Collapse and reorganization patterns of social knowledge representation in evolving semantic networks

Kostas Alexandridis, Yiheyis Maru

Center for Marine and Environmental Studies (CMES), College of Science and Mathematics, University of the Virgin Islands, USA
CSIRO Ecosystem Sciences, Alice Springs, Australia

Abstract

This study introduces semantic network analysis of natural language processing in collective social settings. It utilizes the spreading-activation theory of human long-term memories from social psychology to extract information and graph-theoretic linguistic approximations supporting rational propositional inference and formalisms. Using an empirical case study we demonstrate the process of extracting linguistic concepts from data and training a Hopfield artificial neural network for semantic network classification. We further develop an agent-based computational model of network evolution in order to study the processes and patterns of collective semantic knowledge representation, introducing incidents of collapses in central network structures. Large ensembles of simulation replication experiments are conducted and the resulted networks are analyzed using a variety of estimation techniques. We show how collective social structure emerges from simple interactions among semantic categories. Our findings provide evidence of the significance of collapse and reorganization effects in the structure of collective social knowledge; the statistical importance of the within-factor interactions in network evolution, and; stochastic exploration of whole parameter spaces in large ensembles of simulation runs can reveal important self-organizing aspects of the system's behavior. The last session discusses the results and revisits the issues of generative semantic inference and the semantic networks as inferential formalisms in guiding self-organizing systemic complexity.

1. Introduction and rationale

1.1. Semantic networks and collective knowledge representation

Semantic networks present coherent and cohesive structures for knowledge representation (KR) and qualitative data mining tasks. Although often encountered as complex ontological hierarchies following strict rules of object inheritance [21,101], their most efficient and information-theoretic symbolism is their graph-theoretic representation as patterns of interconnected nodes (vertexes) and links (arcs). In such symbolic form, semantic networks encode a wide variety of semantic relationships, including linguistic approximations of knowledge and different types and scales of connectivity over complex hierarchical structures of subsumptional entities [20,62].

Semantic expressions can introduce knowledge approximations (also known as isomorphisms) to logic-based, rational propositional inference or formalisms [39,75]. According to Quillian [72], semantic networks based on linguistic assertions form an abstraction theory of the structure of the human long-term memory, able to embody knowledge as a computational
model of mind. The spreading-activation theory of semantic processing \([8, 9, 26]\) studies the effects of priming in semantic memory via associative semantic networks of interconnected relationships. These associations are stored and recalled in memory altogether (retention–activation process). Semantic networks of memory processing have recently contributed to a new approach to learning theory, namely the connectivism as a learning theory \([82]\). Connectivism focuses on the understanding of how informal learning patterns and associations (as complex adaptive systems) help people utilize and connect learning and knowledge rather than information itself. In their graph-theoretic properties, semantic networks are proven to follow scale-free distributions \([12, 73]\), and found to be strongly related to weak social ties in social networks of interactions \([43, 59]\).

In this paper, we will provide an evidence-based framework that goes beyond the construction of semantic networks from natural language knowledge representation \([38, 99]\), attempting to explore their evolution and their patterns of resilience and reorganization over a series of simulation-based experimentation techniques. A brief overview of the participatory and bottom-up empirical construction of semantic networks for a case study will be provided, and the use of Hopfield-type Artificial Neural Networks for training and extracting spreading-activation thresholds will be explored. Hopfield ANNs present a type of recurrent NNs with simple neurons and no hidden layers \([17, 52]\). They utilize dynamic computation of auto-associative semantic rules in connecting synaptic nodes in the network and symbolize content-addressable memory systems with binary threshold units. The iterative nature of semantic memory processing \([48, 92]\) and retrieval in Hopfield Semantic Nets allows for finding a lower “energy” level for the network that self-organizes spreading-activation network patterns over knowledge hierarchies. The paper will showcase how semantic processing in semantic memory priming and retention, modeled using Hopfield-type ANNs allows for efficient and correct knowledge approximation and ontological representation of dynamics in human cognitive processing \([see, for example, 4, 11]\). It will also demonstrate how human memory self-organization of semantic patterns improves informal learning and decision-making.

In terms of their social science context, the semantic representation discussed in this paper elicits collective social understanding of contextual relationships. In other words, the emergence of collective social structure as a function of how closely related are the mental models of individuals as members of the community or the social system under study. In this way, our approach to semantic network construction deviates from traditional cognitive approaches to knowledge representation. The strength of the semantic associations measured in our study depends on the degree to which such associations are widely shared by other individuals in the community. The more semantic concepts are shared by more members, the more they function at the collective social level of inference \([38, 81]\).

Sawyer \([78]\) argues that traditional theories of emergence and the structural patterns from widely accepted system theories of complexity such as the chaos theory and the away-from-equilibrium dissipative structures of Prigogine \([68, 69]\) are hard to fit into the sociological and sociocultural realities of social emergence. He argues that the internal interactions among the components of the social system itself (e.g., individuals, institutions, social collective structures and norms) are far more significant from their perspective “sociobiological inflows” \([78]\). He also raises the importance of computational social science techniques on the study of macro-sociological phenomena and social mechanisms as a form of “social explanation” \([77]\). A growing number of articles in the relevant literature illuminate the high added value of adopting and exploring computational methods in social, cognitive and psychological sciences \([23, 35, 76, 86]\). Computational social science methods, techniques and experimentation allow the study of complexity, adaptation and emergence in social systems ranging from the individual to the collective level of inference. Parallel advances in computational algorithms and artificial intelligence methodologies have provided an ever-expanding toolbox for computational social scientists in their attempts to study and explain highly complex patterns of social interactions (e.g., \([61, 71, 80, 94, 97, 100]\)). The volatile and subjective nature of human judgments, attitudes, dispositions and behaviors, coupled with the deep uncertainties and incomplete information characterizing the flows of real-world phenomena and their social and natural environments brings forth often implausible challenges for computational social scientists. It is not simply enough to describe “how” complexity in social systemic interactions emerges, but requires theoretical and evidence-based empirical arguments attempting to address “why” these complexity patterns emerge from social interactions, as well as for which scale prevalence of social inference (i.e., cognitive, individual, collective). Exploring the micro-to-macro links has been the “holy grail” of social science for the last few decades \([32, 35, 44]\).

Patterns of collapse and reorganization in social systems can emerge and persist in parallel in a society at different scales of social inference. For example, at the same time income inequality between the top 20% of the richest and the bottom 20% of the poorest in the US population has increased noticeably both in absolute numbers and in its global rankings, economic growth at the collective level and investment levels in monetary terms has risen considerably \([89, 93]\). Under the social science context, from the perspective of the increasing number of individuals falling under the poverty line in US their individual social experience could be characterized as a social collapse. On the other hand, from the perspective of the top global companies as collective organizations, their collective social experience is characterized as reorganization. Both of these social realities (individual and collective organizational) coexist in parallel social realities and form a part of a social holon, forming complex holarchies, the latter terms borrowed those terms from Lovelock’s and Margulis’s Gaia hypothesis \([60]\).

1.2. Related work and structure of the study

The empirical research findings from studying collective social processes, including collective knowledge representation flows, suggest that the transition from the individual to collective structure emerges alongside with the emergence of
power-law type degree distributions in social semantic networks. Our recent study of collective semantic associations in a remote desert Australian Outback community (Anmatjere region of central-north Australia) measured and analyzed a number of collective social aspects in relation to sustainability of employment-based livelihood outcomes and opportunities using semantic network analysis. Our results [5], show that collective social aspects of livelihoods such as social cohesion, centrality and social brokerage roles, along with their patterns of interactions across individuals and groups forming cohesive subgroups can be studied extensively using semantic network analysis. The Anmatjere community was a part of a set of remote Australian communities subjected to wider policy government interventions [87] aiming on reorganizing and enhancing collective social structures, roles and responsibilities. What became widely known in Australia as the Northern Territory Intervention, consisted of policy measures that removed central community and local governance institutions, often met with wider social and community distrust and a very high level of uncertainty about the future. In the midst of these changes our field work interviews and focus group meetings was conducted, but did not aimed to study the near- and longer-term effects of such changes in the collective social understanding of livelihoods in the region. The research effort we report in this paper utilizes the empirical findings of the Anmatjere semantic network analysis and introduces a series of additional questions that attempt to widen our theoretical and policy-related understanding of social change in relation to collective livelihoods.

