18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Constructing Semantic Knowledge Networks from the ground up: livelihoods and employment outcomes in Anmatjere region, central Australia

Alexandridis, K. [1], **Y. Maru** [2], **J. Davies** [3], **P. Box** [3] **and H. Hueneke** [3]

[1] *CSIRO Sustainable Ecosystems, Davies Laboratory, Townsville, QLD, Australia*
Email: Kostas.Alexandridis@csiro.au
[2] *CSIRO Sustainable Ecosystems, Urrbae, Adelaide, SA, Australia*
[3] *CSIRO Sustainable Ecosystems, Alice Springs, NT, Australia*

Abstract: People need real opportunities to live the kind of life to which they aspire - to undertake livelihood activities they have reason to value, to achieve good health and well being outcomes, and to have resilience to shocks and stresses. A range of stakeholders consider that economic development is constrained by lack of engagement between Aboriginal people and labor markets, particularly given planned expansion of horticultural and mining operations. Aboriginal people of the Anmatjere region of Central Australia speak their own languages at home, have customary responsibilities for care of the region's natural and cultural resources, and have low levels of formal mainstream education. They aspire to jobs in their region and are engaged relatively strongly in employment in the community services sector and seasonal work in the pastoral industry, but not in other private sector employment. Their high dependence for income on social security payments and government funded jobs makes their livelihoods vulnerable to changes in government institutions. The modelling work presented in this paper is based on the views, attitudes and experiences of people living in the Anmatjere region about jobs and livelihoods. We have organized these as a collective knowledge representation, using semantic networks. This has elicited understanding of the structure, strength and quality of connections amongst social, economic, environmental and cultural dimensions important in people's livelihoods. The qualitative data were analysed using (a) natural language processing and linguistic algorithms; (b) exploration of semantic associations among knowledge constructs using a Hopfield-type Artificial Neural Network; and (c) graph-theoretic network analyses. We present the findings of this analysis in light of critical challenges that the Anmatjere community is facing. We show that culturally-explicit local Aboriginal institutions, world views and behaviours play significant and central roles in maintaining the community's knowledge representations. They connect people and establish the social and cultural roles that are critical in people's search for opportunity, income and the sustainability of their livelihoods in the region. 'Top down' actions including changes to government institutions aimed at enhancing individual Aboriginal people's engagement with employment have little chance of success unless they take into account the locally and culturally-specific ways in which the community is collectively functioning.

Keywords: *Indigenous communities; sustainable livelihoods; modelling; semantic networks; social networks; knowledge representation; artificial neural networks; participatory research*

1. INTRODUCTION

1.1. Semantic processing

Semantic processing has been a topic of scientific inquiry in social and community psychology methods since the latter half of the 1960s. Its goal is to represent the content of interactions amongst actors, e.g., the nature of their conversations, natural language communication or implied behaviour. The spreading-activation theory for semantic processing (Anderson, 1983; Collins and Loftus, 1975; Quillian, 1967) provided a first insight on the retention and activation processes of mental representation in conceptual semantic spaces. The theory allows the imputation of the weights of a series of semantic input nodes (activation nodes) and their propagation through a network of two-way mutual relationships among the network nodes via a 'firing' function (aka, 'spreading').

Critical questions have been raised about the ability of traditional social and cognitive models to address issues related to generative mechanisms of behavioural change both at individual and collective levels (Glaser, 2002) and phenomena such as social emergence, paradoxes and cognitive dissonance effects (Antrop, 2006; Gawronski and Strack, 2004). The emerging field of computational social science (Epstein, 2006; Sawyer, 2005) has enabled researchers to simulate and test assumptions and inferences of social science theory with a high degree of rigour, introducing stochasticity by running repeated simulations, enhancing test-retest reliability of estimated results and testing assumptions. Algorithms for semantic processing became functional with advancements in the fields of computational linguistics, neurocomputing, and artificial intelligence (e.g., see Christiansen and Chater, 2001; Jurafsky and Martin, 2000; Sun, 2008).

Computational methods for semantic processing have particularly important application where empirical information on the social settings under study is incomplete or characterized by deep uncertainty (Santos Jr et al., 2003). The social setting of the Anmatjere region of central Australia is one example.

