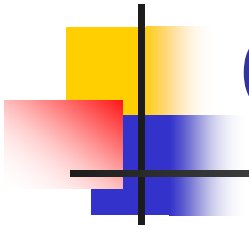
‘Discovering’ Small Worlds in Potentially Biased Networks

Sam MacAulay
John Steen
Tim Kastelle
ARC Centre for Complex Systems/University of Queensland
Business School

ARC Centre for Complex Systems
School of ITEE | The University of Queensland | ST LUCIA QLD 4069 | AUSTRALIA

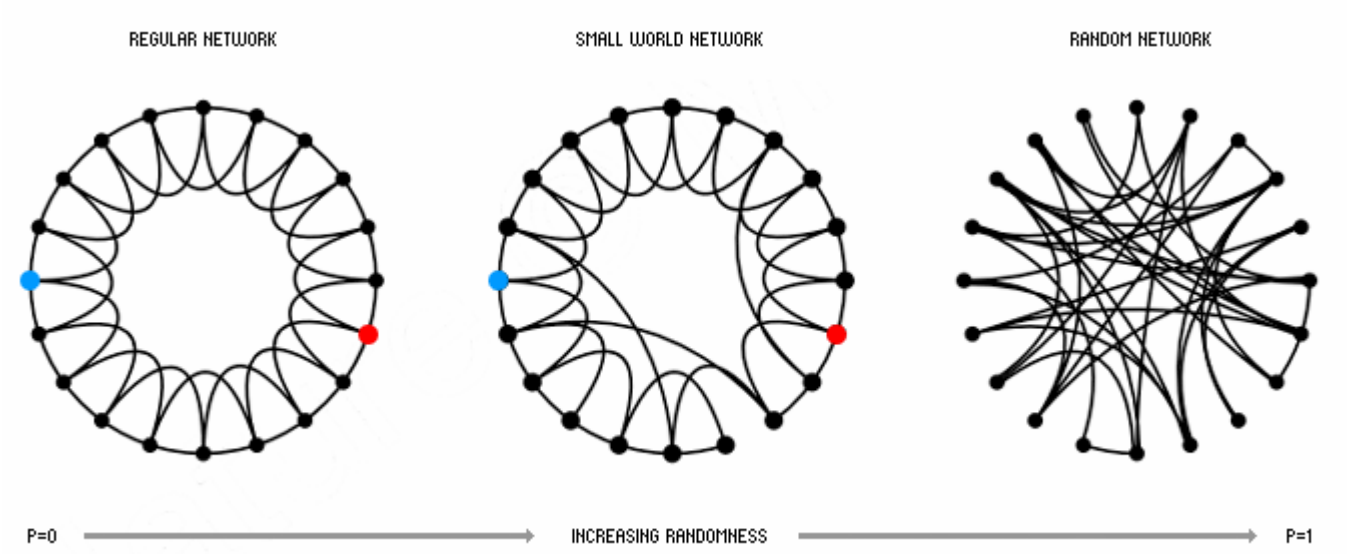
T: +61 7 3365 1003 | F: +61 7 3365 1533 | E: outreach@accs.edu.au



Overview

- Background and Aims
- Research Questions
- Propositions
- Meta-Review
- Results
- Insights from Physics
- Implications

What is a small world network? (Watts and Strogatz, 1998;Watts, 1999)



WS Small Worlds: High degree of local clustering between agents and simultaneous low global path link

Necessity is the mother of all creation

- Beginning a study at Vestas
- Reviewing the literature
  - Local vs. global ‘dichotomy’
- Physical vs. Social Systems: Questions began to emerge regarding the design of methodology and interpretation of empirical findings

Questions Emerging from the Differences Between Physical & social Systems

- How does one appropriately define the system and unit of analysis under investigation when its nature is inherently intangible?
- Why are researchers only examining strong forms of interdependence in the social sciences?
- The use of the term ‘non-linear’ is common within the literature. How does this change the way we design studies and interpret results in the face of missing data?
- **How are these dynamics going to influence the ‘discovery’ of small worlds?**