We thus introduce a computational simulation framework that conducts social experiments over semantic networks structured in ways to resemble the social structure encountered in our empirical social settings. Within such an experimental computational social framework we conceive each collective semantic entity (i.e., network node) as an artificial agent, equipped with a number of key social metrics that signify its collective social importance, role and position in the overall semantic social structure. We also introduce a number of externally imposed changes in the network such as removal of key central nodes and collapses with varying degrees of severity and/or frequency of incidental occurrence. Finally, we provide the simulation agents (network nodes) with a response/feedback mechanism as the ability to restructure their association with other (similar) nodes following externally imposed events. Our aim is to study the following research questions:

(a) How semantic structure emerges from simple semantic interactions/associations (Section 2)?
(b) What are the effects in collective semantic social structure following collapses of key or very central social institutions or concepts (Section 3.1)?
(c) What is the magnitude of these effects when we vary the frequency and the severity of these interventions (Sections 3.1 and 3.2)?
(d) Given a set of simple feedback/response rules for the agents, how does social structure emerges and reorganizes if at all (Section 3.3)?
(e) What are conditions under which global (network-wide, that is, at the whole social system level) or local (cohesive group-specific) reorganization patterns emerge and propagate through the evolving networks (Sections 3 and 4)?
(f) What can we learn about guided-self organization in the context of knowledge representation in collective social systems (Section 4)?

The structure of the paper is as follows: Section 2 presents the basic construction of the semantic network inference, and the basic simulation framework for testing the emergence of self-organization. Specifically in Section 2.1 we present the basic inferential and algorithmic summary of the semantic network construction from natural language textual information, and its calibration for extracting the main network nodes (semantic ontologies) and the network links (semantic weights or strengths). Section 2.2 presents a computational social simulation ensemble using an agent-based model for testing the effect of different factors affecting the emergence of self-organizing patterns in similar semantic networks. Section 3 presents the basic analysis of the simulation ensemble results. The hierarchical decomposition of network trees, the emergence of power law distributions in semantic networks and the effects of different factorizations of network evolution (collapse intensity and frequency, and reorganization potential) are provided in Section 3.1. A longitudinal panel model analysis is presented in Section 3.2 testing and estimating the combined effects and combination of interactions among the network simulation ensemble factorizations, allowing a robust statistical framework that supports the proposition of emergent properties in complex systemic interactions in simulation ensembles. Section 3.3 goes further and test the effects of the parametric complexity in the emerging evolution of semantic network simulation ensembles using a GLM univariate estimator model. The final analysis Section 3.4 uses a multilayer preceptor ANN to group and estimate the accuracy of the factorization in the simulation ensembles network. The final Section 4 of the paper presents and summarizes the important findings of the analysis (Section 4.1) and provides a series of inference in the study of guided self-organization in semantic systems of collective knowledge representation (Section 4.2). The outline of the paper is shown in the following Fig. 1.

2. Simulating semantic network emergence

2.1. Extracting semantic networks from natural language texts in the Anmatjere case study

A three-stage process for extracting and analysing semantic knowledge representation flows has been employed. First, a number of interviews were conducted during field visits in the region with community members (n = 72). Two additional
focus group meetings followed the interviews (for men and women, given social conventions in the Aboriginal culture prominent in the region) that provided the opportunity for discussions and social deliberations among community members. The transcribed texts of the interviews and focus group meetings along with written summaries of discussion points have been used to extract an initial set of concepts present in the natural language texts using frequency-based linguistic data mining techniques (see for example [10,24,31]). The initial set of extracted concepts and their frequency weights was used as training inputs in the implementation of a Hopfield ANN algorithm for grouping those extractions to semantic categories. The Hopfield model produced a set of estimated semantic weights, by calibrating the spreading-activation threshold for inclusion in the network, and then optimizing the weights given that threshold. The use of the Hopfield-type ANN models is shown to produce good estimations in use with the spreading-activation theory of semantic processing [42,98]. The algorithm uses a recursive estimation for optimizing the objective function:

$$E = \frac{1}{2} \sum_{i \neq j} a_{ij} \cdot s_i \cdot s_j + \sum_{i} w_i \cdot s_i$$

where $s_{ij}$ are the initial frequencies of each node $n$ in the text; $w_i$ is the spreading-activation threshold, and $a_{ij}$ is the activation weight function between nodes $n_i, n_j$ and is calibrated such that:

$$a_{ij} = \begin{cases} 1 & \text{if } w_{ij} > w_t \\ 0 & \text{if } w_{ij} < w_t \\ w_{ij} & \text{if } w_{ij} = w_t \end{cases}$$

where $w_{ij}$: semantic weight between nodes $n_i, n_j$.

The objective function shown in Eq. (1) is also known as the Lyapunov equation whose asymptotic attractor properties are widely known and reported [36,42,96]. The auto-associative energy function estimation is capable of global optimization as it is shown to have properties resembling universal Turing computation [19,34]. The final output of the Hopfield ANN estimator is the calibrated semantic activation weights that optimize the objective function in Eq. (1). Further details of the machine learning categorization properties can be found in the relevant literature of semantic networks [6,27,88].

As can be seen from the properties of the Hopfield ANN estimator, the spreading-activation threshold value, $w_t$ is exogenous to the model estimation, that is, has to be determined a priori. In the Anmatjere semantic data, we imputed the algorithm for a vector of thresholds, ranging from 2 to 20.1 The results in terms of the estimated number of semantic categories (concept groups) and the estimated mean number of ties (links among categories), is shown in the top part of Fig. 1. Both trends show the prevalence of power-law distributions in the number of estimated semantic network sizes. We can think of the spreading-activation threshold as an indicator of “collectiveness” in social and community knowledge representation observed from our data. Increasing levels of threshold capture more widely held attitudes, beliefs or norms shared by wider groups of participants in the study. The presence of these power-law distributions shows how only a very small number of semantic categories and their knowledge-flows have wider influences on the broader collective social system.

Computing the ratio between extracted semantic categories (nodes) and semantic ties in our data showing in the bottom part of Fig. 1 allows us to observe the semantic emergence of a tipping point across spreading-activation thresholds. As we move from the individual to the collective level of semantic inference (increased threshold levels), the density of ties increases initially. This means that shared knowledge flows expand among widening cohesive subgroups of community members, until a tipping-point is reached (where $w_t = 8$) above which knowledge flows spread by few key collective semantic categories. Evidence from social network theory suggests that this might be the point where small-world network dynamics emerge [16,91], and the power of weak social ties exceeds collective social structure and institutions [18,46].

An example of the emerging Anmatjere semantic network for a threshold of $w_t = 5$ is shown in the following Fig. 2. Node sizes are varied based on the computed Freeman’s degree centrality coefficient [90], and show how nodes with high degree

---

1 The value of $w_t = 1$ is trivial, as it produces all possible extracted concepts from data, with no classification groupings.
centrality represent higher levels of collective social concepts such as the Anmatjere community, the (Aboriginal cultural) ways of doing things, the concept of livelihood activities, and the Anmatjere Shire Council.

