, settlement, LOC, ...)</td>
<td>1 (member)</td>
</tr>
</tbody>
</table>
We see CCPs (diamonds) and their members (circles) from different regions:

- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

On subsidiary level, we see a tight core with peripheral CCPs and a number of completely disconnected CCPs from Latin America and Middle East.
210 Banking Groups

Largest (# of entities):

1. Citigroup (19)
2. Morgan Stanley (13)
3. Goldman Sachs (12)
4. JPMorgan Chase (12)
5. Bank of America (12)
6. HSBC (11)
7. Credit Suisse (10)
8. Deutsche Bank (10)
9. Nomura (9)
10. Banco Santander (8)
We see CCPs (diamonds) and their members (circles) from different regions:

- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

On parent level we see a completely connected network dominated by a core consisting of CCPs from North America and Europe and global banks.
CCP Interconnectedness on Subsidiary vs Parent Level - Example

**Subsidiary Level**
(Connected to 3 CCPs)

**Parent Level**
(Connected to 23 CCPs)
CCP Interconnectedness on GSIB Level

<table>
<thead>
<tr>
<th>Bank (Parent)</th>
<th># of FMIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citigroup</td>
<td>21</td>
</tr>
<tr>
<td>DEUTSCHE BANK</td>
<td>21</td>
</tr>
<tr>
<td>JPMorgan Chase &amp; Co.</td>
<td>19</td>
</tr>
<tr>
<td>BNP PARIBAS</td>
<td>18</td>
</tr>
<tr>
<td>Bank of America</td>
<td>17</td>
</tr>
<tr>
<td>HSBC</td>
<td>17</td>
</tr>
<tr>
<td>Morgan Stanley</td>
<td>16</td>
</tr>
<tr>
<td>Societe Generale</td>
<td>16</td>
</tr>
<tr>
<td>The Goldman Sachs</td>
<td>15</td>
</tr>
<tr>
<td>Credit Suisse</td>
<td>14</td>
</tr>
</tbody>
</table>
Contagion - CCP Disruption

A disruption in a CCP would affect all of that CCP’s clearing members, thereby affecting the other CCP’s to which the affected CCP’s members belong, possibly creating a cascading cycle as disruption is propagated across members and CCPs.
Footprint of CCPs - SHCL

SHCL's 56 members are connected to 23 other CCPs.

Most members are domestic with a few large global banks based in EU & US.

The most connected CCP is HKEx.
Footprint of CCPs - OCC

OCC's 89 members are connected to 27 other CCPs.

The membership is mostly US with a significant EU base.

The most connected CCP's are DTCC and CME.
LCH Ltd 100 members are connected to 27 other CCPs.

The membership is mostly European with a significant US base.

The most connected CCP is LCH SA and ICE-CLEAR EUROPE.
A member disruption can be felt by up to **458** banking groups or banks (of total of 563, or 80%) that are members of the same CCP as the stricken group.

<table>
<thead>
<tr>
<th>Banking Group</th>
<th># banking groups connected via a CCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deutsche Bank</td>
<td>458</td>
</tr>
<tr>
<td>Citigroup</td>
<td>446</td>
</tr>
<tr>
<td>Morgan Stanley</td>
<td>442</td>
</tr>
<tr>
<td>BNP Paribas</td>
<td>423</td>
</tr>
<tr>
<td>Goldman Sachs</td>
<td>412</td>
</tr>
<tr>
<td>HSBC Holdings</td>
<td>402</td>
</tr>
<tr>
<td>JPMorgan Chase</td>
<td>388</td>
</tr>
<tr>
<td>Bank of America</td>
<td>382</td>
</tr>
<tr>
<td>Credit Suisse</td>
<td>348</td>
</tr>
<tr>
<td>Société Générale</td>
<td>340</td>
</tr>
</tbody>
</table>
Deutsche Bank Group participates in 21 CCPs (of 29 mapped).

458 other banking groups or banks are members of these CCPs.
Morgan Stanley participates in 16 CCPs (of 29 mapped).