1.2. Livelihoods in Anmatjere region

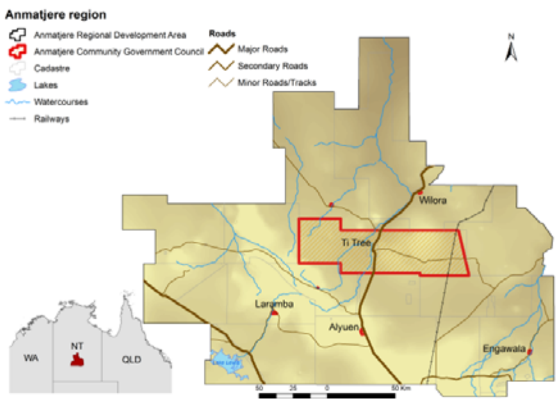

Figure 1: The Anmatjere region of Central Australia

The Anmatjere region, located in arid central Australia, is culturally diverse. Aboriginal people form the majority (86%) of the population of 1,350. Most (54%) speak Anmatyerr at home and more than three other Aboriginal languages, as well as English, are spoken by the region's people in their homes. Administratively the region now comprises the Anmatjere Ward of Central Desert Shire. Up to mid 2008, Anmatjere Community Government Council (ACGC) governed community services and employment development in the small town of Ti-Tree and eight Aboriginal settlements located on Aboriginal owned land around Ti Tree or on small areas of Aboriginal owned land excised from pastoral leases. Other localities (pastoral stations, roadhouses, horticultural areas and some Aboriginal settlements) were not encompassed by any local government structure but were included with the ACGC area by the Northern Territory government in regional development planning.

A key issue for regional sustainability and sustainable livelihoods has been low level of Aboriginal labour participation in established industries, notably horticulture, and in emerging industries and opportunities, notably mining and tourism. Our research has aimed to improve understanding about the relationships between jobs (paid activities) and livelihoods of people in the region, where 'livelihood' "expresses the idea of individuals and groups striving to make a living, attempting to meet their various consumption and economic necessities, coping with uncertainties, responding to new opportunities and choosing between different value positions" (Long 1997, quoted in De Haan and Zoomers, 2003, p. 351; Davies 2009, in preparation).

2. METHODOLOGY

2.1. Primary data

We collected primary data in Anmatjere region between September and December 2007. We interviewed 72 people (50 Aboriginal, 22 non-Aboriginal). The interviewees lived in Ti Tree Township, in three Aboriginal settlements and in some other dispersed localities. Their demographic characteristics were reasonably well-matched to the overall demography of the region. A structured interview format was used though most of the questions were open ended. We asked interviewees about the activities they undertake (whether or not these comprise paid employment), the assets they draw on, their income sources, aspirations and level of satisfaction with aspects of their livelihoods. Interviewers wrote out their responses and typed these up, generally on the same day. A men's and a women's focus group discussion was also held involving nine people in total. This tested some of the preliminary analyses of the qualitative data. Discussions were digitally recorded and transcripts were generated and used for analysis.

Most of the primary data collection was conducted in English. Several community members were engaged casually to assist with cross cultural communication and interpretation. Our use of English language, a structured interview instrument and handwritten notes by interviewers rather than voice recording of interviewees had the effect of filtering much of the primary data through the ontological frameworks of the research team. We are not able to account for the impact this has had on the textual material used for semantic analysis. We recognize these limitations (see discussion section) which arose from constraints of project resources and the developmental nature of the modelling undertaken in the project.

2.2. Extracting semantic categories from textual interview responses

The primary textual data contained 4,836 distinct sentences. The data were trained using a two-stage computational algorithm consistent with the spreading-activation theoretical and methodological framework for semantic processing. Specialised software was used for this purpose (MicroSystems Co. Ltd. and Megaputer Intelligence Inc., 2003; SPSS Inc., 2008). Training involved identifying semantically relevant concepts – that is, words or word combinations found in the textual data that are not meaning-irrelevant words. The excluded words were derived from lexicographic resources lists (for more information and examples, see Fellbaum, 1998; White et al., 2004).