Further analysis on the structure and social properties of the emerging Anmatjere semantic network can be found in the paper by Alexandridis et al. [5], and exceeds the scope of this paper. In the next sessions, we will attempt to go beyond the empirical snapshot of social structure in the Anmatjere region and perform additional analyses that will explore more dynamic dimensions of semantic emergence and evolution.

2.2. From semantic extraction to computational social simulations

The emergence of computational social science techniques, especially in the domain of social simulations has gained significant recognition in the last decade. Social simulations using artificial agents have been used extensively to aid our comprehension of both individual and social dynamics, as well as to unravel the complexity and patterns of interactions emerging at different social settings and situations [7,33,49,57]. When social simulation approaches are combined with attempts to achieve computational robustness especially in the face of policy-related research questions, increases the value of the scientific inference. Lempert, Popper and Bankes from Rand corporation [58] introduce the concept of robustness in simulation ensembles as powerful tools to perform sweeps of parameter spaces for desirable conditions or combination of conditions. They provide four basic principles or rules of thumb for achieving a robust framework of computational simulation ensembles: (a) exploring large parameter spaces by considering ensembles of very large simulation runs; (b) striving for robustness rather than optimality in computational strategies, i.e., conditions under which solutions persist rather than being optimal; (c) utilizing adaptive strategies seeking robustness, and (d) performing backwards-inference, e.g., asking what conditions fail to meet assumptions set forth by the simulations. In this analysis, we employed these rules to develop our simulation framework for the semantic emergence.

We introduced an ensemble of agent-based simulations using the NetLogo modeling environment [63]. Each artificial agent in the model represents a semantic node. At each simulation step a new agent is emerging and through an iterative algorithm links with another node. The linking algorithm was designed in a way to replicate the power-law dynamics observed in our empirical Anmatjere semantic network study. Barabási has demonstrated how preferential attachment gives rise to power-law, scale-free distributions [13,14], thus we used a preferential attachment algorithm for performing this task. Specifically, at each time step a new agent links with a node sampled from a probability distribution proportional to the

![Fig. 2. Top left and right: Distribution of the observed number of nodes and ties across spreading-activation thresholds in the Anmatjere semantic network respectively; bottom: Distribution of the observed tie density across increasing spreading-activation threshold levels in the Anmatjere semantic network data.](image-url)
networks’ tie density. The specific probability distribution is chosen as a model parameter of interest. Agents have the ability to measure and monitor key structural metrics of their status in their network at each step of the simulation, including their nodal centrality, the number of nodes to which they are connected, as well their actor-level degree and betweenness centrality [29,90]. At this reference simulation level, the model runs by evolving the network size. In this instantiation of the model, the network size is proportional to the time steps of the simulation.

In addition to the reference level, we introduced a set of parameters that alter the structure of the network. Specifically, the simulation framework allows for introducing the following additional variables:

(a) A collapse event, with frequency $c_t = t/T$. For a simulation with total steps $T$, at given time step intervals $t/T$, a collapse incident occurs, such that nodes with degree centrality above a certain level are removed from the network. This degree centrality level is determined with binary probability:

$$\Pr(\text{collapse} | c_t) = \begin{cases} 1 & \text{if } d_i > 1 + (1 - s) \cdot \max(d_n) \\ 0 & \text{otherwise} \end{cases}$$

where $d_i$ is the degree centrality of a node $i$; the term $\max(d_n)$ denotes the maximum nodal degree across all actors in the simulation, and $s$ is the severity of the collapse incident. When the severity is low, only the central actor is removed, but as the severity increases, the removal threshold becomes lower, and more actors are removed from the simulation.

(b) A severity of collapse level, $s \in [0, 1]$. The level determines the magnitude and effect of the collapse incident impacts in the network, as seen before.

(c) An actor reorganization level $r \in [0, 1]$. Following every collapse incident, the agents (actors) of the semantic network still remaining in the simulation are given the opportunity to re-establish lost links resulting from the collapse. Only those agents that have centrality above the mean network-level centrality are allowed to reconnect. During this stage, the agents pick randomly from one of their neighboring nodes and establish a connection. An implicit aspect of the simulation (as a result of the spring-embedding network layout) is that neighboring nodes have Euclidean distances inversely proportional to their centrality. Thus, this stage allows similarly central semantic concepts to connect.

These three external parameters (i.e., $c$, $s$, $r$) can be determined interactively from the user. When the model runs continuously at intervals $t' = t_0 + c_t$, collapses are introduced and subsequent nodal reorganizations occur. A snapshot of the user’s interface of the NetLogo model is shown in the following Fig. 3.
2.3. Generating simulation ensembles using repeated Monte Carlo replication runs

For the remaining analysis of this paper, we generated two major simulation ensembles of Monte Carlo experiments used for different purposes of the analysis.

- **Ensemble No. 1**: reference simulation runs (without collapse incidents). A 100 simulation steps × 100 replication runs. Thus $N_1 = 100 \times 100 = 10,000$ data points over 100 semantic networks.

- **Ensemble No. 2**: analysis simulation runs (with collapse incidents). We ran experiments of 100 step simulation over 100 replication runs each and tested all possible combinations for:

  \[
  c = \{10, 40, 70, 100\} \\
  s = \{0.1, 0.25, 0.5, 0.75, 1.0\} \\
  r = \{0.1, 0.25, 0.5, 0.75, 1.0\}
  \]

  thus $N_2 = 100 \times 100 \times 4 \times 5 = 1,000,000$ data points over 10,000 semantic networks.

For each of the data points in the two simulation ensembles, the simulation reported a total of 18 variables related to the state of the actors in the simulation. In addition, at the end of each replication run (after 100 time steps) we exported the full graph network for further analysis. The simulation ensemble results are analyzed in the next session. An important consideration that restricted larger ensembles is the capability of spreadsheets and statistical packages to analyze larger than the ones computed data samples. The second ensemble of simulation runs took approximately 8 h of simulation in an ordinary Intel Core dual-CPU laptop running at 2.00 GHz frequency with 2 GB of RAM.

3. Analysis of simulation results

3.1. Simulation ensemble trend analysis

3.1.1. Semantic network evolution without collapses

An example of the simulated semantic networks at different stages of network evolution is shown in Fig. 4. We used a graph root-embedding algorithm [41,53] to display the hierarchical decompositions of semantic relationships in the network. The nodes and links of the semantic network are arranged from semantic parents to semantic children categories. The effect of power-law distribution over network degrees is apparent in the four reported stages of the network evolution. As the network grows, from 25 nodes to 100 nodes, we can see that the mean degree of separating the top nodes from the majority of the rest of the nodes does not change. Even at $t = 100$, most of the nodes appear to be 2–3 degrees from the most

![Fig. 4. A screenshot of the NetLogo Simulation model used for the analysis of the semantic network patterns.](image-url)
central one. We can also see how weak social ties emerge in the network, as subgroups of nodes form tightly connected clusters that are only weekly connected with more collective social clusters.

Another interesting property of the emergent semantic social structure comes from the spreading-activation theory of long-term memory [26]. As spreading increases from the bottom to the top of the semantic topology, activation only functions in top-down flows (i.e., activating a node in the middle of the hierarchy also spreads to its children, but does not affect its parent nodes). In the social context of our simulation experiments, knowledge flows influence (spread) across sub-hierarchical social structures. Social structures such as the one emerging in Fig. 4 because of their scale-free properties are highly sensitive to structural changes at the top of the hierarchical decomposition.

In the first ensemble of simulation replication experiments we tested the effects of the semantic network evolution to key semantic social network parameters. Plotting the results of the evolution of the simulation span we can reveal the dynamics of temporal emergence in the semantic networks under study. Fig. 5 shows that a core set of semantic nodes are required for observing scale-free distributions. Most of the simulated networks exhibit those properties when maximum nodal degree reaches a threshold between 8 and 12 degrees of centrality. If we recall the results from the spreading-activation threshold calibration of the Anmatjere semantic network in Fig. 1 (θ = 8), we can see that our simulation findings provide a more robust confirmation of the level of semantic activation in which collective social knowledge flows spread throughout the community.