442 other banking groups or banks are members of these CCPs.
**Transactions & Similarity Based Networks**

**Transaction:** payment, trade, exposure, supply, flow, ...

**Similarity:** correlation, partial correlation, granger causality, transfer entropy, ...

Stavroglou et al (2016)

*Causality Networks of Financial Assets*
Typical view of cross asset correlations

**Correlation Matrix Over the Last 15 Years (2001-2015)**

<table>
<thead>
<tr>
<th>Equity</th>
<th>Fixed Income</th>
<th>Alternative Strategies</th>
<th>Alternative Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Cap</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid Cap</td>
<td>0.93</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Small Cap</td>
<td>0.88</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Intl</td>
<td>0.88</td>
<td>0.85</td>
<td>0.79</td>
</tr>
<tr>
<td>Emerg. Meta</td>
<td>0.78</td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Fixed Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corp.</td>
<td>0.17</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td>High Yield</td>
<td>0.66</td>
<td>0.71</td>
<td>0.68</td>
</tr>
<tr>
<td>Treas.</td>
<td>-0.36</td>
<td>-0.35</td>
<td>-0.39</td>
</tr>
<tr>
<td><strong>Alternative Strategies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long/Short</td>
<td>0.75</td>
<td>0.79</td>
<td>0.72</td>
</tr>
<tr>
<td>Mkt Neutral</td>
<td>0.37</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>Event Driven</td>
<td>0.64</td>
<td>0.70</td>
<td>0.84</td>
</tr>
<tr>
<td>FI Arbitrage</td>
<td>0.42</td>
<td>0.48</td>
<td>0.37</td>
</tr>
<tr>
<td>Mgd Futures</td>
<td>-0.13</td>
<td>-0.10</td>
<td>-0.12</td>
</tr>
<tr>
<td><strong>Alternative Assets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Estate</td>
<td>0.06</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Currency</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Commodities</td>
<td>0.32</td>
<td>0.38</td>
<td>0.33</td>
</tr>
</tbody>
</table>

**Color Legend**
- **High (0.9-1.0)**: High Diversification
- **Moderate High (0.7-0.9)**
- **Moderate (0.3-0.7)**
- **Low (0.0-0.3)**
- **Negative (<0.0)**
Correlation Networks

Interconnectivity of markets has increased

We need to be able to understand correlations structures of much larger scale.

Networks help develop intuition, and understand stress tests.
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BND</td>
<td>Total Bond Index</td>
</tr>
<tr>
<td>DBC</td>
<td>Commodities</td>
</tr>
<tr>
<td>DIA</td>
<td>DJIA</td>
</tr>
<tr>
<td>DXJ</td>
<td>Japan Stocks (in JPY)</td>
</tr>
<tr>
<td>EEM</td>
<td>Emerging Markets</td>
</tr>
<tr>
<td>EFA</td>
<td>EAFE</td>
</tr>
<tr>
<td>EMB</td>
<td>EMBI</td>
</tr>
<tr>
<td>EPP</td>
<td>Asia ex Japan</td>
</tr>
<tr>
<td>EWG</td>
<td>Germany</td>
</tr>
<tr>
<td>EWI</td>
<td>Italy</td>
</tr>
<tr>
<td>EWJ</td>
<td>Japan</td>
</tr>
<tr>
<td>EWQ</td>
<td>France</td>
</tr>
<tr>
<td>EWU</td>
<td>UK</td>
</tr>
<tr>
<td>FXB</td>
<td>GBP</td>
</tr>
<tr>
<td>FXC</td>
<td>CAD</td>
</tr>
<tr>
<td>FXE</td>
<td>EUR</td>
</tr>
<tr>
<td>FXI</td>
<td>China</td>
</tr>
<tr>
<td>FXY</td>
<td>JPY</td>
</tr>
<tr>
<td>GDX</td>
<td>Gold Miners</td>
</tr>
<tr>
<td>GLD</td>
<td>Gold</td>
</tr>
<tr>
<td>IEF</td>
<td>Barclays 1-7Y US</td>
</tr>
<tr>
<td>IYR</td>
<td>Real Estate</td>
</tr>
<tr>
<td>JNK</td>
<td>High Yield Bonds</td>
</tr>
<tr>
<td>LQD</td>
<td>Corp Bonds</td>
</tr>
<tr>
<td>SLV</td>
<td>Silver</td>
</tr>
<tr>
<td>SPY</td>
<td>S&amp;P 500</td>
</tr>
<tr>
<td>TIP</td>
<td>TIPS</td>
</tr>
<tr>
<td>TLT</td>
<td>20Y+ Gov't</td>
</tr>
<tr>
<td>USO</td>
<td>Oil</td>
</tr>
<tr>
<td>UUP</td>
<td>USD Index</td>
</tr>
<tr>
<td>VGK</td>
<td>Europe</td>
</tr>
<tr>
<td>VPL</td>
<td>Asia</td>
</tr>
<tr>
<td>VXX</td>
<td>VIX ST Futures</td>
</tr>
<tr>
<td>XIU</td>
<td>TSX 60</td>
</tr>
<tr>
<td>XLB</td>
<td>Materials</td>
</tr>
<tr>
<td>XLE</td>
<td>Energy</td>
</tr>
<tr>
<td>XLF</td>
<td>Financials</td>
</tr>
<tr>
<td>XLK</td>
<td>Tech</td>
</tr>
<tr>
<td>XLU</td>
<td>Utilities</td>
</tr>
<tr>
<td>CSJ</td>
<td>Barclays 1-3Y US</td>
</tr>
<tr>
<td>FXF</td>
<td>CHF</td>
</tr>
</tbody>
</table>
We can view any matrix as a network.

