When Simple Networks Fail: Characterising Social Networks Using Simulation

Bruce Edmonds
Centre for Policy Modelling
Manchester Metropolitan University
bruce@edmonds.name

Edmund Chattoe
Department of Sociology
University of Oxford
edmund.chattoe@sociology.ox.ac.uk

Plan of the talk

- Really only a prospectus for the paper!
- A concern with network measurement and explanation
- A case study of a simulation of a Peer-to-peer system
- A proposed solution

Background/Terminology

- **Generative process**: Gives rise to measured social network (Triad balance, preferential attachment …)
- **Distributive process**: Gives rise to distribution of attributes over network (Information transmission)
- Only conceptually separate (favours)

"Typical" SNA and measurement

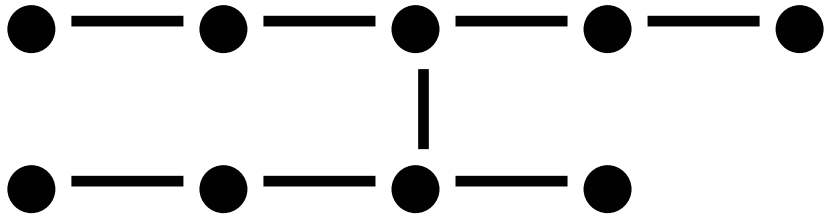
- Associations between network measures and attributes (or other network measures)
- **Generative challenge**: What can we infer from associations about the underlying generative process?
- **Distributive challenge**: What can we infer from associations about the underlying distributive process?
- Not just criticisms of statistics revisited

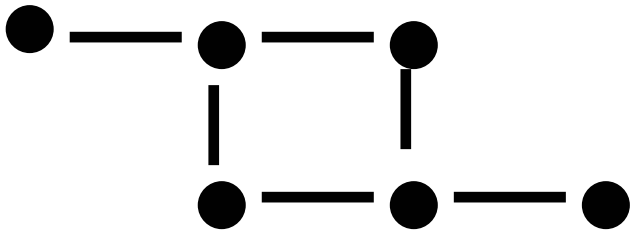
Special concerns

- Are networks complex (not linear) systems? (Where does this leave inference?)
- To what extent do effects of networks depend on whole structure rather than separable characterisations of nodes?
- What do we say about dynamic networks?
- **How do we tell how much of a problem this is?**

Simplistic example

CASE A □□





CASE B □□

Same densities for all ties but one has a loop (alternative routes) and is disjoint. Problem gets more ambiguous when there is missing data.

What does simulation contribute?

- Point is *not* simply that density is an insufficient measure. We can raise the same issue about any measure or set of measures.
- Explicit formulation of generative process
- Explicit formulation of distributive process
- Ability to "sample" the simulated system in more than one way at very low cost.

The problem

- Under what circumstances can existing network measures can tell us useful things about the generative and distributive mechanisms at work in networks?
- What do we do when the existing measures fail in particular applications or classes of cases?

Case study – A Peer-to-Peer (P2P) File-sharing system

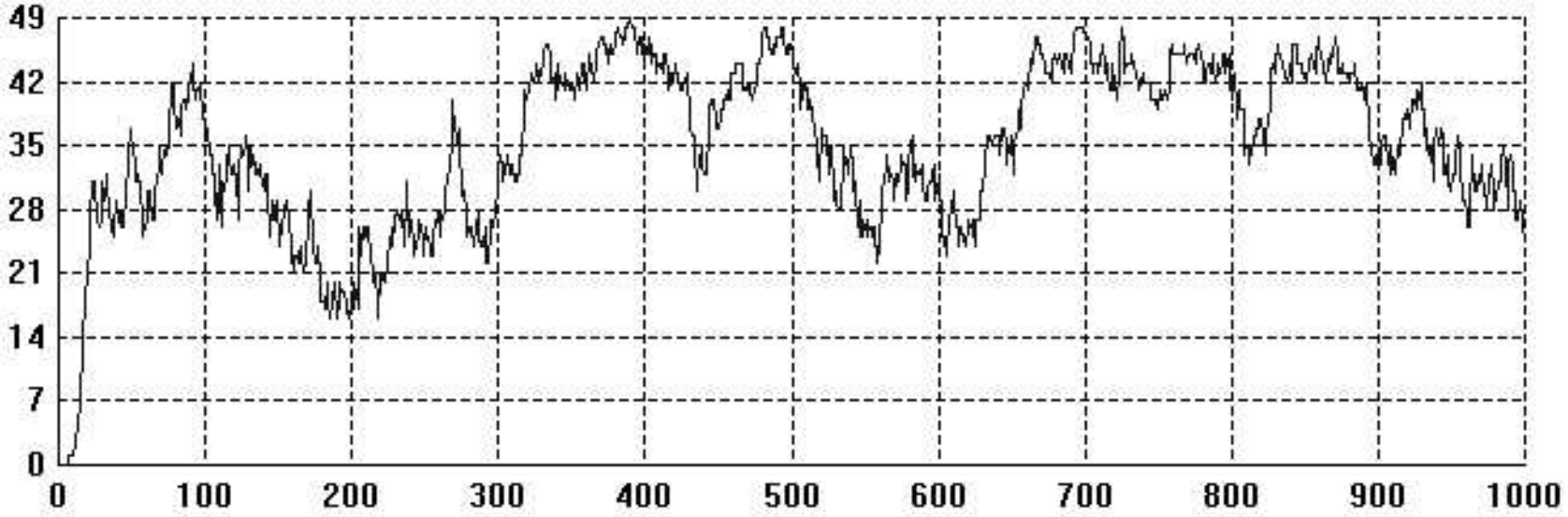
Collection of servers, each of which:

- Is controlled by a user to some extent
- 'Knows' a limited number of servers, with which it can communicate (the network)
- Makes some (or no) files available for download by other servers
- Search for files is by flood-fill: (i.e. send query to n others who send it to n others…)
- If query matches an available file it is sent back to originator

A Simulation of a P2P System

- 50 servers, each can decide to share files (*coop*) or not (*def*) at any time
- Try collect 'sets' of related files stored (initially) randomly by sending queries
- Satisfaction is measured by success at collecting files – (small) cost of dealing with others' queries (but decays over time)
- May look at and copy what a more satisfied server does, or may drop out and be replaced (especially if satisfaction is low)

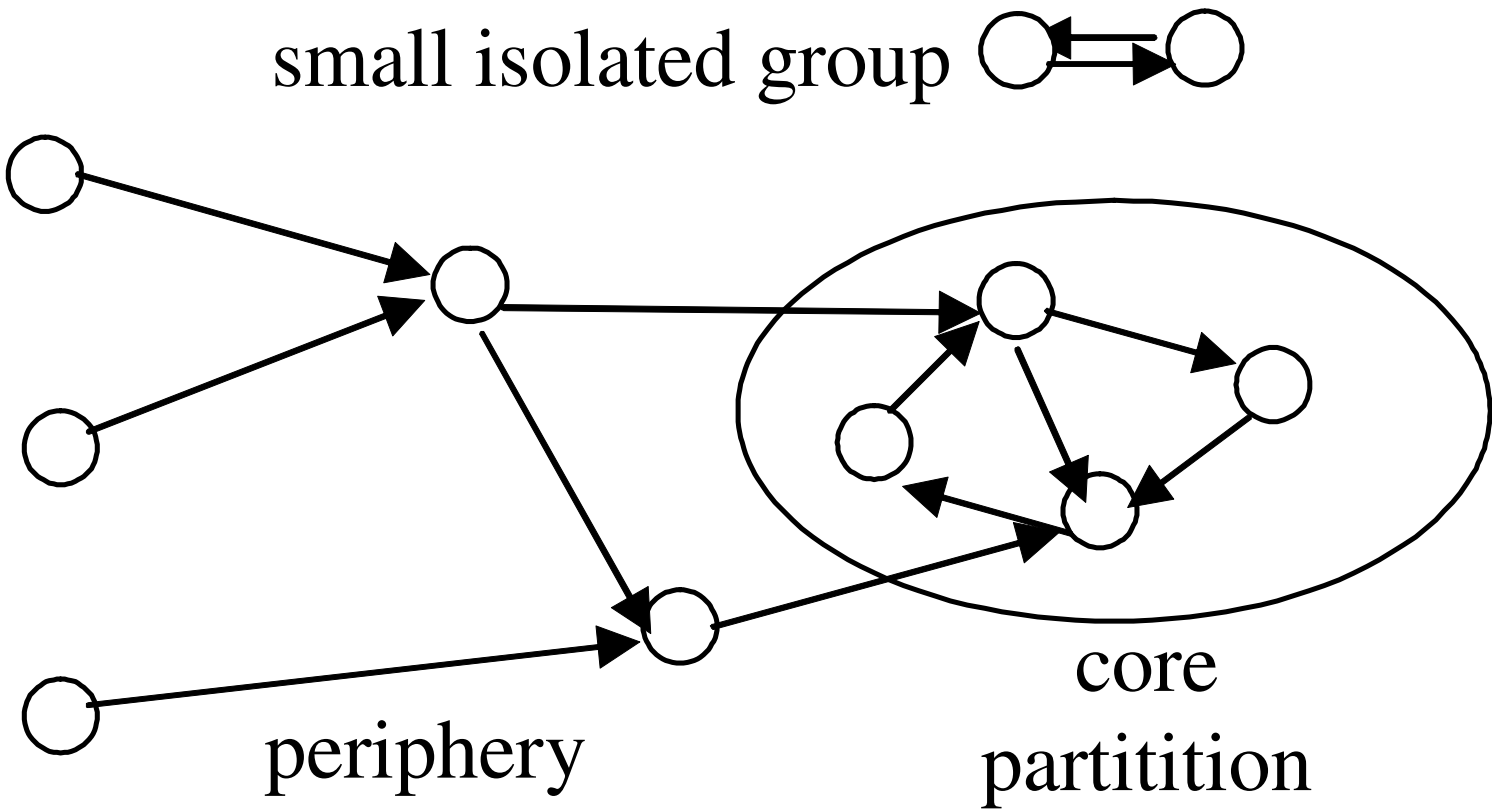
Number of co-operators in a run of the simulation