In order to provide more statistically robust evidence of the previous proposition, we performed a power-log regression of the mean observed network degree over the binned observed frequencies of network size from our simulation experiments. In each time step we computed the frequencies of nodes with various degrees, and averaged over the 10,000 simulated steps (for the exponential family of simulation runs). The regression results shown in Table 1, indicate a very good fit (adjusted $R^2 = .919$), and the $t$-statistic for the effects of the independent variable to explain the variance in degree distribution is statistically significant ($p < .001$).

Fig. 6 plots the predicted power-log regression along with the simulated mean network degrees of the dependent variable.

3.1.2. Introducing collapses in semantic network evolution

The introduction of collapses in semantic network structure has significant effects on the emergent structure of the evolving networks. Depending on different combinations of simulation parameters, those effects vary both in magnitude and

---

**Table 1**

Power regression fit estimation statistics.

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>Std. error</td>
<td>$\beta$</td>
<td></td>
</tr>
<tr>
<td>ln(Network degree (Binned))</td>
<td>–3.06</td>
<td>0.226</td>
<td>–0.961</td>
<td>–13.533</td>
</tr>
<tr>
<td>(Constant)</td>
<td>207.292</td>
<td>99.211</td>
<td>207.292</td>
<td>99.211</td>
</tr>
</tbody>
</table>

$R = .961, R^2 = .924, adj. R^2 = .919$. S.E. = .720.

---

Fig. 5. Hierarchical decomposition of the evolution of a simulated semantic network over 100 simulation steps, using a root-embedding algorithm. The strength of collective social structure increases from children to parent nodes.
spread across simulation time and semantic space. They range from network fragmentation to full systemic global network collapses. On the other hand, under certain conditions, global network-wide systemic reorganization emerges and propagates. Fig. 7 provides a series of visual examples of graph-theoretic layouts of network structures drawn from the results of the ensemble simulation runs. It would be impossible to graph the entire sweep of parameter space in this paper, so we just provide an illustrative example of potential outcomes. Due to the multi-dimensional character of the simulation runs, all four example networks displayed in Fig. 7 have been sampled from results with the same moderate frequency of collapse $c = 40$, that is a collapse incidence occurs every 40 steps of the simulation run. Consequently the networks shown here experienced two collapses over their total number of runs (at $t = 40$, and $t = 80$, respectively). The network growth is sampled at four equal intervals not directly following collapse incidents. We employed a graph-theoretic spring electrical embedding algorithm with inferential distance of 2.0 using the Combinatorica package for Mathematica [67].

In the top row of the graph network matrix, a reference simulation run is provided (no collapses) for comparison reasons. When node reorganization levels are moderate and the severity of collapses is low (second row) despite minor fragmentation of the semantic structure the network recovers and reorganizes into more cohesive social structure. Increasing the severity of collapses in the network has dramatic effects on both the structure of the network and the ability to recover, as we experience system-wide fragmentation of collective structures and only limited and weakly connected cohesive subgroups remain in the network. Further increasing the adaptive ability of the nodes to reorganize at moderate (and high) collapse severity levels does locally recover some centralized network structure, but those networks experience loss of resilience and exhibit structural sensitivity in subsequent collapse incidents.

The primary results of the simulations in the second ensemble of simulations are summarized in the following Figs. 8 and 9, in which a more precise picture is provided for different combinations of parameters in the model ensemble runs.

The interplay of patterns in the simulated mean nodal degree between severity of collapse and reorganization levels is shown in Fig. 8. Four different regions of interest can be identified in the results:

(a) Local reorganizations (low severity, low reorganization potential): the system is robust, but not resilient (relatively sensitive) and bounces back from events with relative minor severity, but the majority of networks remain to the same state.

(b) System-wide (global) reorganizations (low severity, high reorganization potential): the system undertakes state transitions to more efficient network configurations. The relative minor severity levels trigger system-wide reorganizations that exhibit high level of social resilience.

(c) System-wide (global) collapses (high severity, low reorganization potential): the system undertakes state transitions to less efficient network configurations. The yielded networks either physically collapse or are extremely fragmented. The effect of severe events pushes the network in states that cannot recover from. The social system completely lacks social resilience.

(d) Local collapses (high severity, high reorganization potential): the system shows a level of robustness that keeps it from full collapse. The majority of the networks maintain their overall state, but tend to be relative sensitive.

A similar set of interactions is observed for the simulated mean degree between severity and frequency of collapses in the semantic networks (Fig. 9). In this instance, collapse incidents characterized by high frequency of occurrence and low

![Fig. 6. Emergence of scale-free distributions over the growth of semantic networks.](image-url)
severity show higher mean degree centrality than the ones with moderate or low frequency overall. On the other hand, very frequent collapse incidents with high severity levels lead to systemic collapses of semantic social structure, and very unstable dynamics when the frequency is moderate or low.

3.2. Modeling longitudinal effects on panel network variables

One of the important research questions in this analysis is to understand the effects of the severity of collapse and the consequent reorganization patterns on the structure of the collective social structure in the semantic networks under analysis. To the extent that such semantic networks encapsulate the strength and pattern of collective and joint social knowledge representation within a community, the maximum degree of the vertexes in the network captures the power of the most joint held semantic concept. At the Anmatjere semantic network case study examples of central concepts with high degree centrality are the sense of community, the Aboriginal ways of doing things, the Aboriginal Shire Council, and the institution of family at the collective social level. Thus testing the effects in the overall network evolution when one or more of these central concepts are removed and the collective knowledge representation is called to reorganize allows to demonstrate how robust or resilient collective social structures are in such changes.

A useful approach for measuring the quantitative strength of those effects is to perform a panel regression of the maximum degree centrality as a dependent variable over the collapse and reorganization levels as independent variables. Given the structure of the second ensemble of the simulation experiments (100 simulation steps × five collapse levels × five reorganization levels × four collapse frequencies × 100 replication runs = 1,000,000 networks), a longitudinal panel analysis seems to be an appropriate method of analysis. We defined the simulation steps as our time variable and the replication runs as the longitudinal panel variable and we estimated a GEE population-averaged model regression using the frequency of collapse as the exposure (incidence) variable. GEE PA models have been shown to produce valid estimates of the effect of the independent variables to the dependent variable at the sampled population level under study [2,51]. The general equation for the GEE population-averaged model is

$$g(E(y_{it})) = \mathbf{x}_{it} \beta, \quad y \sim F(\theta)$$

where

- $g(\cdot)$ is the assumed link function, e.g., for the power (2) link function $g(y) = y^2$;
- $E(y_{it})$ is the mean value of the estimated parameter of interest across $i = 1, \ldots, m$ group (panel) identifiers and over $t = 1, \ldots, n_i$ simulation steps;
- $\mathbf{x}_{it}$ is the panel vector of $m$ observations by $n_i$ simulation steps for the dependent variable to be estimated, and;
- $F(\theta)$ is the assumed probability distribution family over which the variable of interest is distributed. For example, for the Poisson distribution $F(\theta) \sim \text{Poisson}$ with $\theta = \Pr(X = \exp(z))$.