Maybe…

- **Proposition:** The relatively higher cost of identifying weak ties, when compared to strong ties, suggests that studies would primarily rely on strong tie data when describing system and thus systematically bias the small world statistic. E.G. Corporate Boards
- **Proposition:** The complexity and ambiguity of defining the unit of analysis and drawing system boundaries within socio-economic systems can increase the probability of misspecifying the empirical models used within the study of small world networks and result in bias. E.G. USNH
- **Proposition:** The inherent non-linearity of many systems where small world networks are examined would suggest, unless system boundaries were well known, there would be a high likelihood that missing data would bias the small world statistic. E.G. Singapore/Dubai in the World Trade Network

Meta Review

- Method (Business Source Premier; ABI/Inform; Jstor)
- 334 articles identified > 13 satisfied the criteria
- Results
  - Only two studies used primary data, both were pre-1977
  - Only four studies discussed how they decided on their unit of analysis (e.g. Robins and Alexander, 2004), and of these only one study explicitly discussed how and why it drew system boundaries in the way that it did (Gay & Dousset, 2005).
  - All studies considered elements of the systems under investigation inherently non-linear, but failed to discuss how this influenced study design and the interpretation of results.
- In addition: Recent work within the physical science provides preliminary support to the three propositions forwarded in this paper

The Sensitivity of the Small World Statistic: Insights from the Physical Sciences

- Recent work suggests that node deletion (Deng, Zhoa & Li, 2007), weight randomization of network ties (Li, Fan, Wang, Li, Wu & Di, 2007), changes in the proportion of weak-to-strong ties within a network (Shi, Adamic & Strauss, 2007) and the definition of system boundaries (Arita, 2004) can bias the small world properties of various networks.
- Each of these findings correlates to the three propositions put forward in section one of this paper.

Summary Table

Table 1. Support for Propositions found in the Physical Sciences		
Physical Science	*Corresponding Proposition*	*Relationship*
Li, Fan, Wang, Li, Wu and Di (2007) and Shi, Adamic and Strauss (2007): The introduction of weighted ties between actors can, in some circumstance, bias the small world properties of the network.	**Proposition One**: Strong tie bias	The biased introduced by the introduction of weighted connections within a network is a similar process as would occur by biasing the network towards the inclusion of strong ties. This evidence provides preliminary support to the notion that
Arita (2004): The misspecification of system boundaries meant that the metabolic structure of E.Coli was previously considered a small world. Arita (2004) provides evidence that in fact, when correctly defined, the system is not a small world.	**Proposition Two**: Drawing system boundaries and defining the unit of analysis	This paper demonstrates that even within a system that was previously thought to be well understood (Watts, 2006), system boundaries can still be misspecified. Furthermore, this misspecification can in fact mean that a system ceases to be considered as demonstrating small world properties. This finding demonstrates that if misspecification can occur within the physical science, that are argued to face less ambiguity when defining system boundaries, it is likely to be occurring within the social sciences, but is currently unacknowledged.
Deng, Zhoa and Li (2007): When networks are examined over time, as many of the empirical studies reviewed in Table 1 do, randomly deleting nodes actually biases the small world properties in the network under investigation. The degree and direction of the bias depends on the underlying properties of the network.	**Proposition Three**: Non-linearity and missing data	Whilst Deng, Zhoa and Li (2007) find that randomly deleting nodes biases the small world statistics, they also find that randomly replacing these nodes mitigates this bias. However, in the case of proposition three where there is missing data, which is roughly equivalent to deleting nodes, these nodes are not replaced and thus it can be suggested that some bias could eventuate.

Implications

1. Supplement secondary data with primary data (e.g. interviews in Lee Fleming's work)
2. Make explicit discussion of how the unit of analysis and the system boundaries were determined
3. Greater care must be taken with missing data when conceptual or empirical non-linearity are apparent and data is missing

THANKS!

- Please email me if you would like a copy of the draft working paper on which this presentation is based:

s.macaulay@business.uq.edu.au