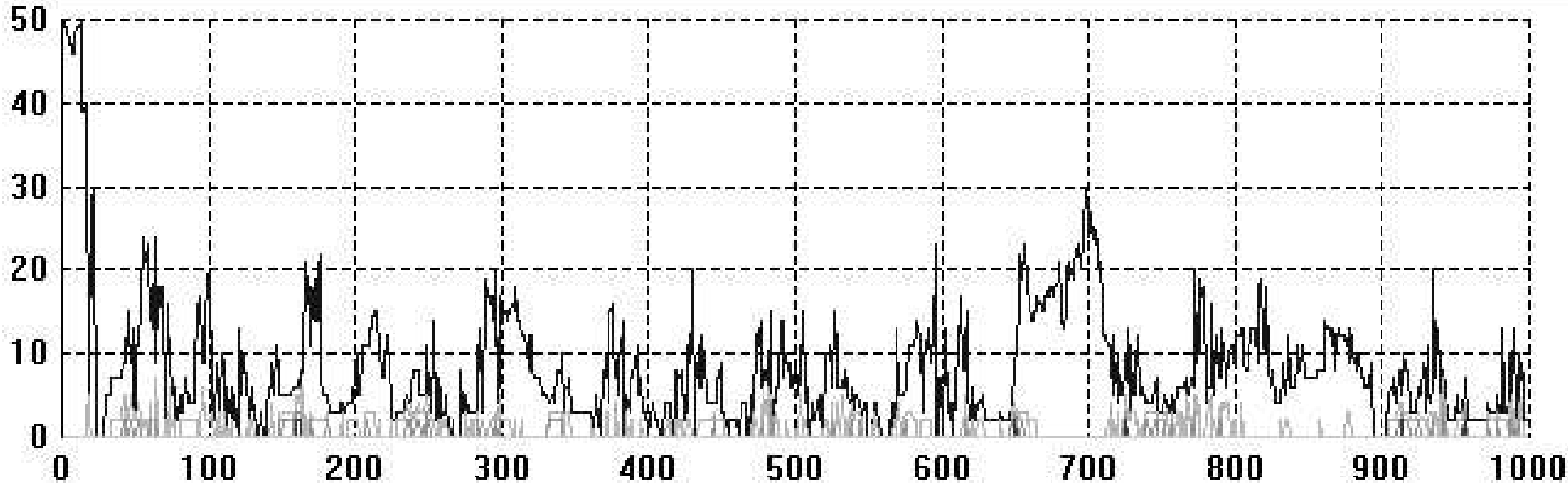
Key issue is number (and *manner*) of cooperation

- Why does anyone cooperate?
- How does network structure impact upon this?

