Environmental Modelling and Software 109 (2018) 202–222

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Semantic knowledge network inference across a range of stakeholders and communities of practice

Kostas Alexandridis[a,b,*,1], Shion Takemura[c,1], Alex Webb[a,2], Barbara Lausche[d,2], Jim Culter[e,2], Tetsu Sato[c,2]

[a] Institute for Geocomputational Analysis and Statistics (GeoCAS), Center for Marine and Environmental Sciences, University of the Virgin Islands, St. Thomas, VI, USA
[b] Computational and Computer Science Department, College of Science and Mathematics, University of the Virgin Islands, St. Thomas, VI, USA
[c] Research Institute for Humanity and Nature (RIHN), Kyoto, Japan
[d] Marine Policy Institute, Mote Marine Laboratory, Sarasota, FL, USA
[e] Benthic Ecology Program, Mote Marine Laboratory, Sarasota, FL, USA

ARTICLE INFO

Keywords:
Semantic networks
Knowledge networks
Social-ecological systems
Integrated local ecological knowledge
Latent semantic analysis
Social network analysis

ABSTRACT

This paper provides empirical and experimental assessments of thematic knowledge discourses based on two case studies in the US Virgin Islands and Florida. We utilize a latent semantic indexing analysis over natural language corpus to classify and categorize knowledge categories. We computed TF*IDF scores and associated co-occurrence Jaccard similarity scores to construct semantic knowledge networks. Using network analysis, we computed structural metrics over four composite groups: neighbor-based, centrality, equivalence and position. The analysis show that structural network characteristics of environmental knowledge can exponentially predict associations between knowledge categories. We show that connectivity play a critical role on acquisition, representation, and diffusion patterns of knowledge within local communities. We provide evidence of a global prevalence of a shared knowledge core. We show that core social-ecological attributes of knowledge follow scale-free, power law distributions and stable, equilibrium network structures. We identify two distinct models of bidirectional translation: a bottom-up and a top-down.

1. Introduction

In the last decade multiple studies have underlined the importance of the relationship between knowledge and environmental conservation. In many of these studies, *traditional ecological knowledge (TEK)* and/or *local ecological knowledge (LEK)* is/are shown to have strong connections with environmental conservation and social-ecological stewardship (Basurto et al., 2013; Becker and Ghimire, 2003; Berkes and Turner, 2006). Others tackle the relationship between scientific knowledge and sustainability (Kristjanson et al., 2009). Villa, Athanasiadis, and Rizzoli (2009) for example, provide a review of semantic knowledge models addressing ecological and environmental modeling applications, and discuss the broader adoption and feasibility of new approaches. Rivera, Minsker, Work, and Roth (2014) use a text mining framework to classify and develop sustainability criteria and indicators at a regional scale. In many of these studies, the concept of knowledge is addressed as a rather abstract and in a somewhat descriptive manner (Kiptot, 2007; Tàbara and Chabay, 2013). In other studies, the content of knowledge was evaluated against decisions and behaviors related to environmental conservation and action in these local systems (Dutta, Morshed, Aryal, D'Este and Das, 2014; Grant and Berkes, 2007). And, in some studies, knowledge was measured as a list on survey responses (Cinner et al., 2010). Often, the concepts of collective knowledge and social learning are used interchangeably and without a clear set of definitions and/or boundaries related to their perspective functions and effective roles in the social-ecological systems. On the other hand, many studies focus on formal or formalized ontologies and ontological frameworks (Martínez-García et al., 2018; Polhill et al., 2016).

Collective knowledge systems interact and operate across the full extent of social-ecological systems. They incorporate individual and cognitive characteristics of knowledge, social perceptions of reality

* Corresponding author. Institute for Geocomputational Analysis and Statistics (GeoCAS), University of the Virgin Islands, 2 John Brewers Bay, St. Thomas, VI 00802, USA.
E-mail address: kalexan@uvi.edu (K. Alexandridis).
[1] KA, ST: These authors contributed equally to this work.
[2] AW, BL, JL, TS: Organized and contributed participant study focus groups/workshops design, implementation, and data collection.

https://doi.org/10.1016/j.envsoft.2018.08.026
Received 3 May 2018; Received in revised form 8 August 2018; Accepted 23 August 2018
Available online 29 August 2018

(Berger and Luckmann, 1967), ecological reflections of reality, as well as institutional and governance dimensions. Knowledge systems exist and function in the heart of informal institutions and social norms, but also directly relate our everyday social realities to formal institutional rules and arrangements. Social-ecological stewardship critically depends and builds upon existing collective knowledge structures.

1.1. Effective institutional governance in managing social-ecological commons

Our ability to examine and analyze the efficiency, efficacy and effectiveness of our governance systems and the institutional arrangements in place and at work towards managing our commons critically depends first on the context, nature and characteristics of the governance systems themselves. It also depends on the weak network or web of connections between institutional processes that form management functions, and key systemic components of the linked social-ecological system. The presence of weak links or ties between core institutional processes and both social and ecosystem functions are often the catalytic drivers of institutional change.

Efficient governance systems with respect to social-ecological system management can function both to the benefit of the ecosystem services and functions of the natural system, and to the strengthening of the social system and the communities of practice within the management scales of reach. From the ecological standpoint, ecosystem health and ecological resilience are among the most important functions and processes that one needs to pay attention to, albeit a number of secondary ecosystem and landscape processes bare significance to the analysis. From the social standpoint, the triplet of cognitive/dispositional, collective or social and institutional interactions represent important components of such an analysis (Fig. 1).

In social-ecological systems, the governance of the commons more often than not emerges as a function of local community social and ecological stewardship. Without it is difficult if not impossible to achieve governance structures that have the necessary legitimacy, power and control to negotiate efficient and adaptive institutions of change (Cowie and Borrett, 2005; de Vos et al., 2013). It is exactly because such adaptive institutional arrangements reflect broader community and societal goals or aspirations, the presence of a synergistic relationship between local stewardship and governance of the commons is critical (Ghorbani et al., 2017). On the other hand, stewardship at the community level alone cannot successfully achieve social-ecological sustainability and resilience in systems where multi-governance of commons is required. This is because of systemic and institutional externalities entering, and likely affecting the social-ecological local system in question.

The degree to which a successful level of management is achieved in the dual stewardship – governance of the commons system perhaps has something to do with the ability and capacity of transference across linked systems and domains of knowledge. The capacity of the dualistic system of knowledge transference is inevitably linked to the capability set and network of interconnected associations across and within the key players or actors of both systems. Such systemic interconnectedness critically depends on the functioning and roles of *knowledge producers, knowledge translators (bi-directional), and knowledge diffusers* in the system. The three roles are complimentary, mutually reinforcing, and non-exclusive of each other.

1.1.1. Knowledge producers

They emerge and function at both scales of the two subsystems (local/global). Yet, the nature of the knowledge produced at each scale differs substantially, albeit dependent of each other. Local knowledge producers operate at the community scale, in the fringe of the natural-social prevalence of the phenomena/problems. They acquire and produce knowledge by merging and combining empirical or observational data/information with local and traditional ecological knowledge, as well as with scientific knowledge. The produced knowledge, more likely than not is raw, unprocessed, and customized to fit the level of local community understanding of reality and the community mental knowledge representation. Global knowledge producers at the level of the governance of the commons often produce knowledge that adheres to strategic and policy objectives and social-ecological imperative realities in which the commons exist and operate. They accumulate and synthesize new knowledge by combining and contextualizing (or re-contextualizing) existing knowledge emerging from the local community mental model interactions. The level of efficiency in such contextualization and synthesizing of new knowledge depends to the degree in which the latter finds its way or reaches their operant level or scale of perception.

1.1.2. Knowledge (bidirectional) translators

They operate often at the margins and "neutral zone" between the local and higher scales of inference. Their role is to bi-directionally translate knowledge from the one subsystem to another and vice-versa (Sato, 2014). They often seek and assist stewardship efforts at the community level, all the while promoting cooperation and coordination among other players and roles at higher levels (e.g., decision makers, managers, policy makers, governance and institutional players,

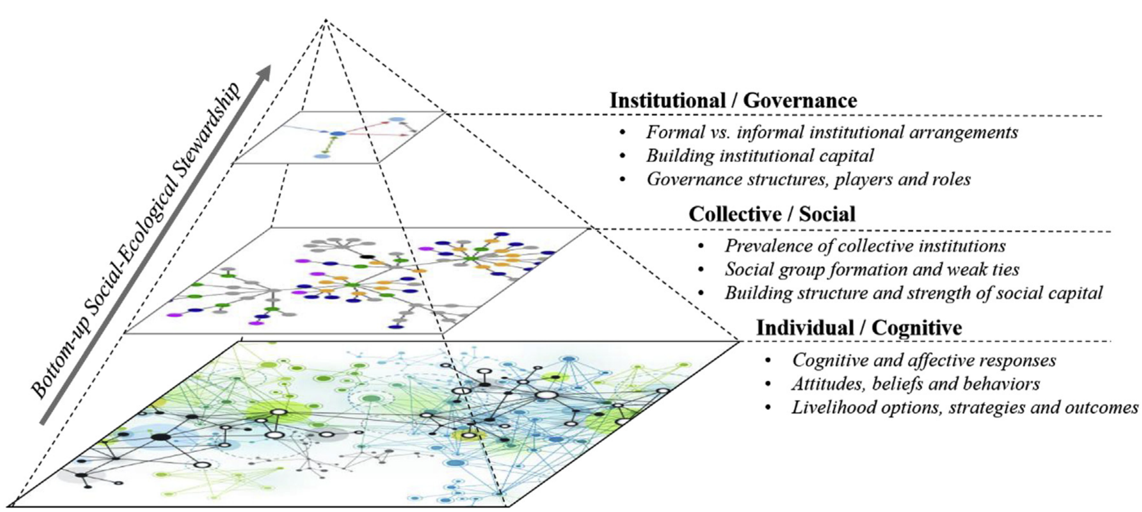

Fig. 1. Conceptual pyramid structure illustrating a bottom-up social-ecological stewardship building from the individual through collective to institutional considerations.

scientists, organizations, etc.). Their role is critical as "enablers" or bridges connecting small-world types of networks that they otherwise would not be connected without their brokerage. By assuming this role, they open knowledge transference pathways capable of producing a wide array of interesting and fascinating systemic phenomena. Such phenomena include the emergence of power-law type of relationships in knowledge networks; generative social emergence across multiple social and ecological scales, and; perhaps importantly, keeping open bi-directional pathways of communication and socially-relevant situational understanding of reality. The latter represents a necessary condition in negotiating appropriate and efficient institutional arrangements required in managing and governing the commons.

1.1.3. Knowledge diffusers

They represent a group of social actors capable of transforming knowledge structures into tangible, real, and beneficial social-ecological outcomes. They are the ones that complete the cycle from information to knowledge to a particular type of actions. Such actions represent knowledge-relevant or knowledge-intensive decision and policy making. The context, content, and extent of their knowledge diffusion establishes the conditions and situations necessary for the emergence of adaptive proactive and reactive action. While action itself does not necessarily implies or requires knowledge transference, the type of action that knowledge diffusers enable is such that ensures that the options and choices available have a direct and undisputable mapping into real and clear outcomes with societal significance. This clear mapping between choices and (distinct) outcomes necessitates knowledge as critical and fundamental element of each strategic mix of choices and outcomes.

1.2. Collective social processes and social-ecological knowledge organization

A significant part of the scientific literature deals with semantic processes as an integral part of semantic cognition. Semantic ability (e.g., activation, association, inter-connectiveness or retrieval) is found to be seriously impaired in the presence of various social illnesses that affect cognitive ability. For example, semantic ability impairment is found to be associated with mental patients suffering from schizophrenia (Chiu et al., 2003; Condray et al., 2003; Paulsen et al., 1996) and other psychotic conditions (Fritzsche, 2003), Alzheimer's disease (Kazui et al., 2003; Rogers and Friedman, 2008), dementia (Graham et al., 2000; Robertson and Köhler, 2007), amnesia and epilepsy (Giovagnoli et al., 2009), various brain or auditory damages (Breese and Hillis, 2004), or other social psychological conditions such as selective retrieval of unwanted memories (Levy and Anderson, 2008). These and other relevant findings indicate how central is our ability to perform cognitive semantic tasks in order to function in a human social environment.

At the collective social level of interactions, the issues go beyond social or cognitive pathologies and indicate strengths or deficiencies of the social system as a whole. If our ability to organize and semantically categorize our knowledge is a proxy of our overall social intelligence (Enfield, 2009), then as societies we benefit and perhaps advance our levels of collective social achievements by taking advantage and exploring this social intelligence in different or higher orders of knowledge organization.

Our ability as organized communities and societies to achieve higher levels of knowledge organization is therefore a characteristic integral perhaps of our social learning capability. In an era where technology and complexity of social interactions advance and develop in parallel, our abilities as social groups to process, reorganize, utilize and functionalize information in semantically rich ways are paramount. New knowledge is generated and required not simply in the form of unprocessed information availability, but as deliberate and functional response to the needs of our social complexity of interactions. The semantic web (Greaves and Mika, 2008) for example does not itself generates new fundamental knowledge, but rather alters its availability and accessibility to wider social audiences, reorganizes it in new and highly functional ways or semantic categorizations, and enhances our ability to perform more and more complex cognitive and socially critical tasks at both the individual and the collective social levels.

In a sense, semantic social knowledge representation encodes and categorizes a world of semantically important objects that belong to the past, present and future alike: the past, through the historical collective evolution of norms, aspirations and beliefs; the present, through values and attitudes; the future through expectations and visions that individuals and local societies hold and pursue. Collective semantic social knowledge representation also encodes both physical and psychological space alike: physical space through socially embedded geographies, historical and cultural spaces and landscapes, geographical narratives and named places; social space through social distances, distinctive social roles and actor-settings, psychological scales and magnitudes, collective dispositions and mapping of shared realities, experiences and norms. This blending of space and time into a unique social continuum is what semantic knowledge representation classifies and encapsulates within its ontologies.

Collective social processes and collective decision making depends on the level of knowledge and its representation in the system of interactions among and across objects of collective inference. In other words, how the system behaves is a matter of complexity of both its parts, but also on its interactions. Systemic interactions, i.e., part-part relationships and part-whole/whole-part interactions, are contextualized in terms of informational or entropic context that is primarily knowledge-based, or at least knowledge-contingent. To the extent that collective networks of interactions (e.g., semantic) encode and encapsulate such interactivity across social actors in the society and community, as evidence suggests in support of this proposition, then collective knowledge representation has the propensity of capturing important social dynamics. Such social dynamics include social emergence, centrality and collective social influence over behavior, decisions, and actions.

If we agree that many social dimensions of change require an inductive approach and cannot be reconstructed (at least empirically) using deductive approaches or social mental models, then we can see that the level of support in often uncertain and incomplete inferences regarding social change can only be achieved by introducing heuristics stemming from evidence at the collective social level of interactions. This is especially true when non-monotonic reasoning is both present and necessary to interpret the evidence at hand. Williams (2009) argues that there cannot be natural necessity in the social world, but only social necessity in probabilistic terms of the word. Natural necessity requires the presence of consistency and regularity (Williams, 2009) which is often and commonly encountered in the physical world (e.g., ecosystem function and processes), but not in the social world of interactions. Non-monotonic phenomena in human judgment (e.g., induction, abduction, belief revision, knowledge-based reasoning) clearly showcase how variation in levels of certainty/uncertainty of evidence or knowledge or information leads to nonlinear and often profound shifts in inferences or conclusions reached (Brewka et al., 2004; Pinkas, 1995; Rott, 2001; Voorbraak, 2004). Furthermore, collective reasoning propositions are probabilistic or probabilogic and variational by default (Chater et al., 2006; Haenni, 2005). In other words, they vary across individuals within populations and across populations within space and time.