The first stage of the algorithm computed the joined co-occurrence frequencies of semantically relevant concepts using iterative imputations of the primary data. Join frequency and Euclidean distance across the text was calculated for each pair of semantically relevant concepts. These were weighted by the number of lines, sentences and paragraphs within which the join occurrence is found. The pairs were then ranked according to their join weighted frequency score. Each score was then divided by the maximum number of joined weighted occurrences to compute frequency probabilities for co-occurrence of each pair of semantically relevant concepts. Concepts with a co-occurrence value ≥ 2 were used for the second stage of analysis. Additional algorithmic details are provided in Alexandridis and Maru (2009, in review).

The second stage computed the calibrated semantic weights of semantically relevant concepts. An optimization algorithm was used to derive the best possible groupings of concepts into relatively distinct classes that most likely encapsulate broader semantic information. We used the *Hopfield* Artificial Neural Network (ANN) since it has been shown to produce reasonable estimations within the spreading-activation theory of semantic processing (Goetz and Walters, 1997). The algorithm attempts to optimize the activation weight for a semantic classifier by minimizing an informational-energy objective function (Alexandridis et al., in preparation; Wang et al., 2006)

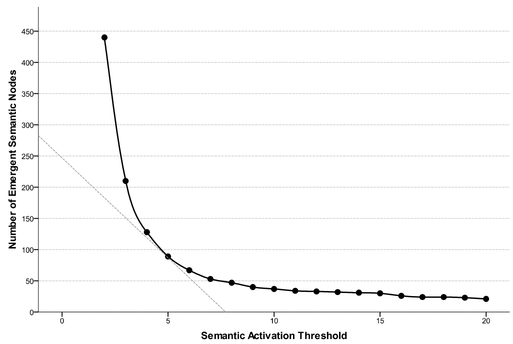

Figure 2: Graph of the number of emerging semantic categories (nodes) as a function of the semantic activation threshold used for ANN learning classification.

The choice of a semantic activation threshold depends on the depth and detail of the semantic analysis desired given that semantic networks are primarily hierarchical and their hierarchical tree span deepens as the level of detail (i.e., the number of semantic nodes) increases (Ravasz and Barabasi, 2003; Rogers, 2008). We used two rules of thumb. First, we desired a manageable number of semantic concepts (less than 100) to promote confidence in the inference of our further analysis. Secondly, we noted that the number of optimized semantic categories follow a negative exponential trend for various possible semantic activation threshold values (Figure 2). We located the point in the curve where the rate of the slope starts to stabilize, being where an orthogonal tangent between number of concepts and possible values for the activation threshold meets the curve. Reduction to the value of the threshold below this point has relatively small effect on the number of extracted semantic classifications. As can be seen in Figure 2, the best possible activation threshold meeting those rules of thumbs is 5.

2.3. Testing for randomness in network structure

To test for randomness in the structure of the Anmatjere semantic network we performed a Monte-Carlo simulation generating an ensemble of 100 random Erdós-Rényi networks (Bollobás, 2001) with the same mean network density as the empirical Anmatjere semantic network. We computed the average number of incoming ties (network in-degree) for each simulation run and the mean in-degree across all 100 simulation runs.