The model estimates via maximum-likelihood the conditional probability:

$$\Pr(Y_{it} = y_{it}|\mathbf{x}_{it}) = F(y_{it}, \mathbf{x}_{it}\beta + v_i)$$

The assumed correlation structure for the error variance term $v_i$ can take various forms. Two within-group error variance correlation matrix ($R_{w}$) forms are examined for this model, namely the exchangeable equal-correlation structure shown in Eq. (6), and the AR (1) auto-regressive correlation structure, given in Eq. (7).
We tested 13 alternative model configurations and selected the model structure with the best overall fit on the population of networks under study. All models were statistically significant but with varying strength measured by their estimated Wald chi-square statistic. The results are shown in the following Table 2, sorted by their overall estimation performance. The identity link represents linear regression models, the log link represents logarithmic estimations, and the power link a power regression with exponent 2. We also tested models from three family of distributions, namely the Normal (Gaussian), the Poisson, and the Gamma distributions. We also tested different correlation structures for the variance error term of the panel regression, the simple (exchangeable) structure, and the auto-regressive terms (of orders 1 and 2).

The model fit comparisons indicate that the best fit for the model is an AR (1) autoregressive power (2) Poisson panel regression model. The best model’s estimated Wald chi-square statistic yields 18.8% better score than the second-best

\[
R_{ts} = \begin{cases} 
1 & \text{if } t = s \\
\rho & \text{otherwise} 
\end{cases}
\]

\[
R_{ts} = \begin{cases} 
1 & \text{if } t = s \\
\rho^{n-s} & \text{otherwise} 
\end{cases}
\]

Fig. 8. A snapshot of the network evolution in the semantic network simulation ensemble. Each row displays the evolution of a single network at 25 simulation step intervals for different parametric conditions of the severity of collapse and reorganization simulation properties. The first row of the graph matrix is a reference simulation (no collapses).
model. The full estimation results for the model with the best fit are shown in the following Table 3. Both model parameters, namely the reorganization level and the severity of collapse level show statistical significant effects on the maximum nodal degree of the network, since $P(r > \chi^2) < 0.001$. Furthermore, the strength and direction of the computed $Z$ statistic shows the negative effect of the severity level and the positive effect of the reorganization level to the achieved max degree of the network in our data.

Overall the analysis in this session demonstrated that (a) the nature of the relationship between the observed distribution of the maximum network degree as a function of the reorganization and severity of collapses can best be modeled using a power-law of exponent $2$. This result reaffirms theoretical arguments related to network degree evolution \cite{13,55}; (b) Both reorganization and collapse levels have statistically significant opposite effects to the observed network degree, and (c) Those statistically significant effects follow are auto-correlated, that is, dependent on the temporal evolution of the network.

![Graphs by Severity level and Reorganization level](image)

Lines represent fitted values over 100 x 100 replication runs

**Fig. 9.** Evolution of mean degree as a function of the severity of collapse and reorganization levels in the simulated semantic networks.

<table>
<thead>
<tr>
<th>Model</th>
<th>Link</th>
<th>Family</th>
<th>Correlation structure</th>
<th>Wald $\chi^2$</th>
<th>Relative change (%)</th>
<th>Prob $&gt; \chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Power (2)</td>
<td>Poisson</td>
<td>AR (1)</td>
<td>3131.54</td>
<td>18.8</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>Power (2)</td>
<td>Gaussian</td>
<td>Exchangeable</td>
<td>2637.02</td>
<td>3.3</td>
<td>0.0000</td>
</tr>
<tr>
<td>13</td>
<td>Power (2)</td>
<td>Poisson</td>
<td>AR (2)</td>
<td>2553.19</td>
<td>0.9</td>
<td>0.0000</td>
</tr>
<tr>
<td>10</td>
<td>Power (2)</td>
<td>Gaussian</td>
<td>AR (1)</td>
<td>2529.91</td>
<td>0.1</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>Power (2)</td>
<td>Poisson</td>
<td>Exchangeable</td>
<td>2528.03</td>
<td>26.2</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>Power (2)</td>
<td>Gamma</td>
<td>Exchangeable</td>
<td>2003.98</td>
<td>3.8</td>
<td>0.0000</td>
</tr>
<tr>
<td>12</td>
<td>Power (2)</td>
<td>Gamma</td>
<td>AR (1)</td>
<td>1931.13</td>
<td>24.5</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>Identity</td>
<td>Gaussian</td>
<td>Exchangeable</td>
<td>1551.04</td>
<td>6.8</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>Identity</td>
<td>Poisson</td>
<td>Exchangeable</td>
<td>1452.77</td>
<td>6.1</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>Log</td>
<td>Poisson</td>
<td>Exchangeable</td>
<td>1369.86</td>
<td>94.2</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>Log</td>
<td>Gaussian</td>
<td>Exchangeable</td>
<td>705.45</td>
<td>10.4</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>Identity</td>
<td>Gamma</td>
<td>Exchangeable</td>
<td>638.82</td>
<td>3833.6</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>Log</td>
<td>Gamma</td>
<td>Exchangeable</td>
<td>16.24</td>
<td>0.0</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
3.3. Predicting effects of parametric complexity to network evolution

In order to demonstrate the relative magnitude and complexity of interactions among the previous model parameters (independent variables) we conducted an additional GLM univariate analysis on the data ensemble. We tested the null hypothesis about the effect magnitude of the independent variables (reorganization level, severity level and frequency of collapse) on the means of alternative groupings of the observed maximum degree. We de-trended the effects of the temporal evolution of the network by specifying the categorical binning of simulation steps (620, 21–40, 41–60, 61–80, 81–100) as frequency weights for the univariate analysis of variance.

The Levene’s test of equality of error variances in the model, testing the null hypothesis that the observed error variance of the maximum network degree is equal across groups produced a statistically significant result ($F = 580.289, p < .0001$). The test statistic allows us to reject the null hypothesis. The analysis of the effects and significance of between-factor interactions in the univariate model is shown in the following Table 4. Whilst almost all of the 1, 2 and 3-way effects (with the exception of the simulation step) were found statistically significant, only the 2-way and 3-way factor interactions were producing estimates that minimized the mean squared error estimates. The estimated partial Eta-squared statistic ($\eta^2_p$) measures the strength of the association among factors, as:

$$\eta^2_p = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}}$$

The computer $\eta^2_p$ results from Table 4 show that the two-way effects of increased severity or reorganization levels as the network evolves (across increasing simulation steps) have very high strength of association (~.9 and above).

The effects of factor interactions in the estimated maximum network degree are shown in the following Fig. 10. As severity level increases the marginal mean of the max degree decreases significantly, but as reorganization levels increase those

---

### Table 3
Best model estimation results for the panel Poisson power auto-regressive model.

<table>
<thead>
<tr>
<th>Group and time vars:</th>
<th>Number of obs = 1,029,836</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link: Power</td>
<td>Number of groups = 10,000</td>
</tr>
<tr>
<td>Family: Poisson</td>
<td></td>
</tr>
<tr>
<td>Correlation: AR (1)</td>
<td></td>
</tr>
<tr>
<td>Scale parameter:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Max Degree</th>
<th>Coeff.</th>
<th>Std. error</th>
<th>Z</th>
<th>Pr &gt; z</th>
<th>95% Conf. interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reorganization level</td>
<td>2.3254</td>
<td>.100468</td>
<td>23.15</td>
<td>0.000</td>
<td>2.1285 2.5224</td>
</tr>
<tr>
<td>Severity level</td>
<td>5.1485</td>
<td>.103321</td>
<td>49.83</td>
<td>0.000</td>
<td>4.9460 5.3510</td>
</tr>
<tr>
<td>Frequency of collapse (exposure)</td>
<td>37.8565</td>
<td>.488887</td>
<td>77.43</td>
<td>0.000</td>
<td>36.8983 38.8147</td>
</tr>
</tbody>
</table>

---

### Table 4
Effects and significance of factor interactions in the GLM univariate model analysis. Tests of between-subjects effects.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III SSE</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>Sig.</th>
<th>Partial eta squared</th>
<th>Noncent. parameter</th>
<th>Observed power$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Hypothesis</td>
<td>17754809.131</td>
<td>1</td>
<td>17754809.131</td>
<td>57.963</td>
<td>.000</td>
<td>.891</td>
<td>57.963</td>
</tr>
<tr>
<td>Reorganization level</td>
<td>Hypothesis</td>
<td>127470.050</td>
<td>4</td>
<td>31867.513</td>
<td>18.128</td>
<td>.000</td>
<td>.793</td>
<td>72.512</td>
</tr>
<tr>
<td>Severity level</td>
<td>Hypothesis</td>
<td>866992.454</td>
<td>4</td>
<td>216748.113</td>
<td>19.497</td>
<td>.000</td>
<td>.826</td>
<td>77.989</td>
</tr>
<tr>
<td>Simulation step</td>
<td>Hypothesis</td>
<td>281694.165</td>
<td>4</td>
<td>70423.541</td>
<td>5.624</td>
<td>.003</td>
<td>.524</td>
<td>22.497</td>
</tr>
</tbody>
</table>

$^a$ Computed using alpha = .05.