1.3. The dual character of collective social representation

There are two key characteristics and functions of collective social representation relevant to this discussion. The first involve collective dynamics as inferential mechanisms that explain and describe social patterns at the collective level. Such mechanisms of inference are necessarily heuristic by nature and serve the purpose of allowing interpretation of social actions in the context of whole communities and societies. The second considers collective dynamics as means of reasoning about societal-wide processes, decisions, behaviors, and actions. Such processes do not necessarily fall under the realm of inferential or heuristic assessments, but rather serve as indicators for explaining the functionality of collective social structures as such, i.e., facilitating the emergence of individual processes by elevating their semiotic or ontological importance and operation at higher levels of collective social hierarchies.

In the first category, we can ask for example, how individuals form and facilitate collective processes such as semantic emergence, social roles and social actors, centrality, prestige, brokerage, etc. In the second category we need to ask how existing collective structures regardless of their constituent formation process allow for individuals to reason about changes, including social change (Oliver et al., 2012).

1.4. Semantic inference and informational propositions

Semantic similarity represents a correlated function of space and time continuum within a certain state of social structure. Semantically similar representations of social reality imply a certain degree of shared understanding of such reality, or otherwise stated, a shared level of knowledge representation within a shared or collective mental mode. Two important parameters and their combined functioning are implied in such propositional formulation.

First, let us examine the case where a number of individuals sharing a certain level of semantic similarity in their knowledge representation, as this emerges through stated or revealed empirical inference. Such semantic similarities are assessed through a socially-explicit or implicit contextual environment or social space. The latter represents an assumptional proposition and relates to the nature of the empirical evidence at hand or the empirical nature of the evidence sampled through our experimental social science methods. Such individuals within such a socially-relevant context are more likely than not to occupy a level of shared functioning within their respective social space or environment. From the individual standpoint, this is to say that there is an expectation of a certain level of overlap within the social universe of their interactions, such as sharing group memberships, having certain similarities on their social and demographic background, or sharing one or more social, economic or cultural characteristics. This social topology albeit not directly assessed in terms of shared spaces, is reflected through their shared semantic knowledge representation implicitly and axiomatically. From the collective standpoint, one can see the social system of individual interactions as a form of topological density distribution, whose boundaries are multidimensional, and whose dimensions in turn represent social characteristics, i.e., the fabric of social structure under study. Regardless the approach taken (from the individual to the social, or from the social to the individual), the level of semantic information encoded in the empirical structure of evidence represents at least in part a judgmental heuristic on the collective social space occupied by the semantic distribution and their associations. This assumptional proposition of course, is rendered more valid as another (testable) assumption of semantic specificity becomes stronger or supported by the empirical data. Consequently, the more general the semantic knowledge representation is, the less warranted may be the assumption of its correspondence with shared/overlapping social space, and therefore, the less likely it may be that semantic similarity relates to collective social interactions. In a latter part of this paper we will explore further the theoretical and empirical consequences of such propositional inference.

Secondly, let us consider the theoretical statistical inference coming from information theory. As early as the late 1950's it is shown by Jaynes that within a statistical (subjective) interpretation of informational entropy, the probability of a certain state can acquire a temporal interpretation, i.e., formulated in the basis of a "(…) fraction of the time that the system spends in state (n)" (Jaynes, 1957, p. 627). Within a socially-relevant analytical framework, such a subjective probability can be viewed as the level of *exposure* conditional to the characteristics of the social structure under study. The matter of temporal conditional exposure inferentially relates to the emergence of semantic similarities within a collective social knowledge framework. As in the previous case, this level of temporal exposure can be viewed both from an individual and from a collective social perspective. From the viewpoint of the individual, the lengthier the exposure (i.e., the fraction of time spend at state n in relation to the total time under study), the more likely it is to either diffuse knowledge or acquire knowledge, or both, at all cases, such knowledge to be assumed to be a probabilistic distribution function influenced by exposure levels. From the viewpoint of the social perspective, the density of social interactions may relate to the density of time exposed to social ideas or social conditional structure. In both viewpoints, exposure can be subsumed to be associated with shared semantic structure of social interactions. On the other hand, the level of association between exposure and semantic cohesion can be negatively influenced by the intensity of the social interactions themselves. Such intensity, like the spatial considerations deepens as specificity increases and weakens with increasing levels of semantic generality.

At the extremes of the spatio-temporal arguments, the informational entropy of the semantic inference is rather trivial. For, as we all exposed to a certain shared level and representation of the world around us (i.e., a level of generality eliciting global or universal social, value or ideological/worldview systems), there is a trivial level of generality that renders scientific inference as uninteresting. Similarly, in the heart of the structure of our social interactions all individuals within a society face the eventuality of having a minimum level of shared understanding of reality, their social environment, and a level of shared exposure at abstract social states (e.g., national or international ideological, general dispositions regarding fundamental aspects of humanity, etc.). Nevertheless, none of the above arguments would mean much if it was not for an important range of statistical discoveries in the areas of computational linguistics and the statistical nature of human language from its semantic knowledge representation point of view. The integration of linguistic, sociological, mathematical and statistical sciences cannot be better served as in the case of the literature observing the theoretical and empirical nature of the semantic structure of human language. Generalized semantic language distributions are shown to follow a mathematical distribution widely known as Zipf's law (Piantadosi et al., 2011).

1.5. Structure of the study and research questions

This study attempts to address three important research questions. First, *what is the structure and characteristics of environmental knowledge across individual stakeholders, communities of practice and institutional arrangements?* Secondly, *how well semantic network representations of environmental knowledge represent collective social processes and interactions?* Thirdly, *How and to what degree the structure and self-organization of semantic networks influence the nature and characteristics of collective knowledge itself?* To answer these key research questions, we employed a multi-stage, multi-dimensional analysis framework, shown in Fig. 2, and explained in detail in the methodology section.

The adopted analysis methodology follows the state-of-the-art literature in terms of semantic network analysis (Dib et al., 2018; Fronzetti Colladon, 2018; L. Li et al., 2017; Reagan et al., 2017; Takase et al., 2016), especially in relation to the body of literature related to environmental knowledge (Redington and Chater, 1997; Wesche and Armitage, 2010) and complex social-ecological systems (Alexandridis and Maru, 2012; Tàbara and Chabay, 2013). Modeling social-ecological knowledge through semantic network inference has provided valuable insights into the complexity of interactions that constitutes our complex coupled human-natural systems of reality. It has similarly enhanced our

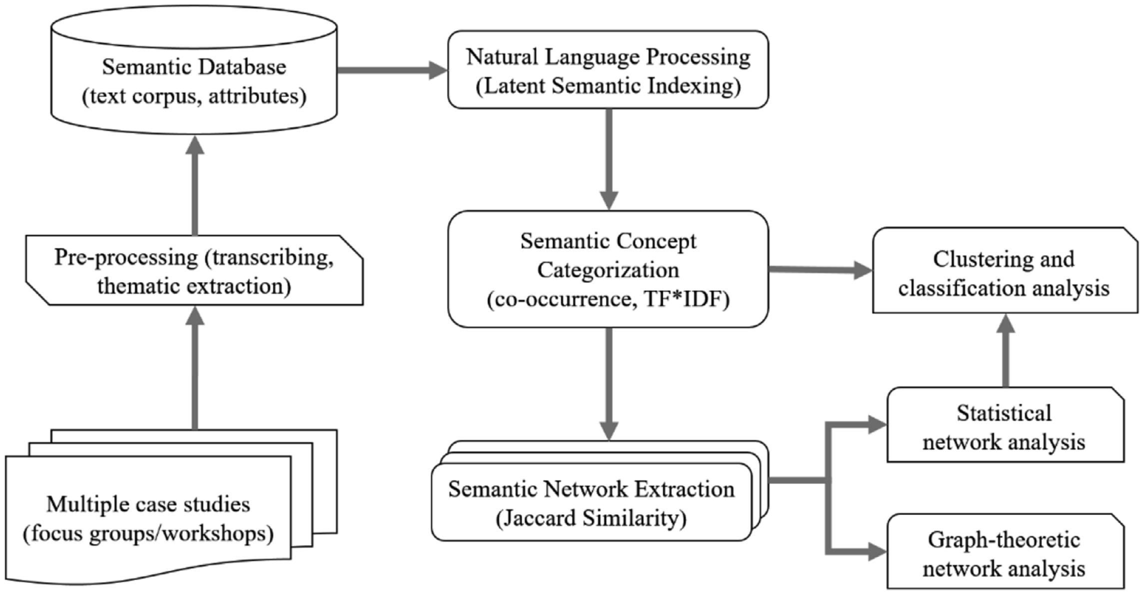

Fig. 2. General overview of the study design framework.

awareness and understanding of the social, institutional and systemic forces that influence and being influenced by environmental change while driving social and economic response mechanisms and strategies (including those related to adaptation to climate change).

2. Materials and methods

The methods used in this study include methods for (a) *data collection methods*, including generating a natural text semantic corpus dataset, with participant and group attributes from qualitative focus group and workshop exercises; (b) *semantic categorization and classification methods*, including performing latent semantic indexing and linguistic categorization, along with clustering and classification of semantically categorized entities, and (c) *semantic network analysis methods*, including graph-theoretic network structural analysis of the data.

2.1. Data collection methods

The data used in the analysis based on five scenario-based focus groups in the island of St. Thomas, US Virgin Islands, and one large multi-stakeholder international workshop in Sarasota, Florida. The US Virgin Islands focus groups were conducted during the period of September 2012 through January 2013, while the Florida workshop was conducted in October 2015. Key characteristics for each of the two case studies are shown in the following Table 1. The key demographic characteristics of participants in the US Virgin Islands are also described further in Webb (2013) and to an extent mirror basic demographic characteristics of the US Virgin Islands at large. The participants in the Florida dataset was based on an international workshop on Integrated Local Environmental Knowledge, and included participants from the scientific community in the Mote Marine Laboratory, local and state agencies, local organizations and NGOs, along with community volunteers. A total of 13 focus group/workshop exercises are included in the analysis (five in VI, and 8 in FL), incorporating narrative perspectives of a total of 57 participants (31 in VI and 26 in FL).

In both case studies, the focus groups and workshop group exercises followed specific thematic protocols in order to ensure consistency of discourses. There were four thematic sections in each of the focus group exercises in the US Virgin Islands, and two thematic sessions in each of the workshop exercises in Florida. Specifically, the thematic entities for the US Virgin Islands were: (i) Discussing drivers of environmental change (DRIVERS); (ii) Defining and discussing environmental sustainability (SUSTAINABILITY); (iii) Developing scenarios for the future (SCENARIOS), and; (iv) Discussing social-ecological resilience (RESILIENCE). Similarly, in the Florida case study, the thematic entities discussed by the workshop participants were: Discussing environmental restoration efforts (RESTORATION), and Discussing environmental stewardship approaches (STEWARDSHIP).

The discourse narratives in both case studies were audio recorded for each group exercise, and were verbatim transcribed. Following transcription, the raw text files underwent a pre-processing sequence:

1. *Identify case studies.* Each case study was assigned a unique study ID (SID).
2. *Separate discourse narrative cases.* In each transcription, a discourse narrative case is a unique set of one or more paragraphs for which a single participant talks during the discourse exchange. Each of these discourse narrative cases represents a unique row on the database, and is identified through a unique variable, named discourse ID

Table 1
Case study participant and group characteristics.

Attribute Characteristics		Virgin Islands	Florida	All Studies
Groups		5	8	13
Participants:	Total	31	26	57
	Per Group	4–9	6–7	4–9
Gender:	Female	16 (51.6%)	9 (34.6%)	25 (43.9%)
	Male	15 (48.4%)	17 (65.4%)	32 (56.1%)
Age Group:	Young adults	17 (54.8%)	7 (26.9%)	24 (42.1%)
	Middle-aged	11 (35.5%)	10 (38.5%)	21 (36.8%)
	Older	3 (9.7%)	9 (34.6%)	12 (21.1%)
Occupation Group:	Government/Agencies	3 (9.7%)	3 (11.5%)	6 (10.5%)
	Local NGO/Community Organizations	7 (22.6%)	9 (34.6%)	16 (28.1%)
	Professional/Business	16 (51.6%)	5 (19.2%)	21 (36.8%)
	Science/Science Organizations	5 (16.1%)	9 (34.6%)	14 (24.6%)

(DID).

3. *Identify discussion theme.* For each transcription we identified and separate each thematic group of the discourse (the theme of the discourse). These were provided with a unique theme ID (TID) across case studies, and across focus group or session replications of the experiments.
4. *Identify participants.* For each discourse ID, we identify the person (participant) who provided it, and was added in each row using a participant ID (PID) variable.
5. *Identify focus group/session.* For each of the case studies, we accumulated transcription data for each of the focus groups/session replications of the experiments. Each group session was given a unique group ID (GID) that identifies the session/group replication.
6. *Identify participant attributes.* Each participant is associated with a number of attribute data, including gender, age, occupation, group (in which they participated). These were added as column variables, for each row associating attributes with participant data.

2.2. Analysis methods

2.2.1. Data pre-processing

The transcribed text data were slightly modified in order to: (a) remove all transcribed instances by the facilitator in the study. During this step, only the discourse statements made by the participants themselves remained in the study; (b) correct grammar or spelling errors and convert spoken linguistic forms to formal ones (e.g., can't to cannot, I'll to I will, etc.), and; (c) converted to a tabular format containing one discourse statement per row, along with key participant, group, and theme characteristics as additional columns. The modified data generated a text corpus database that was used in the semantic analysis of the study.

The final textual corpus database constructed includes 2618 (discourse narrative) cases: 1001 (38.2%) for the Virgin Islands case study, and 1617 (61.8%) for the Florida case study. On average, it contains 45.9 narrative cases per participant (or 3.5% of total narrative cases per participant). While the absolute number of cases per participant differs between Virgin Islands (32.3 narrative cases/participant) and Florida (62.2 cases/participant), their overall fractions as percent of total cases is similar (3.2% for VI and 3.8% for FL). The final dataset also contains on average 45.9 narrative cases per focus group/session or 3.5% of total cases per focus group. In terms of its linguistic content, the textual corpus characteristics and summary statistics are shown in Table 2. The main units of analysis for the main analytical methods used in this study are cases, paragraphs, sentences and words. Thus, the size of the textual corpus dataset has adequate statistical power for the latent semantic indexing and parameter estimation methods used. For more information on the final dataset, see S3 File.

2.2.2. Semantic extraction

Extracting semantically relevant linguistic concepts from the textual corpus involves a number of process steps. First the textual corpus is checked against an exclusionary list, identifying and removing temporarily from the text corpus non-linguistic words, i.e., only linguistically and semantically relevant concepts are kept in the text. The exclusionary list is based on formal English language dictionary. The remaining corpus include verbs, connectors, modalities, adjectives, pronouns, etc. Secondly, the concepts remaining in the linguistically-relevant text corpus are tokenized. Tokenization includes concept substitution through an English lemmatization algorithm, i.e., further reducing any infected spelling of words to its lexical root, reducing words to their canonical forms, including plural/singular forms, past-tense/present tense, etc. (lemma form). Each of the lemma form represents a linguistic token that is used for the categorization process in the next session.

The semantic processing for the textual corpus data in this study reduced the original 188,912 words to 106,144 tokenized words, thus reducing the size of the corpus by 43.8%. The percent reduction was almost identical across the two case studies (43.7% for VI and 43.9% for FL). The process corpus includes 3825 word forms (2427 for VI and 2639 for FL). The ratio between type forms and token words in 0.036 (0.046 for VI and 0.05 for FL). Further summary semantic processing statistics are shown in Table 3.