We encode correlations as links between the correlated nodes/assets.

Red link = negative correlation
Black link = positive correlation

However, this simple encoding does not give us much.
Transactions & Similarity Based Networks

Not all correlations are statistically significantly different from 0.

Absence of link marks that asset is not significantly correlated (here at 95% level).

Due to the large number of estimates, we also need for multiple comparisons correction. Eg. Bonferroni or FDR.
We can use network layouts to better detect patterns from noise.

E.g. we can try a Force-Directed network layout to identify clusters.
Next, we identify the Minimum Spanning Tree and filter out other correlations (Mantegna, ‘99).

We need a distance function, here we look at maximum spanning tree with distance function: abs(cor)

This shows us the backbone correlation structure.
We use a radial tree layout algorithm (Bachmeier et al. ‘05) that places the assets so that:

- Shorter links in the tree indicate higher correlations
- Longer links indicate lower correlations

As a result, we also see how the assets cluster by asset class.
Radial Tree Layout

Focus on the links in the Spanning Tree to highlight clustering structure.

Node color indicates last daily return

- Green = positive
- Red = negative

Node size indicates magnitude of return

Bright colors are VaR exceptions
Financial Cartography

Coordinate system

-> layout algorithm

**System for visual encoding of map data**

-> node sizes & colors

**Dimensionality reduction & filtering**

-> minimum spanning tree
Use Case: Monitoring Housing Markets

In this example we look at US house prices across states. We see the US states as nodes and strong correlation between house prices as link. In 2000 the tree is very spread out and prices are going slightly up. This is a time when ABS are developed with the assumption that real-estate risk can be diversified across US states.
In 2003 we start to see some strong upward movements in prices in states like Nevada and we see a big cluster of bumper returns in Florida and states that have strong correlations with it.
As we move into 2004, into the peak of the housing bubble we see that most states now have outlier price changes and Nevada for example has an almost 12% rise in house prices in one quarter.
As we move into 2005 we look at the length of the tree. It measures the overall correlations in this system. The shorter (smaller value) the tree, the stronger the correlations. We see that the tree has been getting shorter and shorter. The assumptions behind diversification of ABS getting eroded.
In summer 2007 the housing bubble is over and we see the first negative outlier in Florida. Most of the system has become red, except a green branch on the left. We also see that the tree has been getting shorter and shorter, reaching new lows each quarter. The system is becoming highly coupled.
In 2009 we reach the peak crisis. The system has become largely red with many central states as negative outliers.