$^b$ Weighted least squares regression – weighted by frequency of collapse (incidence).
effects are significantly contained, especially when reorganization levels exceed the .75 level. At those levels of reorganization, the network whilst losing their centrality strength (smaller maximum degrees), it does not experience regime shifts or generalized state transitions as the ones emerging at lower levels of reorganization when severity levels are high.

3.4. Using a Multilayer Perceptron (MLP) classification algorithm

In the final analysis stage of the simulated network ensemble we want to examine further the structure and pattern of interactions among the model run parameters (reorganization levels, severity levels, frequency of collapses, and temporal simulation steps). So far, we showed that the effects of these parameters are significant predictors of the semantic structure (in the longitudinal panel analysis), and that cross-factor interaction effects influence the emerging variance of observed patterns in semantic social structure. One further critical question set forth in the introduction is yet to be answered: Which combination of factors enable or disable the emergence of system-wide self-organizing patterns across semantic social structure? Attempting to provide an answer to this question is far from trivial task, as it involves understanding the specific patterns of interactions among the model parameters and how they affect the observed results. The previous analysis steps showed also that such interactions are nonlinear and nonmonotonic. Traditional statistical inference does not provide us with tools to evaluate those patterns of interactions [3,22]. Machine learning algorithms and data mining techniques have been developed and showed able to overcome the problems of increased stochasticity and multidimensionality in databases [50,54]. We employed a pattern learning and estimation of the factorial semantic interactions by the use of an artificial neural network Multilayer Perceptron (MLP) model. Multilayer perceptron models are pattern classifier algorithms capable to solve blind-source separation type problems in an efficient and probabilistic manner [28,45].

We partitioned the data on two datasets, for training and testing purposes of the MLP classification model. Of the total 992,996 valid (non-empty) data points for the dependent and independent variables, 70% of the data (694,752 cases) were used for training the ANN classifier, and 30% (298,244 cases) for testing the classifier’s predictions. The ANN classifier algorithm for the multilayer perceptron model converged on two different activation functions for the classification training task, using a best estimation choice rule [84]. Specifically, the hyperbolic tangent activation function was used for estimating the coefficients (synaptic weights) of the hidden layers in the neural network, and the Softmax activation function was used for estimating the synaptic weights of the output layer. Across the training and testing model classification tasks the best estimator for the error function was the cross-entropy method. The estimated network parameters are shown in the following Fig. 11. Positive larger values of the synaptic activation weights are associated with increased influence of the input layer subcategory or increased influence to the output layer subcategory corresponding to the incoming or outgoing synaptic connection of the hidden layers to inputs or outputs, respectively. Similarly, negative smaller values of the synaptic activation
weights are associated with the reversing of the influence strength to and from the corresponding input (independent variables) or output (dependent variable) layers, respectively.

The profiles of the estimated synaptic activation weights indicate that the strength of the classification varies significantly across the nodes of the hidden layer. As can be seen in Fig. 11, hidden nodes H (1:4,5,6,9) show relatively stronger activation strength than H (1:1,2,3,7,8). For example, from all the activation weight patterns, H (1:6) has very strong positive weights for input conditions characterized by a pattern \((t, c, s, r) = (21–40, <1.43, 0.1, 1.0)\). The same hidden node, exhibits very strong positive synaptic activation weights for output conditions where \(\max(d)_{P_{10}}\) (and relatively low negative weights for all other conditions). In this example, the rule that the classifier imposes is to enhance the probability of observing a relative high maximum degree in the semantic network when relative infrequent and minor collapses occur early on in the network evolution and the reorganization level is very high. A similar rule (increasing the probability of \(\max(d)_{P_{10}}\)) is reinforced by the H (1:4) but when collapses do not emerge in the first 20 simulation steps. If we are to examine the patterns of conditions for which higher maximum degrees are enabled in the evolution of semantic networks, we need to look at which combinations of patterns the synaptic activation weights for the output layer \(\max(d)_{P_{10}}\) are positive or negative. The last subgraph of the bottom panel in Fig. 11 provides the answer to this question. Classification patterns captured by hidden layers H (1:1, 2, 4, 5, 6) yield \(\Pr(\max(d)_{P_{10}} > 0)\), whilst classification patterns H (1:3, 7, 8, 9) yield \(\Pr(\max(d)_{P_{10}} > 10) < 0\).

A more comprehensive picture of the effect of reorganization patterns to the predicted patterns of centrality in the semantic networks following collapses is shown in the following Fig. 12. The y-axis in the graph shows the predicted pseudo-probabilities of the MLP model classifier for the \(\max(d)_{P_{10}}\) output category. The x-axis records the binned network evolution over the 100 simulation steps of the data ensemble. The lines on the panel graph show the variability in the predicted probability for achieving a nodal degree as high as 10 in the simulated network data ensemble, for different reorganization levels. The lines are paneled by severity levels (columns) and frequency of collapse incidents (rows). The structure of the MLP classifier ensures that the pseudo-probabilities across all categories of the output layer sum-up to 1.0, therefore, exceeding a probability of 0.5 for a single class (e.g., \(\Pr(\max(d)_{P_{10}} > 10) > 0.5\)) ensures the classifiers’ selection, and thus, it is the most probable outcome of the simulation. The results clearly show specific conditions under which different reorganization patterns in the network contribute to more central and cohesive semantic social networks.

More specifically, a number of interesting inferences can be drawn from these results:

(a) When the incidental frequency of collapse is very low (=1), that is, in conditions where collapses happen once during the simulation span, the network always achieves high centrality as can be seen from the first panel row in Fig. 12. Some exceptions occur when the severity of the collapse is high to very high (>0.8) and the reorganization level is very low (=0.1) where the probability asymptotically reaches 0.5 but does not exceeds that level.
(b) When the severity level of the collapse incidents is low, the semantic centrality of the emerging networks is not significantly affected by them, but the outcome is highly sensitive to the level of reorganization prevalent in the system. The first column of the panel graph in Fig. 12 shows how different reorganization levels diverge over time in achieving level of semantic centrality. A case of specific interest is the third subgraph of the first column, where the collapse frequency incident rate is 2.5. In terms of the specific simulation conditions this means that collapses emerging every $100/2.5 = 60$ simulation steps. The line patterns show how the effect of a collapse affects the network centrality temporarily, but when the reorganization level is relatively high, the network bounces back and achieves very high centrality levels, even higher than in any other parametric combination of conditions.

(c) In general, at severity levels >0.1 when the frequency of collapse is moderate and above (incidence rate > 1.43), the network centrality experiences long and short term collapses. Whilst some local level resilient patterns can be experienced in the second and third column of the panel graph, the network experiences system-wide collapses at very frequent incident rates (=10). At moderate incident rates, the prevalence of resilience is sensitive to the level of reorganization of the system.