2.2.3. Concept categorization and association

Semantic concept categorization from textual corpus data depends in its original stages on identifying co-occurrence patterns among linguistically-relevant words within cases, paragraphs or sentences of the corpus. Therefore, worlds with higher co-occurrence frequencies not only appear together in the textual corpus, but also do so in high frequencies (B. Li, Wang and Zhang, 2012). Given the fact that our textual corpus dataset includes multiple participant and focus groups, higher co-occurrence frequencies also reflect a higher level of collective mental models of knowledge representation (Alexandridis and Maru, 2012). In other words, they reflect a more social (rather than cognitive or individual) perception of reality.

On the other hand, semantic similarity in patterns of co-occurrence allows us to generate measures of paired association among concepts (words) or categories (groups of words) with relatively high frequencies of co-occurrence within a textual corpus. Such associations, in the forms of semantic networks are shown to be hierarchical by nature (Crestani, 1997; Ravasz and Barabasi, 2003). Thus, the scalability of such networks allows us to retrieve semantic networks from data ranging from almost full graphs, to abstract graphs. The research question that emerges in such cases, is, what is the similarity threshold that maximizes the informational content (i.e., minimizes the informational entropy) of the knowledge contained within them? This question relates closely to the *isomorphism* property of semantic networks (Alexandridis and Maru, 2012; Chia and Ong, 2006) and the scale-free distribution of semantic association in graph-theoretic terms. Therefore, given a scale-free semantic network, the previous research question becomes a matter of finding the minimum cut-off threshold level of semantic similarity for which the scale-free properties of the network (and thus its' semantic isomorphism) remains unchanged. At such level, adding more nodes to a network who's link distribution follows a power-law, will not change its network structural characteristics, and the network becomes simply saturated.

In order to test the latest proposition, we adopted a study design that (a) fixes the number of nodes in the semantically extracted networks to the top 100 concepts based on their analytic TF*IDF index, and (b) varies the threshold of associative semantic link similarity of co-occurrence patterns by using its percentile distribution.

In terms of the fixed number of nodes, this step is necessary in order to perform comparative network analysis (i.e., the networks vary only on their structure and not by their size). In addition, the TF*IDF

Table 2
Summary statistics of the final textual corpus database in the study.

Case Study	Cases	Paragraphs	Sentences	Words	Paragraph/Case	Sentences/Case	Sentences/Paragraph
Virgin Islands	1001	2002	3994	94,095	2.000	3.990	1.995
Florida	1617	3326	4481	94,817	2.058	2.773	1.347
Total (All)	2618	5228	8475	188,912	1.998	3.238	1.621

Table 3
Summary statistics of the textual corpus semantic processing in the study data.

Case Study	Token Words	Type Forms	% excl.	Words/sent.	Words/par.	Words/case
Virgin Islands	52,971	2427	77.6	13.3	26.5	53
Florida	53,173	2639	78.3	11.9	16.5	33
Total (All)	106,144	3825	78.3	12.5	20.3	41

coefficient, representing the product of the term frequency (how frequently a term appears across paragraphs, sentences, etc.), and the inverse document frequency (discounting terms whose expected within-case frequency is larger). TF*IDF is a commonly used information retrieval metric used in latent semantic indexing (Ai et al., 2010; Rivera et al., 2014; Xagi et al., 2010). Symbolically,

$$TF{*}IDF = TF(t, d)\cdot IDF(t, D) = \left(\frac{f_{t,d}}{\max_{\{t'\in d\}} f_{t',d}}\right)\cdot\left(\log\frac{|D|}{1 + |\{d \in D: t \in d\}|}\right) \quad (1)$$

where, t is a given term (word or category); d is a given document (case) in the corpus; $f_{t,d}$ is the frequency of term t in case d; t' is any other term appearing in case d; D is the number of cases in the corpus. The denominator of the inverse document frequency is adjusted (adding 1) to avoid division with zero and computes the number of cases where the term t appears.

In terms of the number of links in the study, following the concept categorization and association process in the next session, we first computed the percentiles of the semantic similarity appearing in all pairs of the top 100 nodes selected by their largest TF*IDF, and then generated three network versions using the 5% (keeping 95% of the pairs), the 50% (keeping 50% of the pairs, those above their mean semantic similarity coefficient), and the 95% (keeping the top 5% of pairs based on their similarity coefficient). In the analysis section we analyzed all three networks for each of the three cases: a network containing both case studies in a unified corpus; a network containing only the Virgin Islands text corpus cases, and a network containing only the Florida text corpus cases.

The primary concept categorization was created by the data using a simple categorization algorithm, after semantic extraction process described in the previous session. The algorithm involves the following steps:

1. Adds word categories with minimum frequency of occurrence in the corpus $f_{min} = 10$. Only words that appear in at least 10 cases are kept.
2. Calculate TF*IDF values for the kept categories and sort by decreasing TF*IDF values.
3. Keep the top 100 concepts with the highest TF*IDF values, and drop the remaining.

The categorization algorithm was run for three corpus configurations: once for all case study data (combined VI and FL), and once for each of the case studies separately. Each of the configurations generated a list of 100 categories.

In addition to the primary categorization algorithm we also implemented alternative categorization algorithms using known semantic categorization dictionaries, to compute categorization statistics for these alternative classifications. The algorithm computes frequencies of occurrence in the textual corpus for each of the dictionary defined categories (and their relative TF*IDF values). The dictionaries used for the alternative categorization are presented in the next paragraphs.

The *WordStat Sentiment Dictionary* (WordStat, 2016) combines positive and negative semantic words from the Harvard IV-4 dictionary (Stone, 1997), the Martindale's Regressive Imagery Dictionary (Martindale, 1981, 1986), and the Linguistic and Word Count dictionary (Tausczik and Pennebaker, 2009).

The *Martindale's Regressive Imagery Dictionary* (Martindale, 1981, 1986), contains 3000 words classified across two fundamental constructs: primary or primordial processes, and secondary or conceptual thinking processes. According to the relevant literature (Martindale, 1999; Smith et al., 1995) textual corpus with higher fraction of secondary processes are representative of more structured and abstract intellectual capabilities grounded in social realities.

The *Laver and Garry Policy Position Dictionary* (Laver and Garry, 2000), contains 415-policy related words and word patterns over 19 level-2 grouped categories. While originally developed to analyze political positions from policy actors, it does have analytical value for this study, albeit not without limitations. It can provide an indicative and comparative measure to the extent to which policy-related issues form a part of the discourse data narratives. It thus, enable us to contextually evaluate the study's narratives with respect to the policy dimensions of knowledge construction.

2.2.4. Semantic network construction

The semantically extracted and categorized concepts from the two previous stages, were coded back to the original textual corpus, by sentence, paragraph and case. In order to generate associative semantic knowledge network, one additional methodological stage of the analysis was performed. We computed a co-occurrence analysis based on similarity of concept patterns of co-occurrence per paragraph in the textual corpus. We used the Jaccard's coefficient of occurrence to calculate the similarity matrix of pattern association. Jaccard's coefficient of similarity is a well-known statistical measure of association (Leydesdorff, 2008; Real and Vargas, 1996). Symbolically, for rows $x_i = 1, \ldots, n$ and columns $y_j = 1, \ldots, n$ in the $n \times n$ concept co-occurrence matrix, the Jaccard's index of similarity for any two cell elements of the matrix $c_{ij} = (x_i, y_j)$ is

$$J(c_{ij}) = \frac{|x_i \cap y_j|}{|x_i \cap y_j| + |x_i| + |y_j|} \quad (2)$$

where the nominator denotes paragraphs where both concepts occur together, and the denominator denotes the sum of: paragraphs in which concepts co-occur, paragraphs in which the first concept occurs but not the second, and paragraphs where the second concept occurs but not the first. The Jaccard's similarity index range is $0 < J \leq 1$. For each of the case study semantic similarity matrices a number of network attributes were also included from the original attributional dataset by accounting for attributes (demographic, group, theme) including case frequencies in which each concept was identified (e.g., number of cases with female versus number of cases with male participant for each of the network nodes). The use of similarity index is also used in network volatility and event link analysis (Hu et al., 2017) to address the predictive ability and evolution of network structures.

Since we are interested in directional semantic knowledge analysis, we compared the difference in TF*IDF of the concepts in each complementary pairing of relationships, i.e., comparing the TF*IDF difference between the (x_i, y_i) and (y_i, x_i). Our methodological convention is to consider knowledge semantic influence moving from nodes with smaller influence to ones with larger influence (i.e., more central nodes attract less central ones). This methodological assumption is also supported by the preferential attachment property present in scale-free network distributions (Alexandridis and Maru, 2012; Newman, 2001). Thus the directional network conversion algorithm was

$$s'_{ij} = \begin{cases} s_{ij} & \text{if} \quad TF{*}IDF_i < TF{*}IDF_j \\ 0 & \text{if} \quad TF{*}IDF_i \geq TF{*}IDF_j \end{cases} \tag{3}$$

where *i, j* are row and column nodes respectively (from, to). The zero condition applies in cases with identical TF*IDF values denoting the diagonal network cases (i.e., where $i = j$, same node).

2.2.5. Semantic network structural analysis

We computed a number of graph-theoretic network metrics in the semantic knowledge networks of the study. In general, these belong to four group types: (a) neighbor metrics, measuring structural link characteristics of the semantic networks, based on their associative connectivity patters (Drieger, 2013; Knoke and Yang, 2008; Scott, 2000); (b) centrality metrics, measuring structural cohesion and centrality of network nodes overall (Jackson, 2008; Xu et al., 2010; Zhang, 2010); (c) equivalence metrics, measuring the efficiency and effect of connectivity patterns in the network (Reichardt and White, 2007; Sailer, 1978), and (d) positional metrics, clarifying the structural role that various node play in the semantic network (Alexandridis and Maru, 2012; Bader et al., 2008; Haybron, 2000). In addition, we considered and analyzed statistical approaches and models of network structure, such as the d*yadic interaction block-model (p1) (Brendel and Krawczyk, 2010*; *White et al., 1976),* the *non-negative matrix factorization (NNMF)* model (Hasan and Zaki, 2011), the *local outlier matrix factorization (LOMF)* model, *singular value decomposition (SVD) (SivaKumar et al., 2011*; *Watkins, 2004)* and *hierarchical clustering* models (Ravasz and Barabasi, 2003) based on both continuous and categorical network association metrics. Fig. 3 below provides a graphical representation of the graph-theoretic network analysis metrics computed and used in this study (left subgraph), along with additional statistical network models (right subgraph). The metric variables (denoted in red circles) was first computed and analyzed from the semantic knowledge networks in the study. The composite coefficients (green stars) were calculated and used in evaluating the effect of semantic network structure on the formation and functioning of knowledge systems (section 3.2.3). The aggregate variables (blue diamonds) are only shown in the graphs as a level of typological grouping categorization of factors, rather than computed variables.

We further evaluated the presence of scale-free distributions in our semantic knowledge networks. We estimated the hypothesis that the connectivity (Jaccard similarity) and the document corpus importance (TF*IDF) in our data follow a power law cumulative distribution. Symbolically, we estimated the α parameter of the power-law model where

$$P(x) = c{\cdot}x^{-\alpha} \tag{4}$$

using the cumulative probability, $P(x \geq x_{\min})$. The evaluation uses a maximum likelihood method to estimate the alpha parameter (exponent) of the model. We also used the Kolmogorov-Smirnov goodness-of-fit statistic to test the null hypothesis that the cumulative probability distribution follows an exponential distribution with exponent α. This methodology for evaluating scale free distributions in empirical data is described in more details the relevant literature (Barabási and Albert, 1999; Clauset et al., 2009).

3. Results

3.1. Semantic categorization

There is an overall 43.7% overlap across semantic categories in the data between the two case studies, Virgin Islands and Florida. As can be seen in Table 4, the mean TF*IDF and narrative case co-occurrence frequency is 70.8% and 57.4% higher in overlapped categories.

The summary of the semantic grammar of the classified knowledge narratives is shown in Fig. 4 for all case studies. In terms of the verbs used, almost 42% reflect *stative* statements, i.e., statements expressing a specific state or situation. Another 31% are *factive* verbs, i.e., used in statements asserting facts or presupposition of truth statements. The third important category, *reflexive* verbs (26%) are mainly used in statements describing personal or collective actions, or processes involving oneself or social groups.

In terms of connectors, *addition* (43.5% of classified narratives) reflects the additive nature of environmental knowledge both as a process

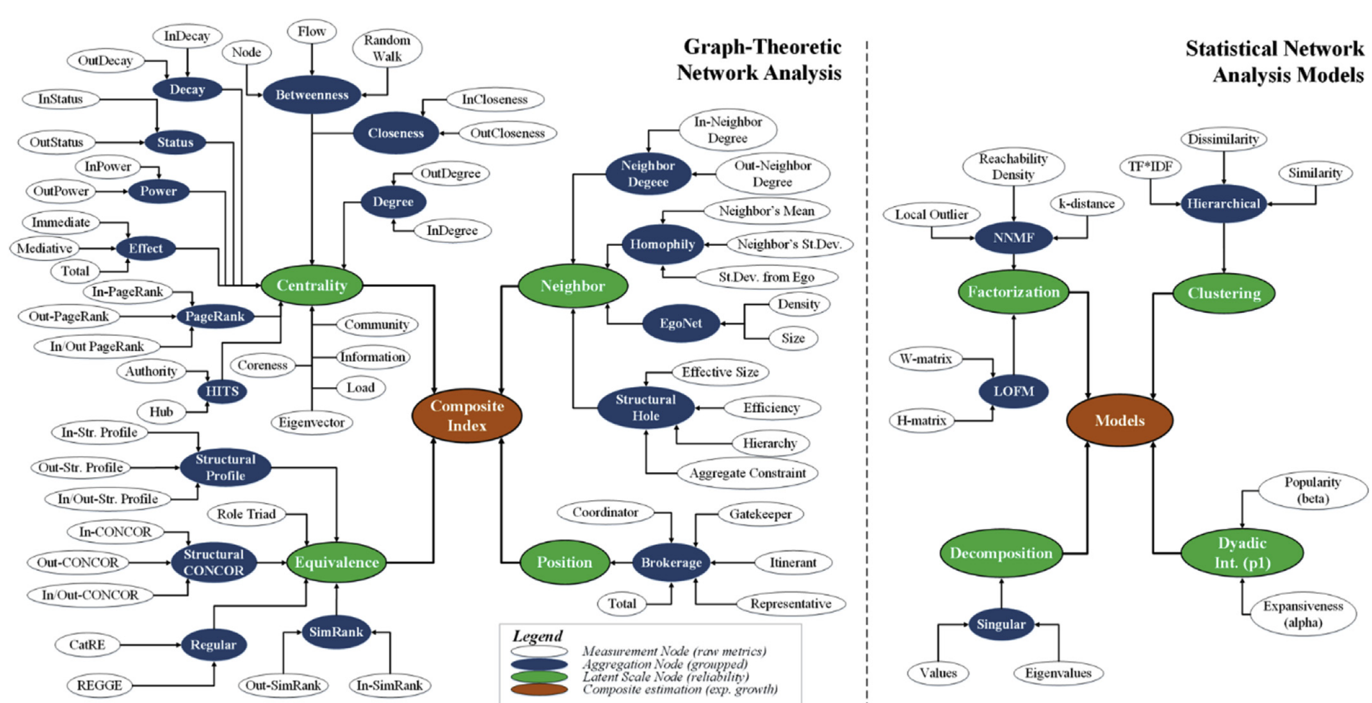

Fig. 3. Semantic graph-theoretic and statistical network analysis structure for the study.

Table 4
Summary statistics for TF*IDF and case frequency by classification category overlap between the two case studies.

Overlap	Variable	Mean	Std.Dev	Min	Max
Unique	TF*IDF	76.13	34.58	37.4	242.3
	Frequency	45.71	25.64	17	177
Both Studies	TF*IDF	107.52	77.38	36.6	566.9
	Frequency	79.69	86.74	19	717

and as an equilibrium state. In terms of semantic modalities, the collective narrative discourse in both case studies reflects a mixture of modal states, including manner, space, time, intensity, negation, assertion and doubt.