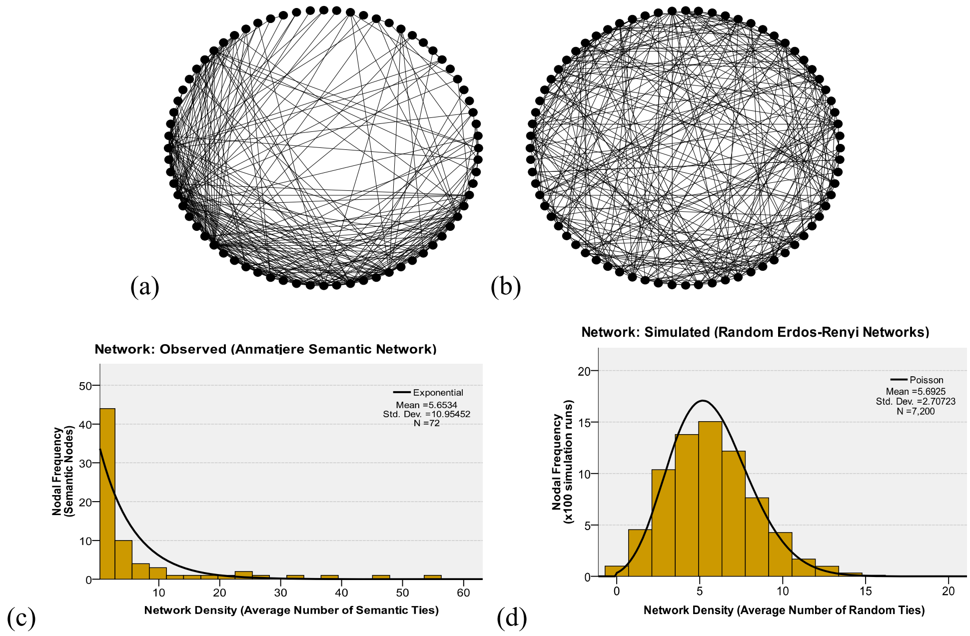

Figure 3: (a) The observed Anmatjere semantic network compared with (b) a theoretical random Erdós-Rényi graph. (c) and (d) show the network degree distribution (average number of ties per node) for network (a) and (b) respectively.

Figure 3 uses an optimized cyclical graph layout to depict (a) the observed Anmatjere network structure and (b) the structure of one of the simulated random networks. A limited number of highly connected nodes dominate the ties of the Anmatjere semantic network, whereas the density of network ties is comparatively uniform in the random network example. Figure 3 (c) plots the network density, or number of ties per node, for the Anmatjere semantic network compared to Figure 3(d) which plots the mean network density for the 100 simulation runs. In the Anmatjere semantic network the relationship of network density to nodal frequency displays a power-law (negative exponential) distribution, completely different to the Poisson distribution that is expected from a random Erdós-Rényi network (Barabási, 2003; Watts, 2003) and displayed in Figure 3(d). The results support other findings that connectivity in networks representing social interactions and processes has specific, non-random social characteristics (see, for example, Barabási and Albert, 1999; Sawyer, 2005).

3. RESULTS: COLLECTIVE KNOWLEDGE REPRESENTATION

3.1. Semantic degree and centrality

The Anmatjere semantic network contains 72 nodes (semantically important concepts) and a total of 676 reciprocal ties. The directionality of the network is inferred by the maximum value of the weighted semantic strength among node pairs. The in- and out-degree metrics of a semantic network denote respectively the degree of semantic influence that a particular concept (node) receives from other concepts and the degree of influence that a particular concept emits on other concepts. In the Anmatjere semantic network, the mean number of ties per node (degree), normalized to scale from the least to most important in- or out-degree, is approximately 5.7, as shown in Table 1. In-degree centralization is relatively high (~50%) and out-degree centralization is relatively low (~7%). This shows that the organization of the semantic concepts in the Anmatjere data involves a very few central semantic concepts that act as couriers of semantic influence and as semantic attractors.

Table 1: Summary statistics for Freeman In-degree and Out-degree distribution in the Anmatjere Semantic Network

Statistic	Freeman's Normalized Outdegree Centrality	Freeman's Normalized Indegree Centrality
Net Centralization (%)	7.122	49.97
Mean Degree	5.653	5.653
SE of Mean	.388	1.291
Standard Deviation	3.288	10.955
Minimum	.000	.000
Maximum	12.676	54.930
Percentile 05	1.408	.000
Percentile 95	11.268	32.394
Sum of semantic ties	676	676
Total N	72	72

The five most central recipients of influence in the Anmatjere semantic network (Figure 4) are: (a) the sense, function and purpose of the Anmatjere *community* as a collective entity; (b) *ways of doing things*, which are associated with local Aboriginal cultures; (c) *activities* that are undertaken by the region's people as part of their livelihoods; (d) the presence and enabling role that the local *council* (then ACGC) plays in negotiating emergence of livelihoods at the collective social level; (e) the central role of *family* as key actor in livelihoods of the region's people.

