Data Issues for Network Analysis

Seeking to examine existing data gaps that hinder the analysis of systemic risks, and discussing potential solutions

Andrei Kirilenko
Commodity Futures Trading Commission

Two parts:

1. Basics of network analysis.

2. Data for network analysis.

Network analysis

Network analysis is based on a representational formalism borrowed from graph theory.

A finite set of **identifiable** entities is represented as a set of nodes.

Pairwise relationships among nodes are represented via edges.

The network is represented by a graph defined as the set of nodes together with the set of pairwise relationships among them.

When is a node a node?

To avoid misleading conclusions, the set of nodes should be defined to include **all distinct entities** that are capable of participating in the relationship under study;

Where no such set of entities can be uniquely identified, it is possible that a finite network representation **will not be** appropriate.

An **alternative framework** such as a continuous spatial representation may prove more fruitful.

Network framework: good for analyzing discrete sets.

When is an edge and edge?

Binary representation is appropriate when a relationship reflects either a **general tendency toward or potential for interaction**.

The relationship under study needs to be sufficiently **stable** to be well-approximated by a constant function over the period of interest.

Episodic relationships that may occur at variable rates require **specialized treatment** – e.g., a weighted graph representation.

Network framework: good for analyzing stable relationships.

What is the time scale?

In determining appropriate node and edge representations, it is vital to consider the **time scales** on which the processes of interest unfold.

Failure to consider dynamics can lead to **extremely misleading** results.

For example, for a given network, everyone may become infected or no one may be infected, depending on **the edge duration and time of onset** .

Network framework: good for analyzing static problems.

To correctly apply network technology, we need to know:

What are the nodes? What are the edges? What is the time scale?

For the analysis of systemic risks:

Nodes are financial institutions – all, systemic, banks?

Edges are linkages – transactions, balance sheet linkages, risks?

Time scale is before, during and after a systemic event – transaction time, reporting time, “risk” time?

Data requirements are different mainly for different definitions of the edges.

Transactions: Need to know the identities of the buyers and sellers.

Balance sheet linkages: Need to know the identities of counterparties.

Risks: Need to know the ultimate holders of risks.

CFTC Research on Trading Networks

Transactions with identity of traders – audit trail data.

Patterns of order execution in electronic limit order markets.

Prototype of automated surveillance of electronic markets.

Used to analyze strategies of algorithmic and high frequency traders.

Many technological and analytical issues solved, many remain (e.g., network formation – hyperbolic time).

Good for systemic events: possible, if they include liquidity issues, May 6, 2010.