The use of adjectives in the collective knowledge narratives in both case studies reflect primarily an objective semantic state (63%), reaffirming the central role of local environmental knowledge as a broader mental representation of the social construction of reality with a degree of subjectivity (26%). This fact is further reflected at the analysis of the pronouns used in the participant's classified narratives. A total of 46.8% of the narratives reflect individual pronouns ("I", "You", "He", "She") covering both objective and subjective adjective statements, while another 33.9% of the narratives reflect purely collective statements ("We", "They", "Somebody"), when participants discuss their knowledge with relation to social realities.

3.1.1. Alternative categorization schemes

In order to understand in better ways the context and content of the narrative environmental knowledge captured in our data, we applied a number of alternative classification schemes in the corpus dataset. Specifically, we classified the narratives using (a) the WordStat Sentiment Dictionary (Loughran and McDonald, 2011; WordStat, 2016); (b) the Martindale's Regressive Imagery Dictionary (Martindale, 1975, 1981), and; (c) the Laver and Gary Policy Dictionary (Laver and Garry, 2000).

The graphs in Fig. 5 below summarize the alternative classification grouped categories for the combined textual corpus data, and for each of the separate case study's corpus subsets. The top row of graphs showcases the application of the *WordStat Sentiments* dictionary classification. The bottom row of the figure summarizes the results of applying the *Regressive Imagery* dictionary classification. In all cases we used the highest-level hierarchical categorization (level 1). In both classification schemes, the darker color bars represent the percent of sentences in the corpus classified within each category, while the lighter color bars show the percent of paragraphs (cases) classified for the same category.

In terms of the positive-negative distinctions captured in the WordStat classification, in all replications of the study, the positive discourse narrative aspects represent at least 70% of the classified cases. The respective negative coverage of the corpus ranges between 50 and 60%.

In terms of the emotional-primary-secondary continuum captured in the RID classification, consistently the % of cases classified as secondary ranges from 62% to 75%, with primary categories ranging from 46% to 60%, and the emotional ones ranging from 17% to 34%.

Similarly, Fig. 6 summarizes the *Laver and Garry Policy Positions* dictionary categories (level 1) for the combined corpus data, and for each of the case studies separately. The classification results show consistently economic aspects of policy-related discourse occupying a significant percentage of the textual corpus coverage (between 24% and 30%), followed by cultural policy dimensions (coverage ranging between 12.5% and 24.5%).

3.1.2. Categorization by attribute characteristics

We analyzed the original semantic categories with respect to the participant, group and theme attribute characteristics. In terms of the key participant demographic characteristics, Table 5 shows the case frequency of the top (up to five) semantic categories in both case studies

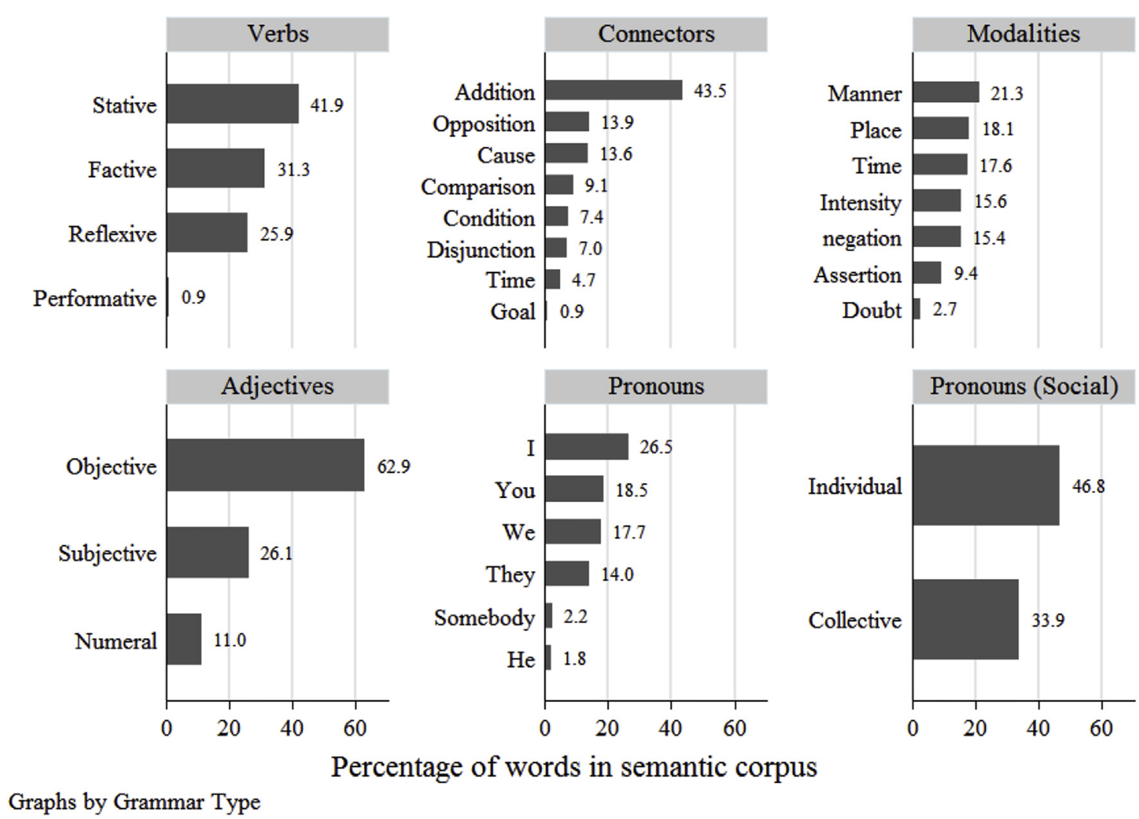

Fig. 4. Linguistic grammar analysis of semantic knowledge narratives corpus data.

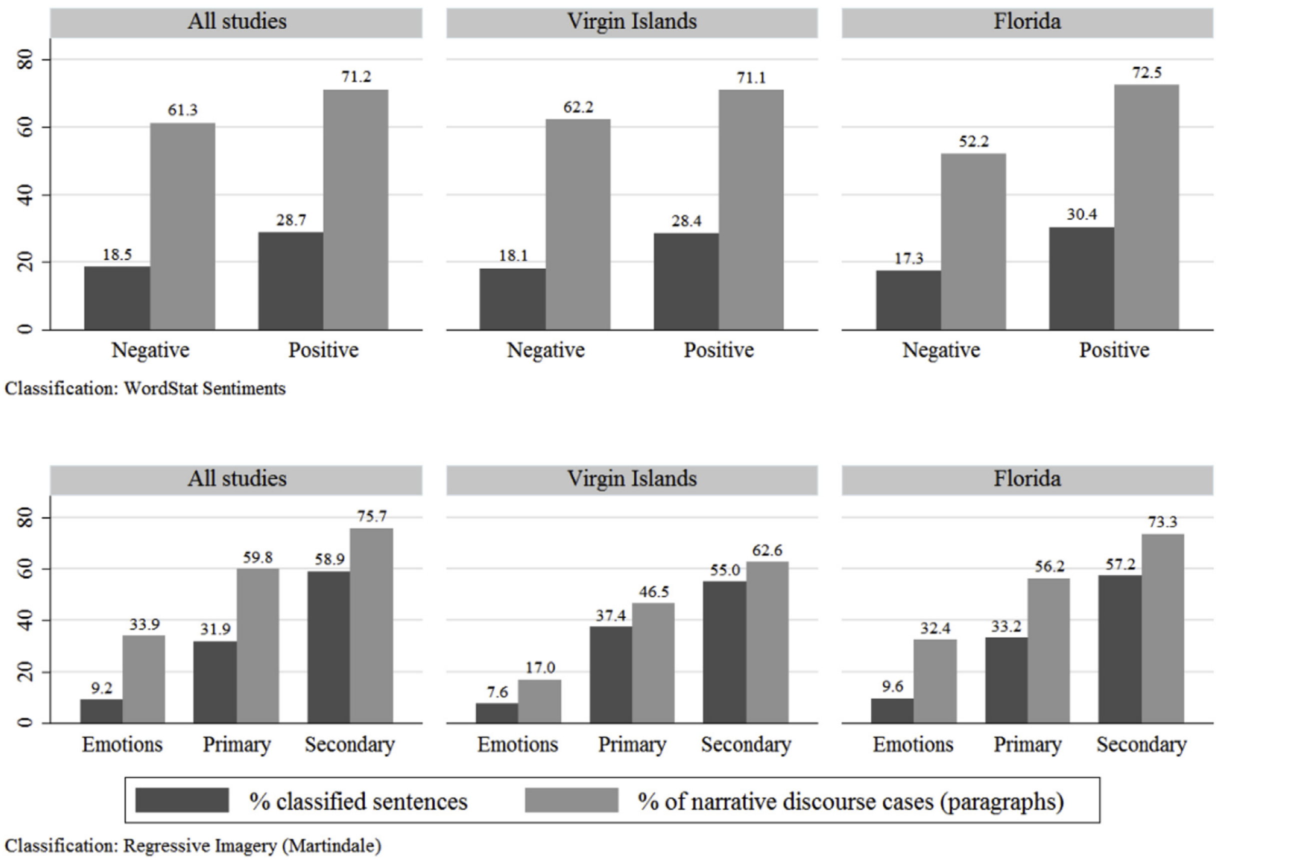

Fig. 5. Alternative document classifications for the semantic knowledge narratives in the study. The top row of graphs uses the WordStat sentiments classification scheme (CIT); the bottom row uses the Martindale's Regressive Imagery Dictionary (RID) classification.

by gender and age group.

In terms of the gender attribute, the top five categories statistically significant favoring male over female participants ($p < 0.001$), are more or less representative of objective outcome realities and situational characteristics of social-ecological realities. They focus on local products, agriculture, or market in general in the Virgin Island case and on restoration, water- and bay-related issues, along with time-related issues in the Florida case. On the other hand, the top categories that favor female over male participants, have generally less statistical power than the ones for the male participants. They focus more on

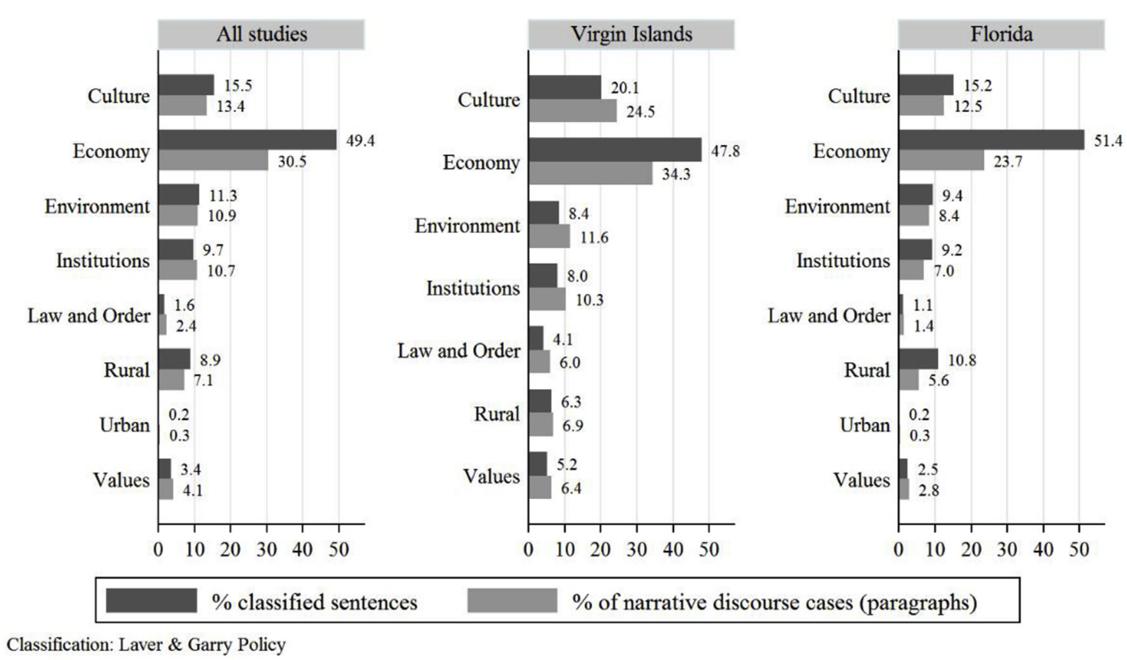

Fig. 6. Policy classification (Laver and Garry, 2000) scheme of the semantic knowledge narratives by case study.

Table 5
Frequency of top (up to five) categories in Virgin Islands and Florida case studies by gender and age group. In each case study the categories favoring a specific gender or age group are shown as boldfaced. Only categories with statistical significance *p < 0.05* for their respective group differences are shown.

Case Study	Category	*Gender*		χ^2	Category	Age group		χ^2
		Female	*Male*			Young adult	Other	
Virgin Islands	*LOCAL*	8	**44**	48.8***	*AGRICULTURE*	8	**29**	36***
	PRODUCT	1	**21**	31.3***	*FARMER*	3	**19**	29.5***
	TRY	30	**59**	31.1***	*GROW*	11	**28**	27.9***
	AGRICULTURE	7	**30**	29.7***	*CULTURE*	8	**24**	26.7***
	STAY	4	**22**	24.4***	*LACK*	9	**23**	23***
	COMMUNITY	**128**	30	23.5**	*JOB*	**48**	6	11.1***
	ENFORCEMENT	**30**	3	11.5**	*START*	**50**	8	9.06***
	CAR	**22**	1	10.9**	*IMPACT*	**29**	29	8.34***
	ISLAND	**152**	128	7.7**	*SCHOOL*	**35**	5	7.15***
	WASTE	**19**	2	7.1**	*ACTION*	**36**	32	6.81***
Florida	*BAY*	11	**106**	21.4***	*STUDENT*	13	**23**	14.4***
	SCALLOP	24	**153**	20.0***	*SCHOOL*	14	**27**	13.6***
	MEETING	**17**	10	15.4***	*EDUCATION*	7	**9**	11.6***
	WATER	23	**134**	15.3***	*MARINE*	11	**25**	8.05***
	RESTORATION	5	**58**	13.3***	*FISHERMAN*	8	**16**	7.33***
	TIME	18	**107**	12.6***				
	FEEL	**14**	12	7.99**				

(*) p < 0.05; (**) p < 0.01; (***) p < 0.001.

institutional arrangements, processes and generally more subjective perceptions of reality. In fact, in the Virgin Islands case, female participants talk more about community and enforcement issues, general environmental issues island-wide. In the Florida case study, female participants are more differentiated by discussing issues related to feelings, meetings and other similar subjective aspects of reality.

In terms of their age group characteristics, the top categories favoring higher frequencies for young adults than any other participants are ones referring to job opportunities, start, general impacts, school and action in the Virgin Islands case study. Given that most of the young adult participants in the VI case are school teachers and young volunteers, it is clear that these discourse topics reflect a more participatory, experiential and bottom-up reality driven narrative. On the other hand, in the Florida case study, the same categories (e.g., school, student, education), differentiate more the other age group categories.

The multi-dimensional scaling analysis by attributes of the semantic text corpus data are shown in the following Figs. 7 and 8. In all cases the first three dimensions, have eigenvalues above 1.0 and they explain on average 84.4% of the variability in the data (76.5% for VI, 80.5% for FL, and 96.1% for the combined case studies). The classified semantic categories are plotted using their estimated multi-dimensional scaling coordinates in each of the paired dimensions (with the third dimension shown as a contour). Then, the mean x, y and z coordinates for each of the focus group in each of the case studies are calculated, and plotted in the same graph. Nodes or categories closer to the axe's center (i.e., axes with value of zero) are closer to the mean values of each dimension.