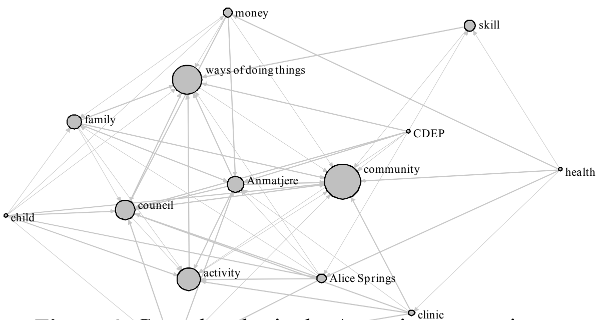

Figure 4: Central nodes in the Anmatjere semantic network, being nodes that have Freeman normalized in-degree values above the network average. Size of node is scaled by in-degree value.

3.2. Semantic Cohesion

The stronger the degree of cohesion among semantic concepts or group of concepts in the Anmatjere semantic network, the more those concepts can be said to form the core of the social construct of livelihoods within that community. We have measured distance-based indicators of cohesion in the Anmatjere semantic network being geodesic distance, compactness and breadth (Table 2). The reported metrics show that 91-95% of the semantic distance is occupied by core nodes whilst 5-9% of the semantic distances is occupied by the peripheral nodes. The analysis (De Nooy et al., 2005; Wasserman and Faust, 1994) indicates that the Anmatjere semantic network has a high degree of cohesion, implying in turn that the collective understanding of the region's residents about livelihood and employment issues is highly coherent. We explored this further by examining the presence and structure of cutpoints in the semantic network. A cutpoint is a node whose removal disconnects the network (Boyd et al., 2006). Hence cutpoints are often considered as threshold nodes in a social network (Janssen et al., 2007; Wang et al., 2006). Figure 5 shows the computed cutpoints for the Anmatjere semantic network. It indicates tight semantic dependency between family, (Anmatjere) region, (livelihood) activities, and cattle (grazing) which is the dominant extensive land use in the region. It also shows that 'opportunities', perhaps the most important enabler of employment, is not only a cutpoint in the network, but is also disconnected from this cluster of other cutpoints. This indicates lack of redundancy in the semantic network, a factor that network research has shown to be related to cascading failures, collapses and loss of resilience in many social and natural

Table 2: Results of the distance-based computed indexes for the Anmatjere semantic network

Measure	Shortest Path	Strongest Path	Lowest Cost Path	Most Probable Path
D_g	1.853	1.091	1.057	1.091
D_c	.097	.955	.972	.955
D_b	.903	.045	.028	.045
P(core)	.954	.909	.943	.909
P(periphery)	.046	.091	.057	.091

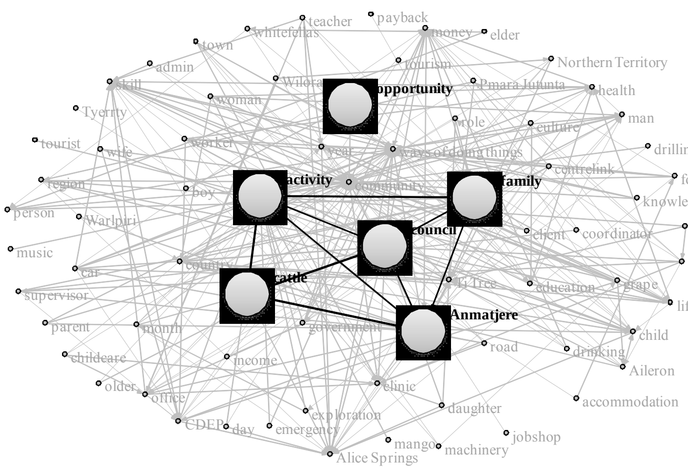

Figure 5: Graph-theoretic cutpoints of the Anmatjere semantic network, shown as darker large nodes.

systems (Neill, 2005; Ormerod and Colbaugh, 2006).

3.3. Exploring critical semantic paths

The Anmatjere semantic network was subjected to the computation of a critical path algorithm (Batagelj and Mvar, 1996). The algorithm searches all possible directed paths of any given length in the network, and tries to finds the one path that minimizes connectivity in the network (De Nooy et al., 2005). It is sensitive to the value of the ties among pairs of nodes. That connectivity path is the one that is most critical in the network. Without it, the network becomes disconnected. Figure 6 shows the results. Figure 6 confirms and reinforces the importance of central network structures in the Anmatjere semantic network. It shows that the flow and connectivity between income, government, council, ways of doing things, and community is unique within the ontology of the region's population and is very critical for maintaining the current social structure of the region. Actions or policies aimed in promoting employment and livelihood sustainability in the region need to take account of this critical path.