We can look at another metric on the left. Systematic risk measures how much changes in the system are driven by the largest single factor, and how much by idiosyncratic - state level - factors. We see that the system is quickly becoming governed by a single factor affecting all states.
The same dynamics continue with the "double dip" in 2011.
In Spring 2012 we see the first positive outlier in North Dakota, likely drive by the fracking boom. The rest of the system is still mostly negative.
If we fast forward to the latest observation, July 2015, we see a period of positive changes in prices with outliers scattered across the network.

We also see both systematic risk and correlations at their peak. We have not returned to the pre-bubble system state but are in a very risky territory still.
We can see this clearly by looking at the size of the tree.

First in 2010.
Then at the peak of the bubble in 2005.
Then at the peak of the crisis in 2009.
And now.

The tree has shrunk during the whole period. The correlations are now stronger than ever.

Such slow moving change is hard to notice when focusing on daily events. Like in the story of the frog put in water that is gradually heated.
Dr. Kimmo Soramäki  
Founder & CEO  
FNA - Financial Network Analysis Ltd.

kimmo@fna.fi  
tel. +44 20 3286 1111

Address  
4-8 Crown Place  
London EC2A 4BT  
United Kingdom
FNA

Introduction to Network Science & Visualization II

Dr. Kimmo Soramäki
Founder & CEO, FNA

www.fna.fi
Agenda

Financial Crime & Cyber Risks

- Fraud, AML & KYCC
- DDoS Attacks
- Related Parties Analysis

Financial Market Infrastructures

- Monitoring Members
- Designing liquidity efficient FMIs
- Predicting Liquidity
- Detecting Anomalous transactions
Intensified regulatory pressures has increased the number of false positives generated by existing software solutions.

It is increasingly difficult for banks and financial institutions to quickly identify fraudsters.

The cost is $50 billion in fraudulent transactions happening each year.
Gartner's Layered Model of Fraud Prevention

- **Layer 1** Endpoint Centric: Analysis of users
- **Layer 2** Navigation Centric: Navigation behavior
- **Layer 3** Account Centric: Anomalies by channel
- **Layer 4** Cross-Channels: Anomalies cross-channel
- **Layer 5** Graph Centric: Analysis of links in data
Background
In 2017 global banks were fined £5B for failures to detect and address money laundering. Current methods are insufficient in identifying money laundering, and costly in terms of large amounts of manual labour needed.

Method
Payments form networks which can be automatically analyzed by network science algorithms. Existing research on large datasets proved that particular graph properties are good predictors of fraudulent transactions.

Benefits
Graphs improve fraud detection by eliminating false positives and identifying true positives more accurately - saving time and money.

Molloy et al (2017), Graph Analytics for Real-time Scoring of Cross-channel Transactional Fraud
We create a link between two account holders if a payment is made between them. Over time these links accumulate to a network.

We can update the network in real-time as payments are being processed.
A payment request comes from Traci to pay to Hazel. Traci has never paid to Hazel before, but has paid to Janet, who has paid to Hazel.

Thus, the payment is relatively normal.
Another payment request comes from Traci to pay to Buffy. Traci and Buffy are very far apart in the network making the payment unusual.

We can operationalise how 'normal' the payment is with a network measure of distance, which in this case is 7 (and in previous case it was 2).
We can also use other graph features of the data in our fraud models, such as centrality.

A node is more central if it has over time accumulated more non-suspicious payment relationships. This is visualized as node size in this dashboard.
AML and Suspicious Activity
Challenge
Understand corporate interconnections for due diligence, fraud/criminal investigation, KYC, KYCC.

Current Situation
Manual investigation.

Solution
FNA has built a connection to Companies House register to automatically build graphs for any UK company.

Benefits
Save time and achieve a systemic view of company interconnectedness.
6 May 2018

NETWORK

CAMBRIDGE ANALYTICA(UK) LIMITED

Alexander James Ashburner NIX
DDoS Attacks

www.fna.fi
Detect Anomalies in Cyber Networks in Real-time
Identify Patterns of DDoS attacks
Monitoring & Simulating FMIs & their Members
Use Case: Understanding Interconnectedness

Mapping SWIFTs global payment network

**Background**
SWIFT message services are used by over 11,000 financial institutions in more than 200 countries. SWIFT was interested in what insights could be drawn from the "Big Data" that it collects when transmitting messages between financial institutions.