Fig. 7 plots the semantically classified knowledge categories against the study's focus groups representing relatively cohesive social, professional and institutional participant groups.

Similarly, Fig. 8 plots the semantically classified knowledge categories against the broader occupational groups of the participants. We generalized specific occupations across both case studies to differentiate between participant roles as scientists, educators, business/professional, government/agency participants, or community/NGO groups.

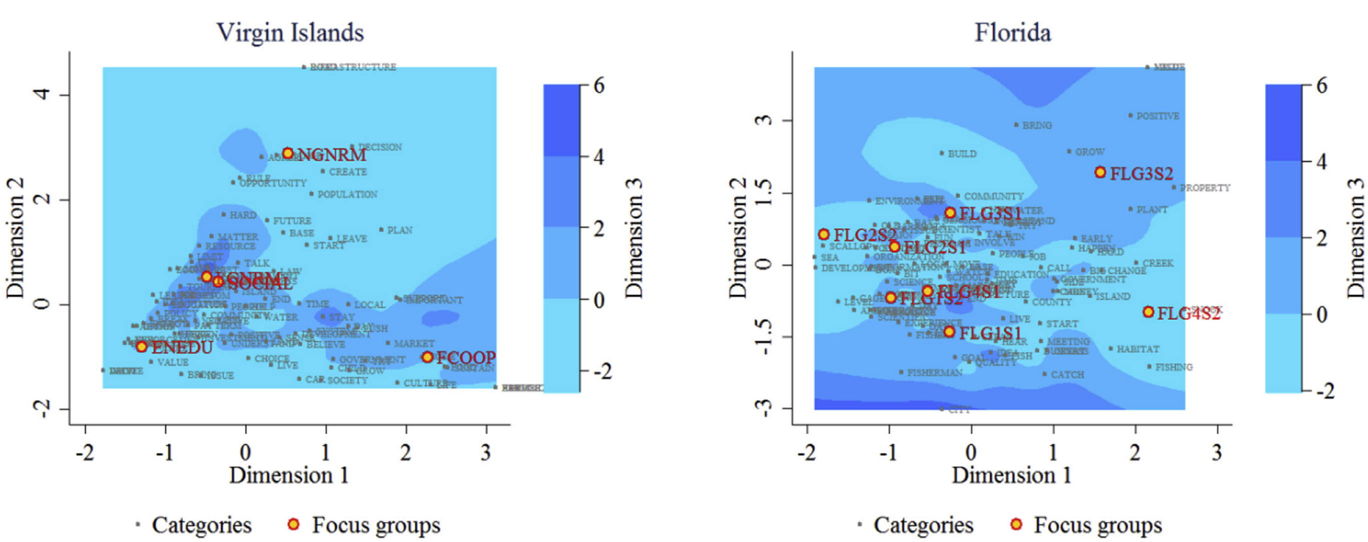

Fig. 7. Multi-dimensional scaling (MDS) plots of the semantic categories by focus groups in the two case studies. The x and y-axes coordinates plot the first and second MDS coordinates, while the color contours plot the third MDS dimension of the analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

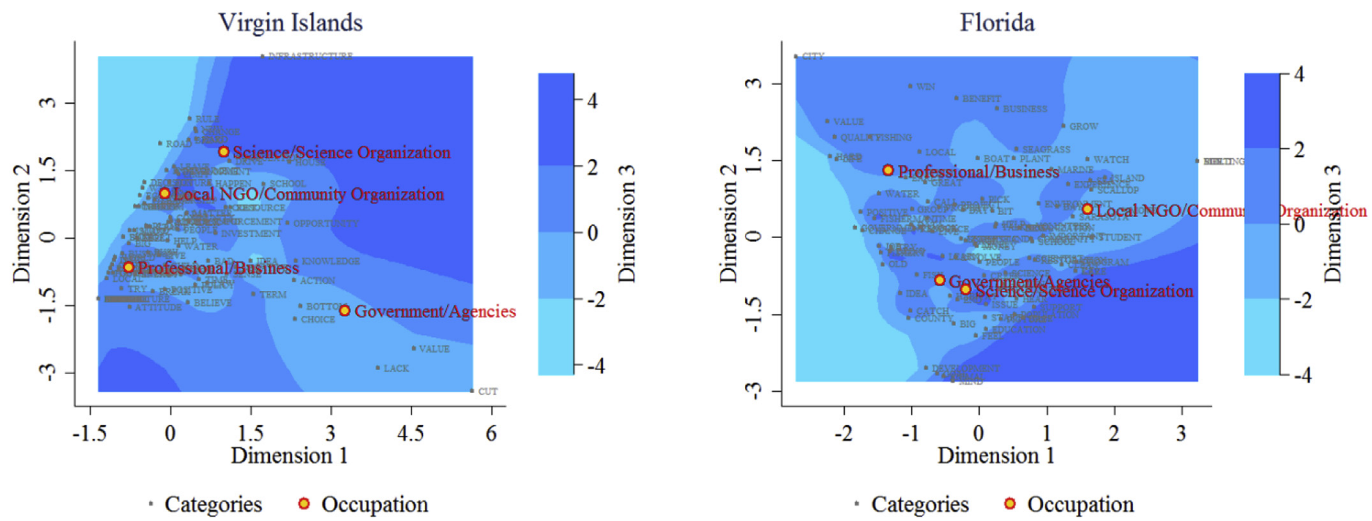

Fig. 8. Multi-dimensional scaling (MDS) plots of the semantic categories by occupation type in the two case studies. The x and y-axes coordinates plot the first and second MDS coordinates, while the color contours plot the third MDS dimension of the analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. *Semantic knowledge network analysis*

3.2.1. *Summary of key semantic network metrics*

Fig. 9 provides a visual network representation of the neighbor in-degree distribution among semantic nodes in the knowledge network for the case study narratives. Nodes with larger diameter have higher neighbor in-degree coefficient, which indicates the degree of semantic knowledge in-flows from the periphery to the center, weighted by the semantic similarity of the links. Three types of nodes based on their coefficients are identified: ordinary nodes, receiver-type nodes, and transmitter-type nodes. The receiver nodes are generally receiving strong semantic knowledge influence flows, while the transmitter nodes are strong sources of semantic influence flows.

The *structural hole network analysis* for the classified narrative data provided some very interesting results. Specifically, the degree of *redundancy* in the network connectivity (link connectivity) is 27.2% for the network constructed with all studies' narratives, 22.1% in the Virgin Islands network and 24.1% in the Florida network. The opposite results hold true for the network's *efficiency*, which represents 1-*redundancy* summed over all network nodes. The efficiency coefficient for all narratives is 67.9%, while the efficiency for the Virgin Islands network is 70.6% and 69.2% for the Florida network.

Another important neighbor network analysis concept is that of *homophily*. Neighbor's homophily analysis shows us the degree to which the nodes of a given network are similar in a certain property with it's neighbor's (i.e., the nodes with which the focal node is connected with). Homophily reflects the property of “birds of a feather, flock together”. We analyzed our semantic knowledge network homophily characteristics with respect to their TF*IDF classification indexes. We compared each node's TF*IDF value (ego) with the mean TF*IDF value of it's neighbors (alters), and computed the mean differences across all nodes in each knowledge network. We finally, compared these differences within the size of the network, using the 5% (expansive), 50% (medium), and 95% (smaller) percentiles of network sizes. Our results indicate a decreasing pattern of differences in TF*IDF values between ego and altars. Specifically, the difference for the expansive network is 30.0% (from 89.84 to 62.93), for the medium-size network drops to 27.4% (from 89.84 to 65.25), and drops further to 18.9% (from 93.97 to 76.25) for the smaller network.

We computed statistics for the *shortest path* between any two nodes in our semantic networks. The shortest path algorithms identify the geodesic (the path with the shortest possible distance) from all possible alternative path connectivity between two nodes in a network. We computed summary statistics of the shortest path distance (measured in standardized Euclidean distance, from 0 to 1, with 1 = network diameter). We used the 95% percentile networks which reflect the top 5% of the connectivity relations in the network. The mean geodesic distance for the combined case study network is 0.115 (0.109 for the Virgin Islands case study, and 0.131 for the Florida case study). The percent of reachable nodes outwards (from the center to the periphery) is 30.7% of the nodes (VI: 28.7%, FL: 27.0%). On the contrary, the percent of reachable nodes inwards (from the periphery to the center) is 12.8% of the nodes (VI: 13.2%, FL: 15.3%). Our results also show that the percentages of out-reachable nodes and in-reachable nodes tend to become more equalized as the network size increases (because of the scale-free nature of link relationships).

The results of implementing a path-finder network algorithm (PFnet) based on Euclidean distance dissimilarity are shown in Fig. 10. The PFnet dissimilarity process implements a minimum spanning tree network algorithm (Nepomniaschaya, 2006; Schvaneveldt and Norwood, 1990), i.e., finds the optimal tree structure in a network that minimizes the sum of TF*IDF similarity coefficients in the semantic network. The resulting pruned networks represent the minimal path structure of the semantic knowledge that keeps the network connected.

In the two case studies, the PFnet dissimilarity algorithm yields a semantic knowledge tree structure that includes 102 out of 3740 links in VI and 102 out of 3875 links in FL (2,73% and 2.63% of the full semantic networks respectively), using an average 7.6% of the total TF*IDF similarity weights.

The *status (Katz) centrality* coefficients for the semantic networks of the two studies show that in-status centrality (from the periphery to the center of the network) is decreasing by almost 50% in comparison with the out-status centrality in both case studies. These results are clearly quantifying the relative importance and influential role that central knowledge structures play in the network.

Four additional network structural analysis metrics yielded statistically significant differences between the two case studies: the neighbor-level *ego-density* (larger in VI), the *information centrality* (larger in VI), the *local reachability density* factor (slightly larger in VI), and the *k-distance* factor (larger in FL). Since the distributional assumption of normality is violated in all four variables (the Kolmogorov-Smirnov test for normality was statistically significant), we used non-parametric testing for the difference. Both the Kruskal-Wallis rank test for equality of populations (KW) and the Wilcoxon (Mann-Whitney) rank test (MW) had statistically significant chi-square (KW) and z (MW) values: for the *ego-density*, $\chi^2 = 98.12$, $z = 9.91$ $(p = 0.0001)$; for the

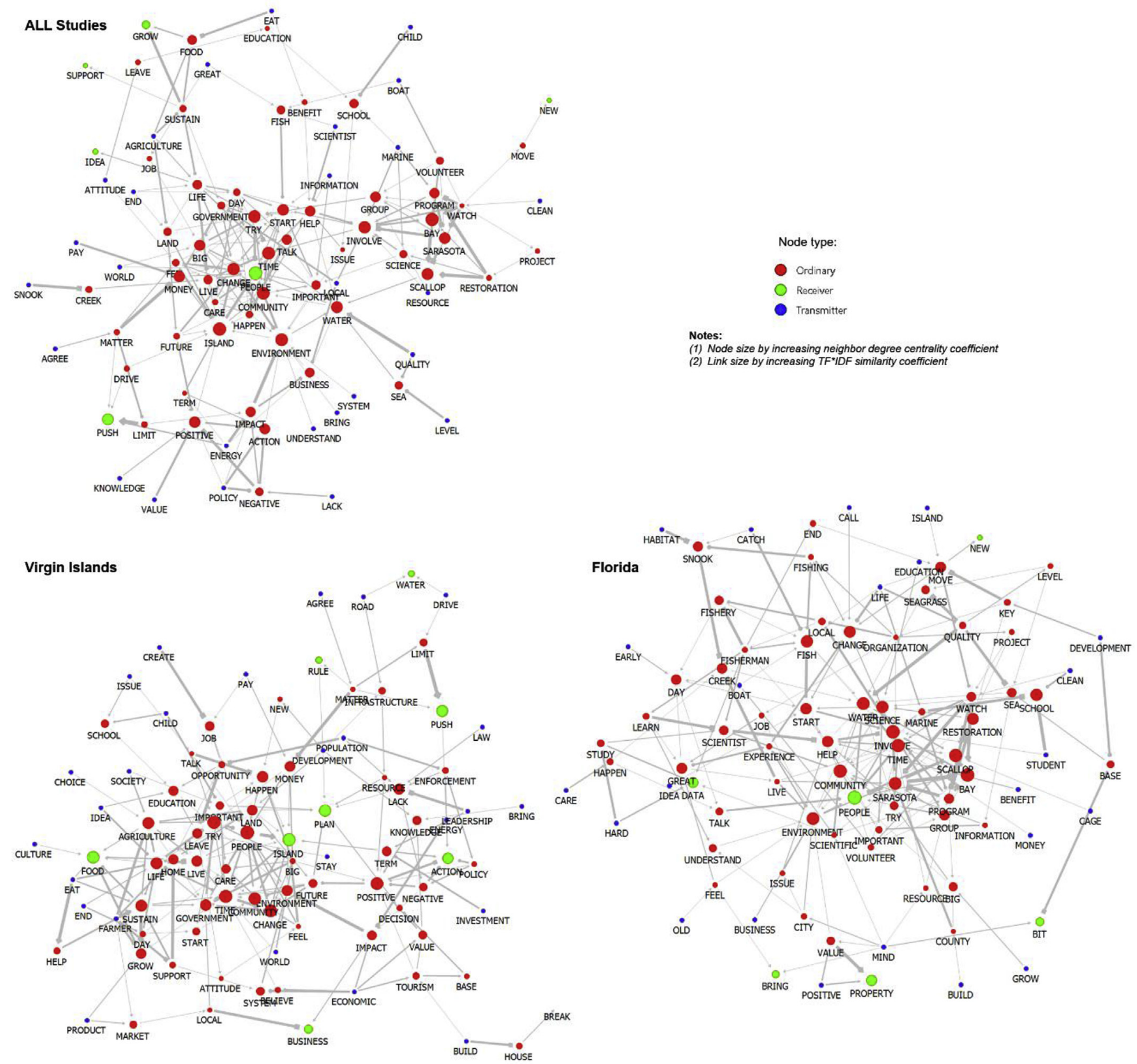

Fig. 9. Semantic knowledge networks (95% percentile) for the neighbor in-degree coefficients (across case studies). Node color differentiates node types, node sizes varying in-degree coefficients, while link width differentiates TF*IDF similarity weights. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

information centrality, $\chi^2=24.94$, $z=4.99$ $(p=0.0001)$; for the *local reachability density*, $\chi^2=27.07$, $z=5.20$ $(p=0.0001)$; for the *k-density*, $\chi^2=22.80$, $z=-4.78$ $(p=0.0001)$.

The ego-density coefficient shows that associated neighbor groups of semantic knowledge concepts tend to be denser in the VI, and more loosely connected in the FL case study. The Information centrality coefficient shows that transmission of knowledge as information flows is stronger in the VI in comparison with the FL case study. The local outlier factor matrix model coefficients (local reachability density and k-distance) show that knowledge is slightly more reachable within k-nearest neighbors in the VI case study, while the average distance of any knowledge node to its k-nearest neighbors is higher in the FL case study. Local reachability (LRD_k) and k-distance (D_k) are closely related, since,

$$LRD_k(x, y) = \max\{D_k(y), d(x, y)\} \tag{5}$$

for any two nodes, x and y in the network (Campos et al., 2016; Jackson, 2008). The box plots comparing the two case studies for the four identified coefficients are shown in Fig. 11.