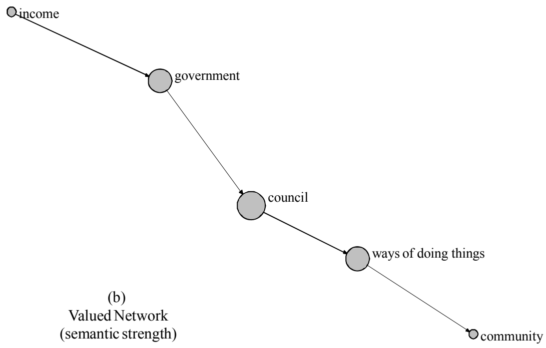

Figure 6: Network visualization of the critical path algorithm (CPM) implementation for the Anmatjere semantic network. The size of the nodes denotes fragmentation potential.

4. DISCUSSION AND CONCLUSIONS

The method employed and the analysis undertaken in this paper allows us to investigate important dimensions of the collective social aspects of livelihoods in the Anmatjere community related to employment outcomes. Some key points from the results and the methodology are:

Very few connected concepts and paths have consistently emerged. These are indicated as being central and critical to how people in Anmatjere region perceive and process basic issues of employment and livelihoods. These emergent connected concepts, paths and their properties form an evidence base for theorizing about employment and livelihood issues in the region.

Evidence based theories, such as grounded theory, are often developed by qualitative aggregation of similar statements to form a hierarchy of concepts. One of the strengths of semantic network analysis is that it allows us to study highly nonlinear and complex semantic knowledge relationships that present traditional social and statistical inference with many challenges. Another strength is that semantic network analysis allows us to focus on both concepts and the strength and structure of their connectivity.

Semantic network analysis provides a rigorous way to understand significant connected concepts that represent the collective knowledge held by research participants about the issues being investigated. The power-law distribution of semantic concepts, with very few highly shared and very many less shared (individual) concepts, demonstrates the robustness of the findings. However the method is sensitive to methods of primary data collection. In-depth interviews preferably using interviewees' first languages would further promote robustness.

The results of the analyses on critical nodes (cutpoints) and critical paths show that employment opportunities in the Anmatjere region are not connected with other important hubs on the network: there is a general lack of redundancies in pathways to valued livelihood options. The connections between income, government, council and community also form a path that is critically sensitive to any policy and program changes that influence one or more of these. This connectivity requires careful attention in policy design.

Finally, from a theoretical standpoint, the hierarchical structure of collective social aspects of livelihoods emerging from our analysis shows how social forces that are often interacting and self-catalysing affect collective realities, the interplay between livelihood choices and outcomes, and social change in general.

5. ACKNOWLEDGMENTS

The research reported here was supported by funding from the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC; the views expressed herein do not necessarily represent the views of Desert Knowledge CRC or its Participants. Willing engagement in the research by a wide range of people who live and/or work in the region is also gratefully acknowledged.

6. REFERENCES

Alexandridis, K., and Maru, Y. (2009, in review). Collapse and Reorganization Patterns of Social Knowledge Representation in Evolving Semantic Networks. *Information Sciences*.

Alexandridis, K., Maru, Y., Davies, J., Box, P., and Hueneke, H. (in preparation). Constructing Semantic Knowledge Networks from the Ground-up: Livelihoods and employment outcomes in Anmatjere region, Central Australia.

Anderson, J. R. (1983). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261-295.

Antrop, M. (2006). Sustainable landscapes: contradiction, fiction or utopia? *Landscape and Urban Planning, 75*(3-4), 187-197.

Barabási, A.-L. (2003). *Linked: How Everything is Connected to Everything Else and What it Means for Business, Science and Everyday Life*. New York: Plume Books, Penguin, pp. 304.

Barabási, A.-L., and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science, 286*(5439), 509-512.