**Objective**
Analyse the payment networks created by flows of SWIFT MT103 messages to draw insights about macroeconomic, geo-political and compliance topics.

**Insights**
Analysis of the SWIFT payment networks revealed a number of insights, including the phenomena of de-risking, payment country blocks relevant for sanctions analysis and how geopolitics shape them, and estimated the cost of the financial crisis at $5 Trillion. The outcome of the research was presented at Sibos 2014 by SWIFT CEO Gottfried Leibbrandt.

SWIFT Institute Research Paper: [The global network of payment flows](#)
Research Paper: [The Impact of Anti-Money Laundering Regulation on Payment Flows](#)
Big data problem: Three billion messages exchanged among banks in 231 countries.

We focus on aggregated links among countries.

Analysis and visualization a challenge. We don’t want to show much information (as on this picture).
The number of messages is 5.5% lower (post-crisis than they would have been had the pre-crisis trend continued unabated throughout the entire period.
Of the 1054 links gained until 2007, in 74% one (or both) were rated as medium or low on the United Nations Human Development Index.

Of the 990 links lost after 2007, 80% involved at least one country listed as an offshore financial center.
Are there meaningful subgroups among the countries?

Can we group the countries so that messages are sent mostly within groups?

Modularity - measure of concentration of links within communities vs. between communities.

Example Communities
We overlay the Minimum Spanning Tree showing the strongest links for each country.

We see US, Germany and France as large hubs.

Background
The Central Bank of Colombia has been using balance sheet and regulatory reporting data to understand the liquidity and solvency of participants in the Colombian financial system. However, the analysis is time consuming and the data comes months late.

Objective
Using network analysis of data from the interbank payment system would allow the Bank to get early warning about risks substantially faster.

Outcomes
Using the FNA Platform, the Bank is now able to monitor its banking system in near real time. Automatic alerts notify the bank of any abnormal behavior in the network. Furthermore, automated stress tests where they fail the two largest participants in the network help to understand the riskiness of the system.

Use Case: Monitoring Liquidity and Solvency of FIs

**Background**
Bank of Korea, South Korea's central bank, was looking for ways to have early warning about intraday liquidity problems in its systemically important BoK-Wire+ interbank payment system.

**Objective**
To develop methods to predict the liquidity position of each member in BoK-Wire+ in real-time, as well as measure the importance of member in terms of the liquidity and operational risk a liquidity shortage would cause.

**Outcomes**
Bank of Korea and FNA developed a framework for identifying bank's liquidity problems in real time and using FNA's SinkRank algorithm to identify most critical banks. The results were published as a research paper.

BoK Research paper: [Network Indicators for Monitoring Intraday Liquidity in BOK-Wire+](#)
Journal article: [SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems](#)
Key issue in payment system is that each bank is dependent on incoming funds to make their own payments.

Objective of this work was to develop measures for ongoing monitoring of systemic risk in payment systems.

Analytics need to be operationalized into a robust and repeatable decision making framework
PSLI (Payment System Liquidity Indicator) is the ratio of projected liquidity demands and projected liquidity supply:

\[
PSLI_{ijt} = \frac{\text{pending}_{ijt} + \text{KRdebits}_{ijt} + \text{ERdebits}_{ijt}}{\text{balance}_{ijt} + \text{limit}_{ijt} + \text{KRcredits}_{ijt} + \text{ERcredits}_{ijt}}
\]
Expected credits and debits are estimated on the basis of a regression model.

The model takes into account the value already settled on the given day, effects related to reserve maintenance and to US holidays and the trade values of bonds and spot exchange.

The model has a good fit.
Payments move liquidity in the network.

Payments take place on links at some given frequency that can be measured (e.g., based on historical or projected flows).

We are concerned on operational failures. The sink can receive payments but cannot send any.

**Example:***

Let’s start by considering one unit of liquidity that is moved by payments in a simple system of three banks.