Typical Emergent Network Structure



Size of partitions during a run



Blue – size of largest partition

Green – 2nd largest (if there is one)

Red, orange, etc. – even smaller ones

Suggests four types of node

- *In-coop* – those who share their files in core partition
- *In-def* – those who don't share their files in core partition
- *Out-coop* – those who share their files but are outside the core partition
- *Out-def* – those who don't share their files but are outside the core partition

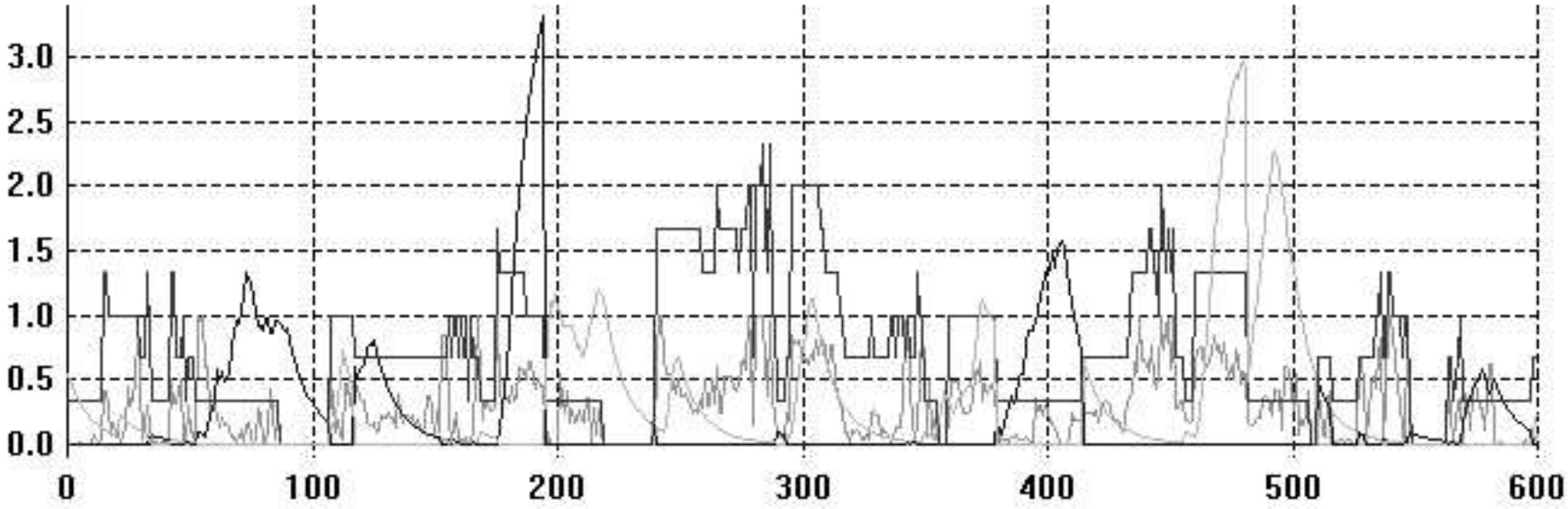
Some Statistics

Type	Average utility	Average number of links	Average centrality
in-coop	0.79	3.0	0.41
out-coop	0.51	2.5	0.31
in-def	0.37	2.0	0.27
out-def	0.32	1.5	0.19

Regression coefficients with satisfaction levels of nodes

Type	Number of links	Number of links lagged 6 periods	Centrality	Centrality lagged 6 periods
in-coop	-0.058	0.13	-0.062	0.12
out-coop	0.073	0.17	0.065	0.16
in-def	0.039	0.074	0.067	0.087
out-def	-0.15	-0.053	0.066	0.13

A history of a single node



Blue/green – is level of satisfaction (blue when *coop* green when *def*)
Red – number of outgoing arcs / 5
Orange – measure of centrality (0 – least to 1–most central)

Conclusion of Case-study

- The global measures were not very useful in providing ‘leverage’ on what was happening
- Rather a structural analysis based on a detailed understanding of the dynamics created a more useful categorisation of node types.
- It can be unsafe to assume that such measures derived from empirical studies give a helpful picture of the role of networks

Caveats

- Implausibility critique: Simulations not very "social" so SNA (developed on "real" networks) not challenged
- Naivety critique: Real practitioners of SNA would not have used those measures.
- BUT an attempt to raise a general problem and thus will require dialogue with traditional SNA to avoid these critiques. Dialogue has to start somewhere.

A Way Forward – staging abstraction with simulation