Batagelj, V., and Mvar, A. (1996). Pajek: Program for Large Network Analysis (Version 1.19, 2007) [Computer Software]. Ljubljana, Slovenia: University of Ljubljana, Web.

Bollobás, B. (2001). *Random graphs* (2nd ed.). Cambridge ; New York: Cambridge University Press, pp. xviii, 498 p.

Boyd, J. P., Fitzgerald, W. J., and Beck, R. J. (2006). Computing core/periphery structures and permutation tests for social relations data. *Social Networks, 28*(2), 165-178.

Christiansen, M. H., and Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences, 5*(2), 82-88.

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Davies, J., Maru, Y., Hueneke, H., Grey-Gardner, R., Box, P., and Chewings, V. (2009, in preparation). *Outback Livelihoods in Anmatjere Region* (Research Report). Alice Springs, NT: Desert Knowledge CRC.

De Haan, L., and Zoomers, A. (2003). Development Geography at the Crossroads of Livelihood and Globalization. *Tijdschrift Voor Economische En Sociale Geografie, 94*(3), 350-362.

De Nooy, W., Mrvar, A., and Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press, pp. xxvii, 311.

Epstein, J. M. (2006). Agent-Based Computational Models and Generative Social Science. In Epstein, J. M. (Ed.), *Generative Social Science* (pp. 4-46): Princeton University Press.

Fellbaum, C. (1998). *WordNet : an electronic lexical database*. Cambridge, Mass: MIT Press, pp. xxii, 423 p.

Gawronski, B., and Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40*(4), 535-542.

Glaser, B. G. (2002). Conceptualization: On Theory and Theorizing Using Grounded Theory. *International Journal of Qualitative Methods, 1*(2), 31.

Goetz, P., and Walters, D. (1997). The Dynamics of Recurrent Behavior Networks. *Adaptive Behavior, 6*(2), 247-283.

Janssen, M. A., Anderies, J. M., and Ostrom, E. (2007). Robustness of Social-Ecological Systems to Spatial and Temporal Variability. *Society & Natural Resources, 20*(4), 307 - 322.

Jurafsky, D., and Martin, J. H. (2000). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, N.J.: Prentice Hall, pp. xxvi, 934.

MicroSystems Co. Ltd., and Megaputer Intelligence Inc. (2003). TextAnalyst (Version 2.3), Web: www.megaputer.com.

Neill, D. B. (2005). Cascade Effects in Heterogeneous Populations. *Rationality and Society, 17*(2), 191-241.

Ormerod, P., and Colbaugh, R. (2006). Cascades of Failure and Extinction in Evolving Complex Systems. *Journal of Artificial Societies and Social Simulation, 9*(4), art. 9.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12*(5), 410-430.

Ravasz, E., and Barabasi, A. (2003). Hierarchical organization in complex networks. *Physical Review E, 67*(026112), 7.

Rogers, T. T. (2008). Computational Models of Semantic Memory. In Sun, R. (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 226-266). Cambridge ; New York: Cambridge University Press.

Santos Jr, E., Santos, E. S., and Shimony, S. E. (2003). Implicitly preserving semantics during incremental knowledge base acquisition under uncertainty. *International Journal of Approximate Reasoning, 33*(1), 71-94.

Sawyer, R. K. (2005). *Social emergence : societies as complex systems*: Cambridge University Press, pp. ix, 276 p.

SPSS Inc. (2008). SPSS Text Analysis for Surveys (Version 3.0), Web: http://www.spss.com/textanalysis_surveys/.

Sun, R. (Ed.). (2008). *The Cambridge Handbook of Computational Psychology*. Cambridge ; New York: Cambridge University Press, pp. 753.

Wang, B., Tang, H., Guo, C., and Xiu, Z. (2006). Entropy optimization of scale-free networks' robustness to random failures. *Physica A: Statistical Mechanics and its Applications, 363*(2), 591-596.

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge; New York: Cambridge University Press, pp. xxxi, 825.

Watts, D. J. (2003). *Six Degrees : the science of a connected age* (1st ed.). New York: Norton, pp. 368.

White, H. D., Lin, X., Buzydlowski, J. W., and Chen, C. (2004). User-controlled mapping of significant literatures. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl 1), 5297-5302.