The estimated and predicted results for the AROC of the MLP classifier (Table 5) indicate that for all the classification categories of the dependent variable the AROC is very high (>0.8). The high value of AROC signifies a departure from classification patterns expected by chance (diagonal reference line of the ROC curve). Furthermore, as can be seen also in the lift plot in
Table 5
Area under the curve (AROC) statistics for the MLP classification.

<table>
<thead>
<tr>
<th>Measured statistic</th>
<th>Variable</th>
<th>AROC</th>
<th>Importance</th>
<th>Normalized importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td>Max nodal degree</td>
<td>≤ 4</td>
<td>.914</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5–6</td>
<td>.809</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7–9</td>
<td>.805</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10+</td>
<td>.890</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>Reorganization level</td>
<td></td>
<td>.116</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Severity level</td>
<td></td>
<td>.246</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Frequency of collapse</td>
<td></td>
<td>.309</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Simulation step (binned)</td>
<td></td>
<td>.329</td>
<td>100</td>
</tr>
</tbody>
</table>
collective semantic knowledge depends on the availability of additional future empirical settings and experiments. Nevertheless we lack further empirical findings from real-world participatory settings anthropogenic systems especially in the face of broader social and environmental challenges for the future.

Social science needs to discover and explore further the real power that the bridging of semantic knowledge representation of knowledge flows and less cohesive social semantic structures that achieve limited collective social outcomes. The latter social systems exhibit bifurcations and easily shift (or oscillate) across possible outcomes. State transitions.

Policy interventions aimed on affecting central social structures should take into account the ability of social actors and institutions to adapt to changes. Social self-organization could be feasible but not without state transitions.

The analysis performed on the simulated ensembles of network data allowed us to answer a number of critical questions regarding patterns of self-organization in networks of semantic collective knowledge representation. We demonstrated the central role of scale-free distributions in allowing the emergence and propagation of knowledge patterns at the collective social level. We also provided additional robustness in the empirical results obtained via our case studies, as well as independent validation of key emerging results involving the existence of tipping-points and scalability of collective semantic social structures. Throughout the statistical analysis of this paper we are able to raise a number of important points.

Firstly, we showed that collapses without the ability to reorganize have significant fragmentation effects on semantic network structure. Social change involving removal of key collective social institutions without examining the adaptive ability of the remaining social structures can lead to wider fragmentation of social knowledge with highly unpredictable effects. Policy interventions aimed on affecting central social structures should take into account the ability of social actors and institutions to adapt to changes. Social self-organization could be feasible but not without state transitions.

Secondly, sequential collapses on such systems can be quite volatile and unpredictable. It often leads either to total loss of semantic representation, or to reconstruction of the knowledge and re-emergence of scale free distributions. The bimodality of those interactions is sensitive on the balance levels of parametric mixes. Limited evidence of the presence of nonlinear thresholds and tipping points of the interactions between collapse severity and self-organizing potential of the network actors has been provided throughout this analysis. Under certain conditions, especially when repeated collapses occur, semantic structures exhibit bifurcations and easily shift (or oscillate) across possible outcomes.

Thirdly the potential for self-organization of semantic knowledge often has undesired effects, as it results in over-clustering of knowledge flows and less cohesive social semantic structures that achieve limited collective social outcomes. The latter relates to findings emerging from connectionism theory of knowledge affirming the potential to semantically activate knowledge in socially meaningful ways.

Going beyond the technical characteristics of the semantic network analysis provided in this paper, some broader methodological issues require our attention. Computational approaches to semantic network classification and analysis using qualitative methodologies and psycholinguistic techniques have not been widely adopted and tested insofar. Computational social science needs to discover and explore further the real power that the bridging of semantic knowledge representation with qualitative social methods used in sociology, anthropology social psychology and other domains involved in the modeling of human dimensions. This can lead to enhanced understanding of the complexity of interactions in coupled natural-anthropogenic systems especially in the face of broader social and environmental challenges for the future.

Our study shows that we can recreate semantic representations of knowledge capable of emulating semantic emergence matching our empirical findings. Nevertheless we lack further empirical findings from real-world participatory settings capable to form a core evidence-base. Our scientific ability to disentangle theoretical and epistemological frameworks of collective semantic knowledge depends on the availability of additional future empirical settings and experiments.

Table 6
A summary of the inputs and output of the study results by analysis stage and type.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Factorial input</th>
<th>Factorial output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic network extraction</td>
<td>Natural language text from qualitative discourse and interview transcribed texts</td>
<td>Initial activation weights based on co-occurrence frequencies and linguistic weights based on extracted 3-g</td>
</tr>
<tr>
<td>ANN semantic knowledge representation</td>
<td>Semantic activation weights from previous step</td>
<td>Trained optimized semantic weights for semantic network</td>
</tr>
<tr>
<td>Agent-based ensemble simulation</td>
<td>Semantic network structure (#nodes, #links, centrality, preferential attachment based on power law distribution)</td>
<td>10,000 + Simulated semantic network ensembles based on time-steps, severity of collapse, frequency of collapse and self-organization potential</td>
</tr>
<tr>
<td>Longitudinal panel variable analysis</td>
<td>Semantic network ensembles from stage 3 (r = [1, 100]; s = [0, 1]; r = [0, 1]; c = [10, 100])</td>
<td>Panel factor statistical estimation using a GEE population-average model selection</td>
</tr>
<tr>
<td>MLPer</td>
<td>Semantic network ensembles from stage 3 (r = [1, 100]; s = [0, 1]; r = [0, 1]; c = [10, 100])</td>
<td>GLM univariate model analysis for estimating marginal effects of mean ensemble panel parameters simulation ensemble pattern estimation, and classification accuracy estimation of the MLP classifier</td>
</tr>
</tbody>
</table>

**Fig. 13**, and the very high AROC values for the two extreme maximum degree classes (where max(d) < 4 or max(d) > 10), the MLP classification performs very well for the top 50–70% of the cases classified.

In summary, the following Table 6 encapsulates all the input factors and outputs of the different analysis stages used in this analysis, demonstrating a true multi and cross-disciplinary and holistic approach to understanding the complexity of interactions in large computational social science simulation ensembles. Our analysis aims to demonstrate the inferential power of combining a long tradition of computational social sciences and simulation ensembles framework with a robust statistical estimation that enhances our ability to enhance the causal nature of our estimations.

4. Discussion

4.1. Important analysis findings

The analysis performed on the simulated ensembles of network data allowed us to answer a number of critical questions regarding patterns of self-organization in networks of semantic collective knowledge representation. We demonstrated the central role of scale-free distributions in allowing the emergence and propagation of knowledge patterns at the collective social level. We also provided additional robustness in the empirical results obtained via our case studies, as well as independent validation of key emerging results involving the existence of tipping-points and scalability of collective semantic social structures. Throughout the statistical analysis of this paper we are able to raise a number of important points.

Firstly, we showed that collapses without the ability to reorganize have significant fragmentation effects on semantic network structure. Social change involving removal of key collective social institutions without examining the adaptive ability of the remaining social structures can lead to wider fragmentation of social knowledge with highly unpredictable effects. Policy interventions aimed on affecting central social structures should take into account the ability of social actors and institutions to adapt to changes. Social self-organization could be feasible but not without state transitions.

Secondly, sequential collapses on such systems can be quite volatile and unpredictable. It often leads either to total loss of semantic representation, or to reconstruction of the knowledge and re-emergence of scale free distributions. The bimodality of those interactions is sensitive on the balance levels of parametric mixes. Limited evidence of the presence of nonlinear thresholds and tipping points of the interactions between collapse severity and self-organizing potential of the network actors has been provided throughout this analysis. Under certain conditions, especially when repeated collapses occur, semantic structures exhibit bifurcations and easily shift (or oscillate) across possible outcomes.