At the time of analysis, the unit of liquidity can be at either A, B or C.

What is the distance of the unit to the different ‘sink nodes’?

---

Soramaki and Cook (2013) *SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems*
SinkRank is suited for Predictive Modeling

Given an observed distribution of liquidity, and a historical pattern of payment flows

- What is the distribution if bank A has an operational disruption at noon?
- Who is affected first?
- Who is affected most?
- How is Bank C affected in an hour?

Valuable information for decision making

- Crisis management
- Participant behavior
Using Network Simulations to Design FMIs
What are Simulations

Methodology to understand complex systems – systems that are large with many interacting elements and or non-linearities (such as payment systems)

In contrast to traditional statistical models, which attempt to find analytical solutions

Usually a special purpose computer program is used that takes granular inputs, applies the simulation rules and generates outputs

Take into account second rounds effects, third round, ...

Inputs can be stochastic or deterministic. Behavior can be static, pre-programmed, evolving or co-learning
Short History of FMI Simulations

1997 : Bank of Finland
Evaluate liquidity needs of banks when Finland’s RTGS system was joined with TARGET

2000 : Bank of Japan and FRBNY
Test LSM features for BoJ-Net/Fedwire

2001 - : CLS approval process and ongoing oversight
Test CLS risk management
Evaluate settlement’ members capacity for pay-ins
Understand how the system works

Since: Bank of Canada, Banque de France, Nederlandsche Bank, Norges Bank, TARGET2, and many others

2010 - : Bank of England, CHAPS
Evaluate alternative liquidity saving mechanisms
Use as platform for discussions with banks
Analytical models need to make many simplifying assumptions.

Problem with static simulations based on historical records is that behavior of banks is not taken into account.

This behavior may have material impact on results in most simulation questions, eg:

- When system features are changed
- In stress situations
- As a reaction to other behavioral changes

-> Agent Based Modeling
Agent Based Models

Each agent has a set of rules that define its behavior
-> system level emergent behavior

Choices

- design of rules
- homogeneous vs heterogeneous agents
- static vs learning agents

Pros

- ability to model complex behaviors
- flexible and realistic
- real systems are sensitive to details of implementation

Cons

- time consuming to set up
- need many input parameters
- results very sensitive to modeling assumptions

Agent Based Models

Existing literature very short

- Arciero et al (2009)
- McLafferty-Denbee (2013)
- Soramäki and Cook (2015)

Results

- Behavior has material impact on results
- Behavior increases delays (or moves away from social liquidity/delay optimum)

Questions

- Money market model
- One vs multiperiod, learning vs fixed populations
- Which payments are discretionary / known
- What is the cost of liquidity/delay tradeoff
- Human vs machine behavior
FNA

Data Needs

www.fna.fi
Data Needs

**Historical transaction data**

- From interbank payment systems
- At minimum: date, time, sender, receiver, value
- More data on type of payment, economic purpose, second tier (if any), type of institution, etc. useful

**Representative transaction data**

- Based on aggregates or sampling of real data
- Based on a network model (defining bilateral flows)
- Assumptions about:
  - Timing of payments
  - Value distribution
  - Correlations (eg do larger participant send larger payments)
- System stability (net flows over longer times)
**Background**
Real transaction data held by FMI's and Banks is highly confidential and hard to get access to. Also as historical records, it cannot be used as input data in simulations about future infrastructures that may process very different flows.

**Method**
FNA has developed and vetted in several client projects a method for generating representative transaction data that contains all known network and statistical properties of the real transaction data.

**Outcome**
The cost of simulations is lower and the speed at which projects can be completed is higher - lowering the entry barriers to start simulations. Often results with representative data prove the value of the simulations and real data can be used for sensitivity analysis.