3.2.2. Evaluating power law distributions in semantic networks

We assessed the presence of scale-free distributions both at the semantic knowledge network level, as well as the node attribute distribution level. As described in the methodology section, we used the Kolmogorov-Smirnov statistic to evaluate the presence of power-law relations (Clauset et al., 2009). Table 6 shows the results of the statistical evaluation. The α coefficient represents the estimated exponent of the power-law relationship. The x_{min} coefficient denotes the probability estimation of the model, i.e., $P(x \geq x_{min})$. The K-S coefficients provide

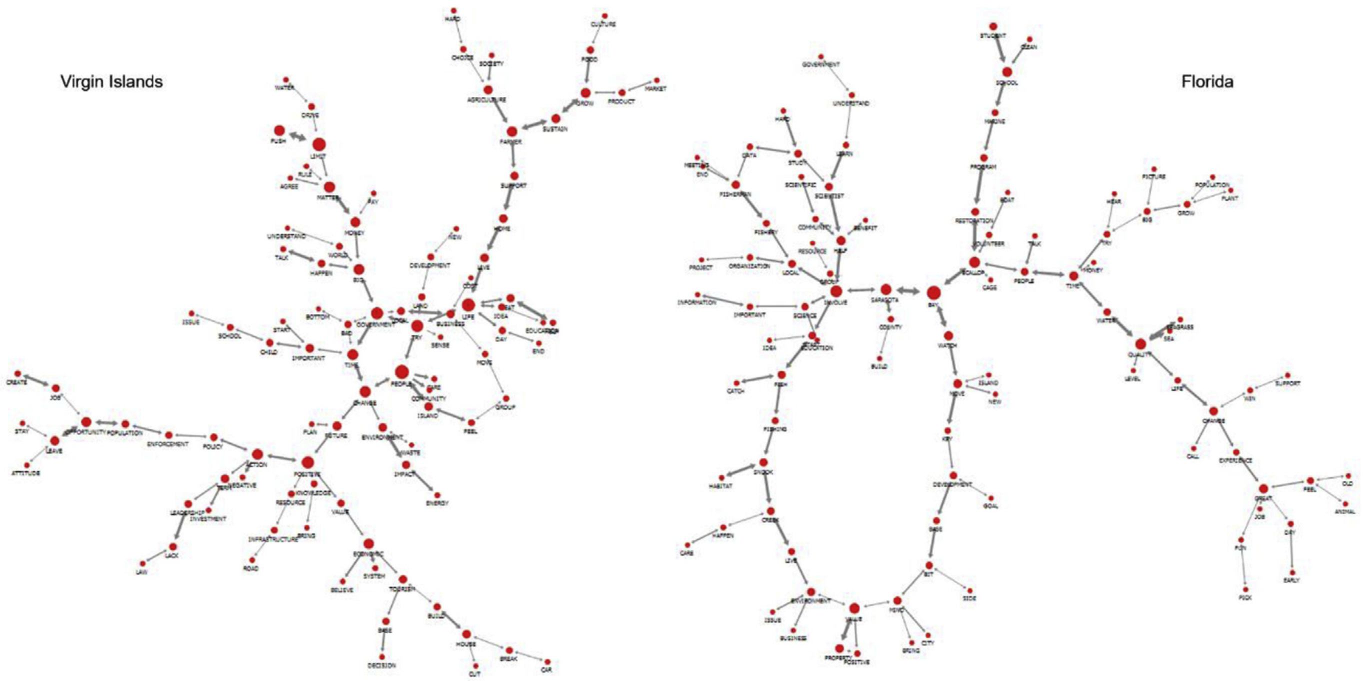

Fig. 10. Semantic network Path-finder (PFnet) algorithm results based on Euclidean distance dissimilarity for the two case studies. Node sizes denote higher average path dissimilarity. Link width denotes higher case frequency similarity in the network.

the values of the Kolgomorov-Smirnov statistic text. The p-values of the statistical evaluation with the exception of the 50% and 95% percentiles (reduced networks) in the combined case study networks are not statistically significant, not allow us to reject the null hypothesis that the network Jaccard similarity in-degree coefficient follows a power-law (scale free) distribution. In other words, the semantic knowledge influence increases exponentially as more we move from peripheral towards collectively shared knowledge concepts.

Visually, the presence of scale free distributions in our case study knowledge networks is also shown in the graphs of the following

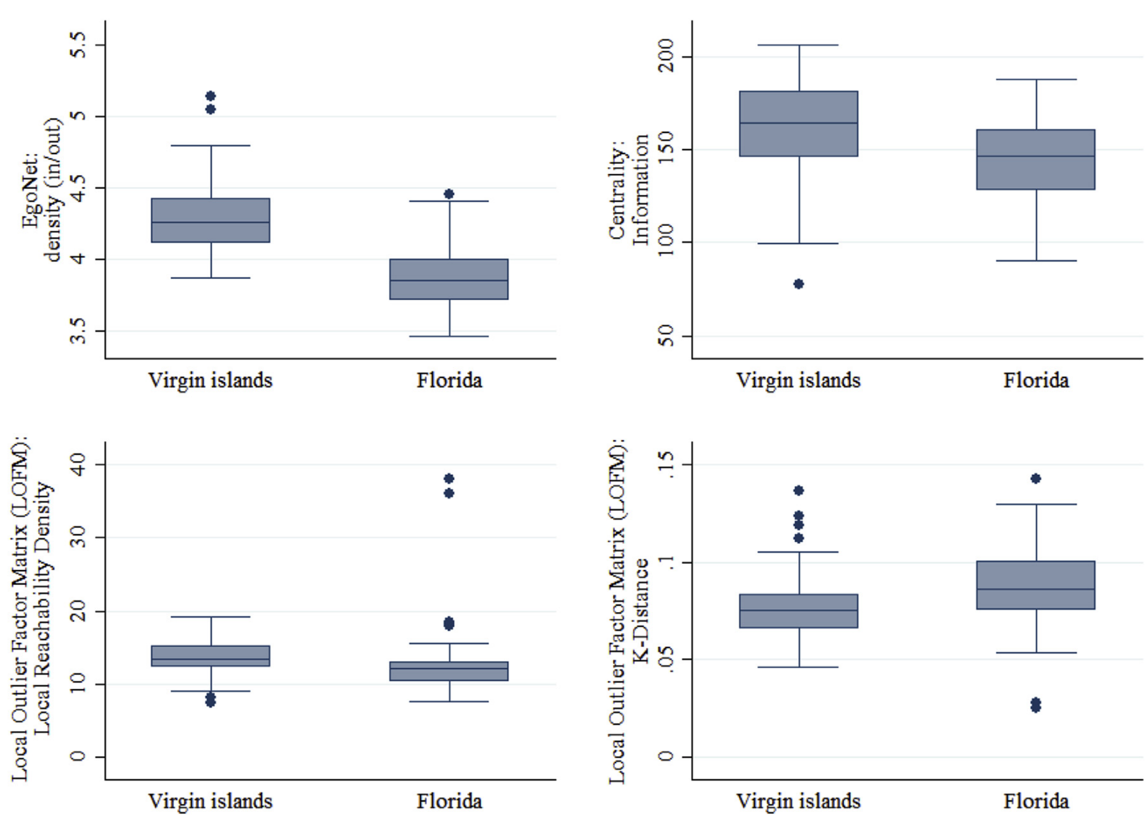

Fig. 11. Semantic knowledge network comparison of four selected graph-theoretic metrics: (a) Mean ego-density coefficients (neighbor-level); (b) Mean information centrality coefficients (centrality-level); (c) Mean local reachability density (local outlier factor matrix), and; (d) Mean k-distance (local outlier factor matrix).

Table 6
Power-law estimation statistics for each of the case study and network percentile in the study. The power-law network distribution is evaluated at the network matrix level (in-degree of Jaccard similarity coefficients) and at the node attribute level (TF*IDF).

Case Study and percentile		In-degree (network similarity)				TF*IDF (attribute vector)			
		α	x_{min}	K-S	p	α	x_{min}	K-S	p
Virgin Islands	5%	3.875	268.2	0.121	0.07	3.261	38.9	0.076	0.08
	50%	3.326	188.3	0.111	0.17	3.261	38.9	0.076	0.04
	95%	4.771	76.1	0.115	0.70	3.157	40.3	0.091	0.04
Florida	5%	4.103	257.9	0.103	0.32	2.922	40.1	0.08	0.04
	50%	3.532	190.8	0.129	0.09	2.922	40.1	0.08	0.01
	95%	2.517	32.5	0.134	0.06	4.283	102.1	0.088	0.59
All studies	5%	2.892	156.8	0.138	0.01	3.894	121.9	0.067	0.46
	50%	2.588	116.0	0.135	0.00	3.894	121.9	0.067	0.44
	95%	1.876	8.2	0.133	0.00	3.894	121.9	0.067	0.63

Fig. 12. The y-axis of the graphs shows the cumulative probability distribution function (CDF) of the in-degree Jaccard similarity in each of the case studies. The x-axis shows the raking of the 100 semantic network nodes (categories). Rank 1 is assigned to the node with the highest in-degree, and rank 100 to the node with the lowest one. The axes are plotted on a log-scale. The graphs show how few categories (marked as dots) have the highest in-degree similarity, while the majority of the nodes have a moderate or low in-degree similarity.

3.2.3. Evaluating the effect of semantic network structure on the formation of knowledge systems

The computed raw semantic network graph-theoretic metrics are divided into four major types: *neighbor* metrics (measuring neighbor link-based characteristics), *centrality* metrics (measuring node and network-based characteristics), *equivalence* metrics (measuring structural network equivalence characteristics), and *position* metrics (measuring role-based network characteristics). These graph-theoretic types of analytical groups represent some unobserved latent variables related to each of their described characteristics. We constructed latent composite scale factor models of these types, using a reliability analysis. Specifically, we evaluated the best subset of the metrics that maximize their Cronbach's Alpha reliability coefficient. Cronbach's Alpha estimation allow us to assess item (graph metrics) reliability towards a latent scale. We perform each Alpha estimation for each of the case studies, as well to the combined corpus narrative data. Consequently, four latent scale coefficients were computed for our network metrics data: *neighbor scale, centrality scale, equivalence scale,* and *position scale* coefficients. The alpha values ranged from 0.943 to 0.958 for the neighbor scale (21 items/metrics); 0.960 to 0.963 for the centrality scale (26 items/metrics); 0.751 to 0.922 for the equivalence scale (10 items/metrics); and nearing 1.0 for the position scale (2 items/metrics). The average inter-item correlation ranged between 0.31 and 0.54 across case studies and scales. The full reliability estimate statistics for the four composite network scales can be found in Appendix 2.

We fitted a *nonlinear exponential growth model* between the TF*IDF values measured from the latent semantic analysis of the narratives and the four composite semantic knowledge network scale indexes, for each one of the main network systemic characteristics (neighbor characteristics, network centrality, network equivalence, and network position). The symbolic form of the estimation is:

$$y = c + e^{\beta_0+\beta_1 x_1+\ldots+\beta_i x_i} \tag{6}$$

or

$$TF{*}IDF = e^{\beta_0+\beta_1 x_{neighbor}+\beta_2 x_{centrality}+\beta_3 x_{equivalence}+\beta_4 x_{position}} + \varepsilon \tag{7}$$

The results of the exponential growth nonlinear estimation are shown in the following Table 7. The top part of the table provides the model estimates, and the bottom part the parameter estimation of the coefficients. All case studies had excellent adjusted fit (R^2) was excellent: 0.9946 for the combined case studies semantic knowledge network; 0.9879 for the Virgin Islands semantic knowledge network, and 0.9956 for the Florida semantic knowledge network. The mean exponent (i.e., the predicted values for the linear exponent component):

$$\beta_0 + \beta_1 x_{neighbor} + \beta_2 x_{centrality} + \beta_1 x_{equivalence} + \beta_4 x_{position} \tag{8}$$

is around 4 (4.1 for VI, and 4.2 for FL), with standard deviation of 0.4. The exponent's values range from 3.6 (VI) to 5.8 (FL).

As can be seen from the parameter estimates in Table 7, the effects

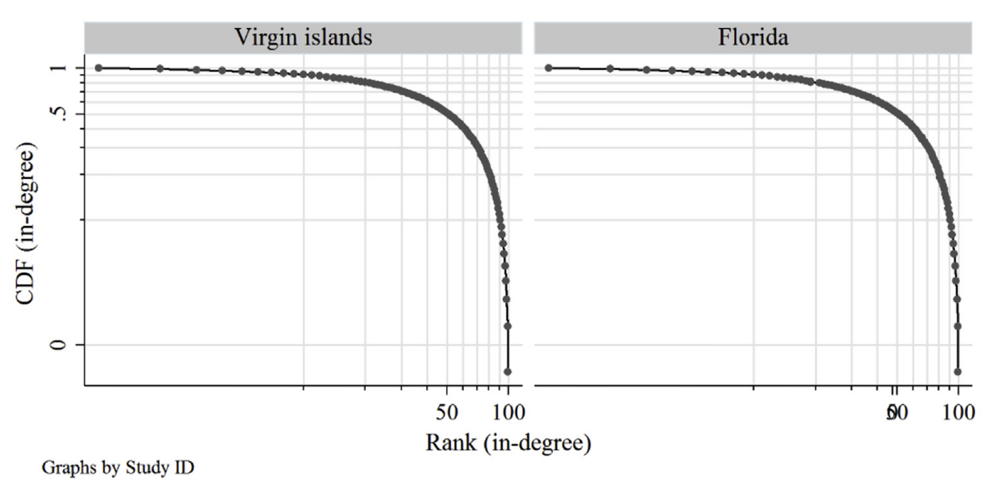

Fig. 12. Log-log graphs showcasing the presence of power-law relationships in the semantic knowledge networks in this study. The cumulative distribution function and value ranking is based on the neighbor's in-degree Jaccard similarity coefficients.

Table 7
Exponential growth nonlinear estimates between TF*IDF (dependent) and the four composite network structure indexes (neighbor, centrality, equivalence and position).

Dependent variable: TF*IDF; Model: Exponential growth

Case Study	Model estimation									
	N	SS_{model}	SS_{resid}	*df*	MS_{model}	MS_{resid}	R^2	$Adj.R^2$	*RMSE*	*Res.Dev*
ALL	100	2167929.9	11093.3	5, 95	433586.0	116.7	0.9949	0.9946	10.806	754.68
VI	100	545664.8	6335.0	5, 95	109132.9	66.7	0.9885	0.9879	8.166	698.66
FL	100	739204.5	3097.0	5, 95	147840.9	32.6	0.9958	0.9956	5.709	627.09

	Parameter estimates						
	Independent	Coeff. (β)	Std. Error	t-value	Pr > |t|	95% Conf. Interval	
ALL	Constant (β_0)	4.7729	0.0100	479.34	0.000	4.7530	4.7928
	Neighbor (β_1)	0.1711	0.0301	5.68	0.000	0.1113	0.2310
	Centrality (β_2)	0.3867	0.0304	12.71	0.000	0.3263	0.4472
	Equivalence (β_3)	−0.0709	0.0200	−3.54	0.001	−0.1106	−0.0312
	Position (β_4)	−0.0789	0.0123	−6.40	0.000	−0.1033	−0.0544
VI	Constant (β_0)	4.0797	0.0151	270.95	0.000	4.0498	4.1096
	Neighbor (β_1)	0.2651	0.0546	4.85	0.000	0.1567	0.3736
	Centrality (β_2)	0.2970	0.0491	6.06	0.000	0.1995	0.3946
	Equivalence (β_3)	−0.0363	0.0340	−1.07	0.288	−0.1038	0.0311
	Position (β_4)	−0.0598	0.0193	−3.10	0.003	−0.0980	−0.0214
FL	Constant (β_0)	4.2010	0.0969	433.68	0.000	4.1818	4.2202
	Neighbor (β_1)	0.1627	0.0339	4.81	0.000	0.0955	0.2299
	Centrality (β_2)	0.3553	0.0316	11.25	0.000	0.2926	0.4180
	Equivalence (β_3)	0.0711	0.0197	3.61	0.000	0.0320	0.1102
	Position (β_4)	−0.0651	0.0097	−6.73	0.000	−0.0843	−0.0459

of the neighbor coefficient increase TF*IDF by 26.5% in the Virgin Islands, versus only 16.3% in the Florida case study. The results are opposite for the centrality coefficient. Its effect is smaller in the VI (29.7%) than in the FL case study (35.5%). The equivalence coefficient is marginally negative in the VI, decreasing TF*IDF by 3.6% and marginally positive in FL, increasing TF*IDF by 7.1%. Finally, the position coefficient is marginally negative, decreasing TF*IDF in both case studies (by 6% and 6.5% in VI and FL respectively).