Thirdly the potential for self-organization of semantic knowledge often has undesired effects, as it results in over-clustering of knowledge flows and less cohesive social semantic structures that achieve limited collective social outcomes. The latter relates to findings emerging from connectionism theory of knowledge affirming the potential to semantically activate knowledge in socially meaningful ways.
4.2. Inferences for guided self-organization

Semantic network analysis overlaps with the aims of guided self-organization as a disciplinary scientific domain in more than one ways. Elements of guidance in self-organizing patterns are present in the form of some of the algorithms used in this study. Training and recursive estimation of the auto-associative memory systems performed by the Hopfield ANNs for calibrating weights of semantic classification falls within the domain of simple self-organizing systems [37,64,74]. Extracted linguistic concepts self-organize their knowledge flows into broader clusters or semantic categories based on objective information-theoretic optimization [30,56,95] of spreading-activation thresholds. To the extent to which such categorizations reflect wider collective knowledge dynamics of social systems, semantic self-organization provides the seeds aiding and improving our understanding of the micro-to-macro link between individual cognitive interactions on the one hand, and societal or institutional interactions on the other [79].

The proposed methodology for studying knowledge flows and knowledge representation at varying social scales of semantic inference is not a product of deterministic social engineering at the theoretical level. It rather emerges simply from mining a set of natural language textual data sources. In this sense, any patterns of emergence and knowledge organization are not designed a priori. The latter comes to play in the “design versus organization” discussion [40,70], supported by empirical evidence in the domains of linked social-ecological systems [1,47]. The network complexity observed in collective semantic knowledge flows both in our empirical findings (Anmatjere semantic network) and our simulated ensembles support the proposition that there exists a level of self-similarity in patterns and structure of semantic organization [83]. Patterns of structural self-organization such as the scale-free network distributions and the small-world effects observed in this analysis, as well as patterns of temporal self-organization (e.g., collapses-reorganization cycles) across network evolution, are all driven by changes in the informational context of the semantic context.

Nevertheless, guided evolution of such self-organizing knowledge systems not always result in lower entropic outcomes (or such of higher informational contexts). Phase transitions and regime shifts are neither uncommon nor improbable outcomes of guided self-organization. Our analysis explored sets of conditions under which semantic knowledge networks fall in states with lower informational context or rapidly collapse. A characteristic example stemming from empirical work is the effect of government interventions in culturally sensitive settings such as small and remote Aboriginal communities in Australia. One of the unintended and high-risk potential outcomes from collapsing collective social structures related to community livelihoods is that those same structures in turn support other collective institutions (such as Aboriginal cultural knowledge, family, and community support structures) that can become fragmented or even distinct over time. In the best case, they experience lack of resilience to adapt to future changes, a fact that the history of lost Aboriginal knowledge and culture across Australia should remind us.

Acknowledgments

The authors wish to acknowledge research contributions from Jocelyn Davies, Paul Box and Hannah Hueneke for the Anmatjere case study field work and analysis.

This research was supported by funding from Commonwealth Scientific and Industrial Research Organization (CSIRO) – Sustainable Regional Development/Indigenous Livelihoods Focal Research Area Project. Additional partial funding was provided by the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. Supplementary funding for the main author has been provided by the NSF Virgin Islands Experimental Program to Stimulate Competitive Research (VI-EPSCoR), award no. 0814417. The views expressed herein do not necessarily represent the views of the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. Supplementary funding was provided by the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. The views expressed herein do not necessarily represent the views of the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC.

The authors wish to acknowledge research contributions from Jocelyn Davies, Paul Box and Hannah Hueneke for the Anmatjere case study field work and analysis.

This research was supported by funding from Commonwealth Scientific and Industrial Research Organization (CSIRO) – Sustainable Regional Development/Indigenous Livelihoods Focal Research Area Project. Additional partial funding was provided by the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. Supplementary funding for the main author has been provided by the NSF Virgin Islands Experimental Program to Stimulate Competitive Research (VI-EPSCoR), award no. 0814417. The views expressed herein do not necessarily represent the views of the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC.

Nevertheless, guided evolution of such self-organizing knowledge systems not always result in lower entropic outcomes (or such of higher informational contexts). Phase transitions and regime shifts are neither uncommon nor improbable outcomes of guided self-organization. Our analysis explored sets of conditions under which semantic knowledge networks fall in states with lower informational context or rapidly collapse. A characteristic example stemming from empirical work is the effect of government interventions in culturally sensitive settings such as small and remote Aboriginal communities in Australia. One of the unintended and high-risk potential outcomes from collapsing collective social structures related to community livelihoods is that those same structures in turn support other collective institutions (such as Aboriginal cultural knowledge, family, and community support structures) that can become fragmented or even distinct over time. In the best case, they experience lack of resilience to adapt to future changes, a fact that the history of lost Aboriginal knowledge and culture across Australia should remind us.

Acknowledgments

The authors wish to acknowledge research contributions from Jocelyn Davies, Paul Box and Hannah Hueneke for the Anmatjere case study field work and analysis.

This research was supported by funding from Commonwealth Scientific and Industrial Research Organization (CSIRO) – Sustainable Regional Development/Indigenous Livelihoods Focal Research Area Project. Additional partial funding was provided by the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. Supplementary funding for the main author has been provided by the NSF Virgin Islands Experimental Program to Stimulate Competitive Research (VI-EPSCoR), award no. 0814417. The views expressed herein do not necessarily represent the views of the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC.

Nevertheless, guided evolution of such self-organizing knowledge systems not always result in lower entropic outcomes (or such of higher informational contexts). Phase transitions and regime shifts are neither uncommon nor improbable outcomes of guided self-organization. Our analysis explored sets of conditions under which semantic knowledge networks fall in states with lower informational context or rapidly collapse. A characteristic example stemming from empirical work is the effect of government interventions in culturally sensitive settings such as small and remote Aboriginal communities in Australia. One of the unintended and high-risk potential outcomes from collapsing collective social structures related to community livelihoods is that those same structures in turn support other collective institutions (such as Aboriginal cultural knowledge, family, and community support structures) that can become fragmented or even distinct over time. In the best case, they experience lack of resilience to adapt to future changes, a fact that the history of lost Aboriginal knowledge and culture across Australia should remind us.

Acknowledgments

The authors wish to acknowledge research contributions from Jocelyn Davies, Paul Box and Hannah Hueneke for the Anmatjere case study field work and analysis.

This research was supported by funding from Commonwealth Scientific and Industrial Research Organization (CSIRO) – Sustainable Regional Development/Indigenous Livelihoods Focal Research Area Project. Additional partial funding was provided by the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC. Supplementary funding for the main author has been provided by the NSF Virgin Islands Experimental Program to Stimulate Competitive Research (VI-EPSCoR), award no. 0814417. The views expressed herein do not necessarily represent the views of the Australian Government Cooperative Research Centre Program through the Desert Knowledge CRC.

Nevertheless, guided evolution of such self-organizing knowledge systems not always result in lower entropic outcomes (or such of higher informational contexts). Phase transitions and regime shifts are neither uncommon nor improbable outcomes of guided self-organization. Our analysis explored sets of conditions under which semantic knowledge networks fall in states with lower informational context or rapidly collapse. A characteristic example stemming from empirical work is the effect of government interventions in culturally sensitive settings such as small and remote Aboriginal communities in Australia. One of the unintended and high-risk potential outcomes from collapsing collective social structures related to community livelihoods is that those same structures in turn support other collective institutions (such as Aboriginal cultural knowledge, family, and community support structures) that can become fragmented or even distinct over time. In the best case, they experience lack of resilience to adapt to future changes, a fact that the history of lost Aboriginal knowledge and culture across Australia should remind us.

Acknowledgments

The authors wish to acknowledge research contributions from Jocelyn Davies, Paul Box and Hannah Hueneke for the Anmatjere case study field work and analysis.