---

**Summary of Research** [slides]
1. Evaluate Changes in Environment
2. Stress Testing & Scenarios
3. Payment System Design
4. Model Validation
5. Monitoring
Framework for evaluating trade-off between liquidity and delay
Motivation

Settlement in RTGS consumes large amounts of cash

Cash/liquidity is not free

Customers' increasing demands for faster payments means delays cost too

The tradeoff is not going away even with Blockchain

There is no natural co-operative outcome

A complex system, hard to analyse

**Bottom line impact**

Baek, Soramäki and Yoon (2014).
J. of Financial Market Infrastructures
There is an amount of liquidity each bank must have to complete settlement. And another amount above which adding more has no impact.
Liquidity

Delay

End-of-day net settlement, 5-20%

RTGS without delays, 100%

Delay

Liquidity
Initially reducing liquidity has little impact. Eventually the system becomes increasingly gridlocked (chaotic).


A convex shape for trade-off
Institutional aspects matter

System with many time-critical payments or other constraints
Network topology matters

- Small number of highly connected participants
- Large number of evenly connected participants
In theory we have some optimal allocation of liquidity that depends on the costs of liquidity and delays.

Where should we be on this curve?
Liquidity distribution matters

But the liquidity will likely not be optimally allocated among participants
Liquidity distribution matters

So we find a point that has more delays than optimal.
Incentives matter

In addition, if liquidity is sufficiently costly, participants are likely use less liquidity than optimal for the whole system.

Galbiati and Soramaki (2011). J. of Econ. Dynamics and Control

How do we foster co-operation?

Throughput guidelines, liquidity groups, building confidence, norms, monitor free-riding, etc.
Liquidity Saving Mechanisms (LSMs)

We can find new trade-off curves using different system designs.

Leinonen and Soramaki (2011).
Bank of Finland WP
Liquidity Saving Mechanisms (LSMs)

Which will allow the system to settle faster with a given amount of liquidity
Liquidity Saving Mechanisms (LSMs)

Or operate with less liquidity for a given amount of delays
How about stress situations?

Stress situations reduce liquidity in the system. LSM help alleviate the impact.

Bech and Soramaki (2002). E-money & Payment Systems Review
Which LSM is the best?

Need to simulate ...
Simulations with Payments Canada for modernization of Canada's Payment System

**Problem**: FIFO order may 'block' settlement if a large payment is at the front of the queue.

Bypass FIFO tries to settle payments down the queue and selects the first one that it finds.

**Example**: A has liquidity available 200. A has queued payment: 300, 150 and 100.

Payment 150 can be settled.
Bilateral Offset

Problem: Liquidity may be unnecessarily used when receiver has payments to sender in its own queue.

Bilateral Offset finds payments from receiver's queue that can be offset with sender's payment.

Example: A has 200 liquidity available and a payment of 500 to B. B has payments to A in queue: 300, 150 and 100.

Payment 500 can be settled offsetting 300 of the 500 against B's payment and the remaining 200 with available liquidity.
Queue Optimization

**Problem:** A system may become gridlocked

Queue optimization tries to find a subset of payments that can be settled by all banks with available liquidity through multilateral netting.

**Example:** A cycle where no bank has liquidity but all payments could be made simultaneously.
Summary

**Bypass** alone brings good benefits in reducing delays (or liquidity at a given delay level).

Other LSM's further improve on it.

**Bypass + Bilateral offset** and **Bypass + Queue Optimization** are equally efficient.

Having all LSM's running, brings best outcome from a liquidity-delay perspective.
Typical use cases:

- Systemic risk analysis
- System monitoring
- System design
- System stress testing
- Clustering/Classification
- Early warning
- Anomaly detection
Typical use cases:

- Criminal investigation
- Terrorist networks
- Money laundering
- KYC & KYCC
- Fundamental investment analysis
- Supply chain analysis
Network Features of Data

Typical use cases:

- AI/ML
- Fraud algorithms
- Recommendation engines
- Algorithmic investment

FNA Research: Comparison of Graph Computing Platform Performance
Typical use cases:

- Central Counterparty Clearing
- Payment Systems
- FX Settlement
- Financial Markets
- Housing Markets
Dr. Kimmo Soramäki
Founder & CEO
FNA - Financial Network Analysis Ltd.

kimmo@fna.fi
tel. +44 20 3286 1111

Address
4-8 Crown Place
London EC2A 4BT
United Kingdom