The predicted values of the exponential growth model estimation are shown in the following Fig. 13. In all case studies, the structure of the semantic network characteristics has exponentially increasing combined effects on the semantic importance of the narratives.

3.3. Network clustering classification

The following Fig. 14 below, presents a visualization of the major cluster group of contents classified through a *community graph plot algorithm* (CIT). The algorithm enabled the identification of seven broader cluster groups in the Virgin Islands case study, and eight cluster groups in the Florida case study. The algorithm maximizes cross-group semantic distances, and minimizes within-group semantic distances based on estimated cluster membership profile.

Heuristically, as can be seen from the graphs, the two case studies have a strong degree of overlap in four key areas: (a) knowledge related to economic considerations (socio-economics and business in both case studies); (b) educational knowledge (education/community outreach in

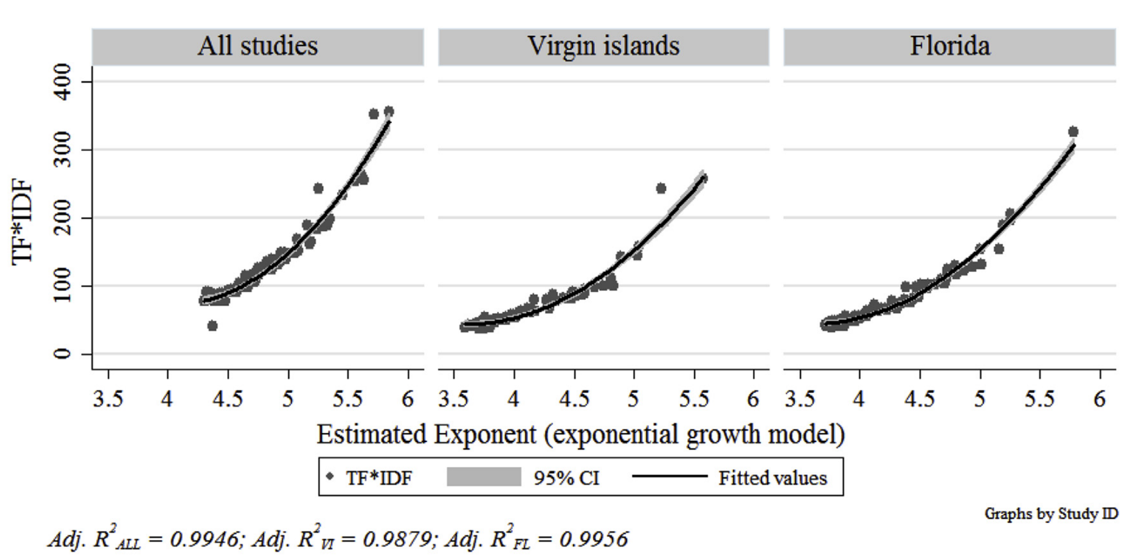

Fig. 13. Case study regression plots between TF*IDF values (y-axis) and the exponential growth model's predicted exponent of the four composite network characteristic variables (x-axis). The fitted lines and 95% confidence intervals are plotted against predicted values.

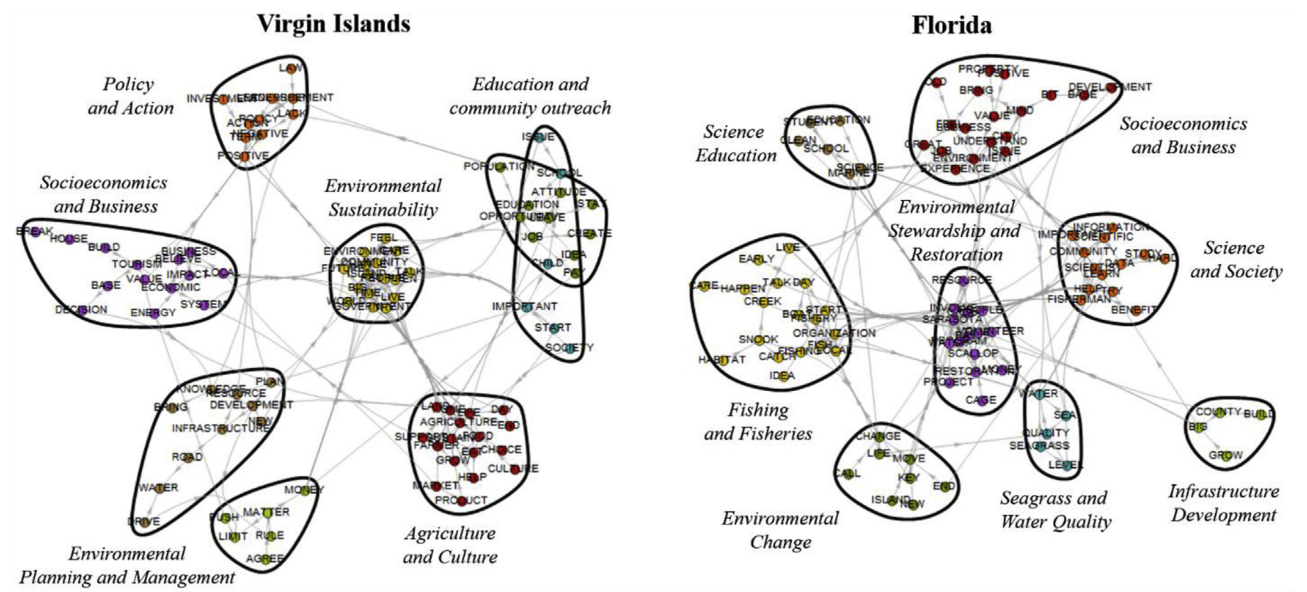

Fig. 14. Community graph plot clustering algorithm visualization of the two case studies: Virgin Islands (left subgraph) and Florida (right subgraph). The thick line encircling groups represent community graph clusters. Group cluster labels are based on content interpretation of semantic category memberships.

VI, vs. Science education in FL); (c) knowledge related to agriculture and fishing livelihoods (agriculture and culture in VI, vs. fishing and fisheries in FL), and; (d) local ecological knowledge related to sustainability and stewardship (central components in both graphs). On the other hand, in the Virgin Islands, the community participants focus more on knowledge narratives related to policy and action, as well as environmental planning and management, themes that are not emerging in the Florida case study. Florida participants focus on ecosystem and ecological aspects of marine and environmental management (including water quality), along with emphasizing the relationship between science and society. Both of these generative themes are not emerging in the Virgin Islands case study.

In addition to the network-level clustering of semantic knowledge, we also applied an alternative clustering classification, based on natural language processing. Specifically, we performed a factor analysis on the 100 extracted concepts using their co-occurrence frequency matrix by paragraph in the corpus document. In each of the case studies, we selected the ten factors with the highest eigenvalues. The factor scores were rotated using a varimax method. The eigenvalues range from 8.5 to 1.72 in the Virgin Islands case study, and from 8.1 to 1.69 in the Florida case study. The category with the most extensive coverage of the textual corpus (percentage of cases) in the Virgin Islands case study was *Social-Ecological Impacts* (33.1%), while the one for the Florida case study, was *Science and Community* (17.4%).

Following the factor analysis, a *co-occurrence (Jaccard) similarity analysis* was performed over the factor groupings (classes) across the two case studies. The following Fig. 15 shows a heatmap graph of the standardized (z-score) Jaccard similarity coefficients. We can see that the stronger similarity occurs between the *Science and Community* in the Florida case study, and the [*Environmental Policy & Management, Drivers of Environmental Change*] pair of factors in the Virgin Islands case study. Strong similarities also occur in general economic impacts in both case studies, and between *Environmental Education* (FL) and *Livelihoods and Opportunities* (VI). Overall, the total overlap between the two factor categories in terms of the semantic textual corpus is 60.1%. Compared to the 43.7% overlap observed in the original categories (see section 3.1), there is an additional 16.4% overlap that is due to the generalization of semantic knowledge categories.

4. Conclusions and discussion

In terms of the overarching goals and research questions of the study, the results provide a clear and concise description of the structure and characteristics of the environmental knowledge across participants (i.e., stakeholders, communities of practice). Similarly, through the network classification and groupings of semantic networks, our model results raise a series of important dimensions related to the institutional and social arrangements that are related to core environmental knowledge. The study results clearly provide a data-driven semantic network representation of both environmental knowledge structures at the socio-linguistic and cognitive level and the collective social level of discourse interactions. Finally, the social network analysis results in terms of centrality, density, neighborhood effects, structural hole equivalence (redundancy/efficiency), homophily, shortest-path analysis, and power-law analysis provide evidence of the strong relationship and association between the structure of self-organization in semantic inference and the characteristics of collective knowledge as a social-ecological system of interactions.

The categorization overlap analysis indicates that categories that appear in both case studies generally have significantly higher co-occurrence and TF*IDF values than the categories that uniquely appear in each case study. Thus, the central core of the semantic knowledge network is shared between the two studies, reflecting the extent to which environmental local knowledge represents a broader, global veridicality of social-ecological knowledge. Conversely, the remaining uniquely identified semantic categories in each case study (approximately 56% of the concepts) reflect discourse narratives that are either specific with respect to the locality, or the semantic discourse theme, or the group participant characteristics.

The application of the three alternative categorization schemes to the textual narrative corpus data allows us to examine some general contextual aspects of the collective knowledge discourse captured in our data. Specifically, it appears that knowledge involves both positive and negative aspects of social-ecological change, with the former group being more prominent than the second. Overall, we can see that in general a positive row balance characterizes the participants' environmental knowledge, reflecting a more optimistic outlook. We can also infer (from the RID classification results) that local environmental knowledge reflects by large a significant proportion of secondary processes. Such processes involve analytical, logical and abstract thinking,

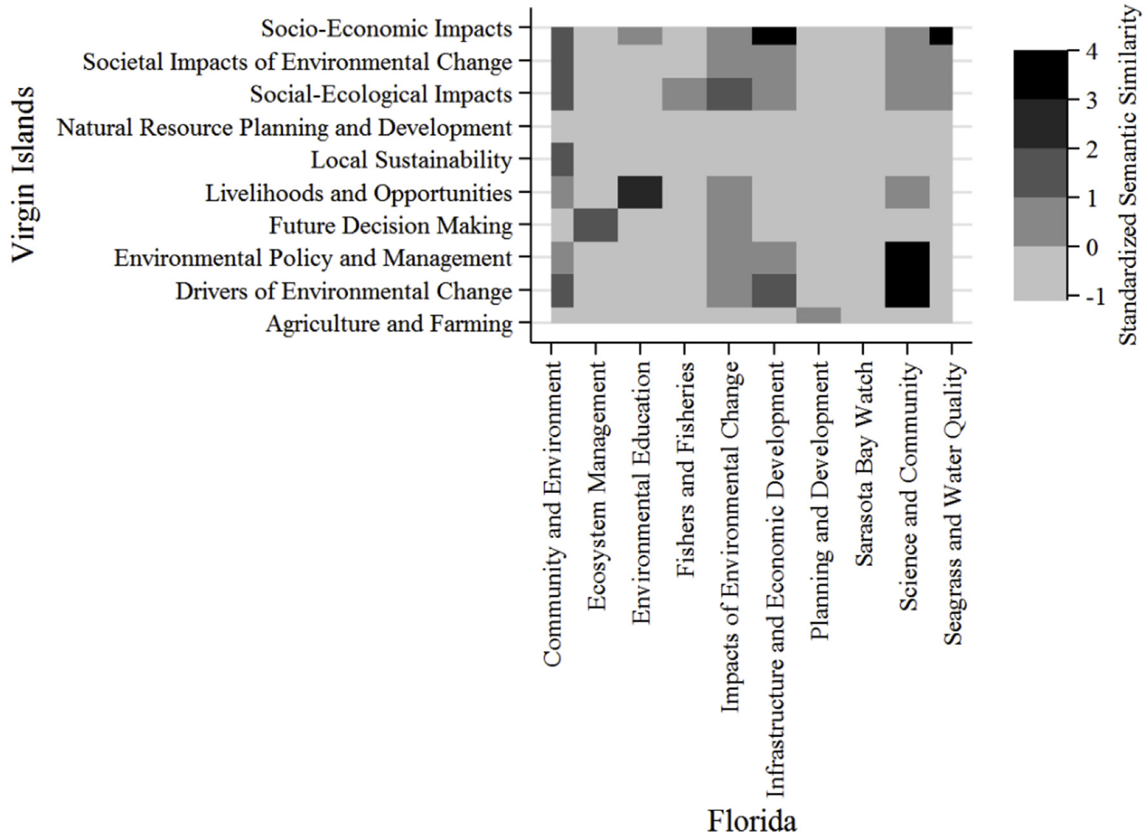

Fig. 15. Cross case-study associations based on standardized semantic similarity of co-occurrence patterns in the textual corpus.

rather than primary and emotional responses. While the latter do have a place in the composition of our collective knowledge structure, their role is not as prevalent as the secondary aspects of knowledge processing. These results also allow us to clearly understand why knowledge structures represents a more advanced and valuable system than simply processing information. The qualitative difference between information and knowledge rests by large on the ability of our communities and social groups to engage in secondary processing of raw cognitive or situational stimuli (e.g., emotional, primary).

In comparing the attribute characteristics of study participants with respect of their semantic categories in the categorization by attributes session, we identified how there exists a gender-based gap in knowledge compositional characteristics. Female participants differ in terms of their focus on subjective and procedural aspects of knowledge as constructs of their social realities, while male participants differ mostly by their focus on objective and situational outcomes. This distinction is generally present in both case studies, thus it is unlikely that it has much local cultural significant, and more related to the cognitive and mental attributes of male and female psychology in general. In that sense may represent an intrinsic characteristic of knowledge rather than one that is exogenously driven.

When we compare the age-based attributes of the participants with respect to their respective semantic categorization frequencies, the observed gap is of a different nature. It has to do mostly in terms of the qualitative origins of knowledge. On the one hand we have knowledge processes that are formed and driven by experience and situational perceptions of reality. These experiential and situational forces represent generally bottom-up, participatory mechanisms. On the other hand, we have knowledge processes that are informed and driven by deontic processes (i.e., what ought-to or what could be, rather than what-is). The latter knowledge processes in contrast, represent top-down, often policy or institutional rule-driven mechanisms. This is the case when comparing the Virgin Islands young adult differentiating categories (focusing mainly on career opportunities, job, school, etc.) with the Florida case study, where similar categories favor the opposite age group. More likely than not, these results provide support for the proposition that the Florida participants have a view of reality from outside experiential or situational settings.

The high degree of redundancy we found as part of the structural hole network analysis (approx. 27%) allows us to make inferences on the structural stability and robustness of the local environmental knowledge as represented by the narrative knowledge semantic networks. The higher the redundancy, the more paths exists connecting any two random nodes in the network, thus making difficult to disconnect parts of the knowledge network. Conversely, the higher the redundancy of a network, the lower it's efficiency. The structure of the semantic networks analyzed for this study, appears to maintain an approximate relationship of 2.5:1 between efficiency and redundancy, i.e., a balance between marginal network connectivity and full network connectivity. This is an additional finding re-affirming our network results related to the presence of power-law distribution in the network structure. Combined together, they indicate an equilibrium-state of the semantic knowledge network, and therefore increases our inferential ability about the structure of the social-ecological knowledge represented in our data.

The results obtained from the analysis of homophily for our semantic knowledge networks allows us to support the proposition that local social-ecological knowledge structures become more cohesive in value and pattern as we move from the periphery towards the center of the knowledge network. In other words, the more central knowledge categories are, the more they are similar to their connected neighbors and vice versa. This proposition also supports the idea of an increasing homogeneity in semantic knowledge character as we move from the periphery to the center. Knowledge becomes more and more homogenous as it assimilates and spreads within the local community (i.e., becomes more central). The opposite is true with regards to knowledge heterogeneity. Heterogeneous knowledge structures are found more towards the network's periphery. This is important especially when new knowledge and ideas enter the knowledge structures. They first represent heterogeneous entities. As they move up the network isomorphic hierarchy (become more central) they trade heterogeneity with homogeneity by acquiring values more closer to their connected neighbors.

The geodesic distance and reachability results of the shortest path network analysis of the data shows us that, at least closer to the core of the semantic knowledge, the percentage of nodes that are outwards-reachable is significantly higher than the percentage of nodes that are inwards-reachable. In other words, starting from the center node we can reach more nodes as opposed from starting from the periphery. In terms of semantic knowledge structure, these results provide support of the proposition that knowledge diffusion patterns (how the central knowledge is diffused in the local community) are stronger than knowledge acquisition patterns (how new knowledge is acquired and assimilated in the community). New knowledge patterns can reach and connect with smaller amount of existing knowledge than existing knowledge affecting new ideas. This may be perhaps another, quantitative way to show that new and innovative knowledge needs to reach a certain threshold in order to assimilate fully (in the case of our analysis, connecting with more than 12% of existing knowledge structures).

The multidimensional scaling for the focus groups reveal that certain focus groups have coordinates significantly deviating from the mean. For example, in the Virgin Islands case study groups representing more localized community groups (such as the non-government NRM group or the farmer's coop) appear to be placed in the periphery of the multidimensional graph. In the Florida case study the second session of both the third and fourth focus groups also appear to have peripheral values, at least on the second dimension. These results allow us to infer that the discourse narratives captured by the semantic analysis have substantially diverging knowledge content, and thus contribute more to the heterogeneity of knowledge.

In terms of the multidimensional scaling analysis of the occupational role of the participant's groups, one important result warrants discussion. By comparing the same group position across the two case studies, we can see that while in the Virgin Islands the role of scientists and science organizations is semantically closer to those of community/business groups, in the Florida case study, scientists and science organization roles are more closely aligned with government/agency views and knowledge discourse. These diverging comparative results are perhaps indicative of two distinct models of the role of science or scientists as bidirectional translators of knowledge. The first is characteristic of a bottom-up approach where scientists and scientific organizations work closely and more tightly with the local community and locally-embedded organizations. The second, provides evidence of a top-down approach where scientists serve in support of government policy interventions, or regulatory agency approaches to environmental and natural resource management. While the efficiency of these two generative approaches is not evaluated as part of this study, it is an important distinction that has both theoretical and empirical significance in the relevant scientific literature.

The results obtained by fitting a nonlinear exponential growth regression of the four network composite scale metric categories (neighbor, centrality, equivalence and position) to the semantic importance of node categories (TF*IDF) allowed us to assess both the combined and the relative effects of each of the four coefficients in predicting the structure of the semantic knowledge networks. Overall, more than 95% of the TF*IDF in the semantic corpus can be predicted by the four network coefficients. In other words, the graph-theoretic structure of the semantic networks represents can predict the importance of knowledge categories within a given thematic narrative. Of these four coefficients, neighbor characteristics (i.e., the structural connectivity patterns of knowledge), and centrality (i.e., the cohesiveness and structure of knowledge categories) are the most important predictors. In the Virgin Islands case study, the most important predictor is connectivity (neighbor), but in the Florida case study, the most important predictor is cohesion (centrality). The latest results are also indicative of the two perceptual models of knowledge, the bottom-up (in the VI) and the top-down (in FL).

While both of the case studies allow for a mix of both centrality and neighbor influences to strongly influence knowledge structures, the particulars of each mix allow us to make certain inferences. The higher importance of how participants connect and associate their knowledge structure within their knowledge networks provides comparatively stronger evidence of a bottom-up knowledge system. It is more likely than not, that new environmental knowledge generation in the Virgin Islands flows by new knowledge contributes to the overall connectivity of a network. This type of relationship characteristic in the structure of network is particularly present in the case weak ties in small-world networks (Omidi and Masoudi-Nejad, 2010; Watts, 1999). Similarly, the higher importance on how centrally cohesive are knowledge concepts in a network provides evidence of a top-down knowledge system for the Florida case study. This type of network characteristics are more prevalent in centralized and tightly knit network structures (Borgatti, 2005). These two distinct modes of knowledge structural dynamics, can be reflective of the focal role of knowledge as a social construct of reality. The first one favors knowledge acquisition patterns more than knowledge diffusion patterns (e.g., Virgin Islands), essentially implying that knowledge generation emerges closer to the fringe of the knowledge networks and flows toward the center. The second, reverses these trends: favors more knowledge diffusion flows than knowledge acquisition, implying knowledge generation emergence near the core of the knowledge structure, flowing outwards.

The results of both the community graph plot clustering algorithm classification, and the semantic factor analysis classification on the original textual corpus provide us with a broader, bird-eye understanding of the grouped thematic associations present in the two case study knowledge discourses. While both of the methodologies demonstrate a strong power to identify and discriminate groups of knowledge concepts, they also help us understand how thematically-driven knowledge groups associate with each other. There is a significant percentage of heuristic overlap in grouping classifications between the two case studies, especially given the fact that the discourse narrative topics are thematically independent from each other at the time of data collection. This, in turn, provides a strong indicator and argument in support of the theoretical proposition that we share a core of our environmental knowledge across communities and cultures. We argue that more likely than not, this part represents a more “global” part of environmental knowledge, one that is part of our general socio-cognitive understanding of environmental change. It reflects our general cross-societal and cross-cultural knowledge related to global environmental change, and it is likely to be more influenced by exogenous or broader factors that determine our social understanding of reality within a globalized world. No matter what our individual, not shared components of our local environmental knowledge are, there will always be a fertile core of knowledge where one can plant the seed of true global environmental stewardship, the one that connects communities of purpose, sense of place and communities of practice with respect to social-ecological systems.

Acknowledgments

Partial funding for this research was provided from the US National Science Foundation, Office of Integrated Activities, through the Virgin Islands Experimental Program for Stimulating Competitive Research (VI-EPSCoR), award no.1355437: “Mare Nostrum Caribbean: Stewardship through Strategic Research and Workforce Development”.

The authors wish to acknowledge the contribution of a number of colleagues, including Dr. Michael Crosby (Mote Marine Laboratory, Florida, USA) for hosting and organizing the Florida workshop; Dr.

Hiroshi Miki (RIHN institute) for earlier and latter comments on the analysis trajectories and draft.

We also want to gratefully acknowledge the contributions of the study participants, representing professionals, agency and government officials, community members, volunteers and non-government organizations for their participation in the workshop and focus group exercises.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2018.08.026.

References

Ai, D., Zheng, Y., Zhang, D., 2010. Automatic text summarization based on latent semantic indexing. Artif. Life Robot. 15 (1), 25–29. https://doi.org/10.1007/s10015-010-0759-x.
Alexandridis, K., Maru, Y., 2012. Collapse and reorganization patterns of social knowledge representation in evolving semantic networks. Inf. Sci. 200 (1), 1–21. https://doi.org/10.1016/j.ins.2012.02.053.
Bader, S., Hitzler, P., Hölldobler, S., 2008. Connectionist model generation: a first-order approach. Neurocomputing 71 (13–15), 2420–2432.
Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512. https://doi.org/10.1126/science.286.5439.509.
Basurto, X., Gelcich, S., Ostrom, E., 2013. The social–ecological system framework as a knowledge classificatory system for benthic small-scale fisheries. Global Environ. Change 23 (6), 1366–1380. https://doi.org/10.1016/j.gloenvcha.2013.08.001.
Becker, C.D., Ghimire, K., 2003. Synergy between traditional ecological knowledge and conservation science supports forest preservation in Ecuador. Ecol. Soc. 8 (1).
Berger, P.L., Luckmann, T., 1967. The Social Construction of Reality: a Treatise in the Sociology of Knowledge, 1990 ed. Anchor Books, Random House, New York.
Berkes, F., Turner, N., 2006. Knowledge, learning and the evolution of conservation practice for social-ecological system resilience. Hum. Ecol. 34 (4), 479–494. https://doi.org/10.1007/s10745-006-9008-2.
Borgatti, S.P., 2005. Centrality and network flow. Soc. Network. 27 (1), 55–71.
Breese, E.L., Hillis, A.E., 2004. Auditory comprehension: is multiple choice really good enough? Brain Lang. 89 (1), 3–8.
Brendel, R., Krawczyk, H., 2010. Extended generalized blockmodeling for compound communities and external actors. In: Abraham, A., Hassanien, A.-E., Snasel, V. (Eds.), Computational Social Network Analysis: Trends, Tools and Research Advances. Springer, London, pp. 371–385.
Brewka, G., Benferhat, S., Le Berre, D., 2004. Qualitative choice logic. Artif. Intell. 157 (1–2), 203–237.
Campos, G.O., Zimek, A., Sander, J., Campello, R.J.G.B., Micenková, B., Schubert, E., Houle, M.E., 2016. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Min. Knowl. Discov. 1–37. https://doi.org/10.1007/s10618-015-0444-8.
Chater, N., Tenenbaum, J.B., Yuille, A., 2006. Probabilistic models of cognition: conceptual foundations. Trends Cognit. Sci. 10 (7), 287–291.
Chia, G.L., Ong, P.-H., 2006. Isomorphic factorization of the complete graph into Cayley digraphs. J. Graph Theor. 52 (3), 243–256.
Chiu, C., Gruber, S.A., Simpson, N., Yurgelun-Todd, D.A., 2003. Indirect semantic priming in schizophrenic patients. Schizophr. Res. 60 (1), 129.
Cinner, J.E., McClanahan, T.R., Wamukota, A., 2010. Differences in livelihoods, socioeconomic characteristics, and knowledge about the sea between Fishers and non-Fishers living near and far from marine parks on the Kenyan coast. Mar. Pol. 34 (1), 22–28.
Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703. https://doi.org/10.1137/070710111.
Condray, R., Siegle, G.J., Cohen, J.D., van Kammen, D.P., Steinhauer, S.R., 2003. Automatic activation of the semantic network in schizophrenia: evidence from event-related brain potentials. Biol. Psychiatr. 54 (11), 1134–1148.
Cowie, G.M., Borrett, S.R., 2005. Institutional perspectives on participation and information in water management. Environ. Model. Software 20 (4), 469–483. https://doi.org/10.1016/j.envsoft.2004.02.006.
Crestani, F., 1997. Application of spreading activation techniques in information retrieval. Artif. Intell. Rev. 11 (6), 453–482.
de Vos, M.G., Janssen, P.H.M., Kok, M.T.J., Frantzi, S., Dellas, E., Pattberg, P., ... Biermann, F., 2013. Formalizing knowledge on international environmental regimes: a first step towards integrating political science in integrated assessments of global environmental change. Environ. Model. Software 44, 101–112. https://doi.org/10.1016/j.envsoft.2012.08.004.
Dib, B., Kalloubi, F., Nfaoui, E.H., Boulaalam, A., 2018. Semantic-based followee recommendations on twitter network. Procedia Computer Science 127, 505–510. https://doi.org/10.1016/j.procs.2018.01.149.
Drieger, P., 2013. Semantic network analysis as a method for visual text analytics. Procedia - Social and Behavioral Sciences 79, 4–17. https://doi.org/10.1016/j.sbspro.2013.05.053.
Dutta, R., Morshed, A., Aryal, J., D'Este, C., Das, A., 2014. Development of an intelligent environmental knowledge system for sustainable agricultural decision support. Environ. Model. Software 52, 264–272. https://doi.org/10.1016/j.envsoft.2013.10.004.
Enfield, N.J., 2009. Relationship thinking and human pragmatics. J. Pragmat. 41 (1), 60–78.
Fritzsche, M., 2003. The origin of brain asymmetry and its psychotic reversal. Med. Hypotheses 60 (4), 468–480.
Fronzetti Colladon, A., 2018. The semantic brand score. J. Bus. Res. 88, 150–160. https://doi.org/10.1016/j.jbusres.2018.03.026.
Ghorbani, A., Bravo, G., Frey, U., Theesfeld, I., 2017. Self-organization in the commons: an empirically-tested model. Environ. Model. Software 96, 30–45. https://doi.org/10.1016/j.envsoft.2017.06.039.
Giovagnoli, A.R., Villani, F., Bell, B., Erbetta, A., Avanzini, G., 2009. The chicken with four legs: a case of semantic amnesia and cryptogenic epilepsy. Epilepsy Behav. 14 (1), 261–268.
Graham, K.S., Simons, J.S., Pratt, K.H., Patterson, K., Hodges, J.R., 2000. Insights from semantic dementia on the relationship between episodic and semantic memory. Neuropsychologia 38 (3), 313–324.
Grant, S., Berkes, F., 2007. Fisher knowledge as expert system: a case from the longline fishery of Grenada, the Eastern Caribbean. Fish. Res. 84 (2), 162–170.
Greaves, M., Mika, P., 2008. Semantic web and web 2.0. In: Web Semantics: Science, Services and Agents on the World Wide Web, . Retrieved from https://ejournal.csiro.au/cgi-bin/sciserv.pl?collection=journals&journal=15708268&issue=v06i0001.
Haenni, R., 2005. Unifying logical and probabilistic reasoning. In: Lecture Notes in Computer Science, vol. 3571. Springer-Verlag GmbH, pp. 788–799.
Hasan, M.A., Zaki, M.J., 2011. A survey of link prediction in social networks. In: Aggarwal, C.C. (Ed.), Social Network Data Analytics (pp. 243-275). Springer US.
Haybron, D.M., 2000. The causal and explanatory role of information stored in connectionist networks. Minds Mach. 10 (3), 361–380.
Hu, W., Wang, H., Peng, C., Liang, H., Du, B., 2017. An event detection method for social networks based on link prediction. Inf. Syst. 71, 16–26. https://doi.org/10.1016/j.is.2017.06.003.
Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press, Princeton, NJ.
Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106 (4), 620–630.
Kazui, H., Hashimoto, M., Hirono, N., Mori, E., 2003. Nature of personal semantic memory: evidence from Alzheimer's disease. Neuropsychologia 41 (8), 981–988.
Kiptot, E., 2007. Eliciting indigenous knowledge on tree fodder among Maasai pastoralists via a multi-method sequencing approach. Agric. Hum. Val. 24 (2), 231–243.
Knoke, D., Yang, S., 2008. Social Network Analysis, second ed. Sage Publications, Los Angeles.
Kristjanson, P., Reid, R.S., Dickson, N., Clark, W.C., Romney, D., Puskur, R., Grace, D., 2009. Linking international agricultural research knowledge with action for sustainable development. Proc. Natl. Acad. Sci. Unit. States Am. 106 (13), 5047–5052. https://doi.org/10.1073/pnas.0807414106.
Laver, M., Garry, J., 2000. Estimating policy positions from political texts. Am. J. Polit. Sci. 44 (3), 619–634. https://doi.org/10.2307/2669268.
Levy, B.J., Anderson, M.C., 2008. Individual differences in the suppression of unwanted memories: the executive deficit hypothesis. Acta Psychol. 127 (3), 623–635.
Leydesdorff, L., 2008. On the normalization and visualization of author co-citation data: salton's Cosine versus the Jaccard index. J. Am. Soc. Inf. Sci. Technol. 59 (1), 77–85. https://doi.org/10.1002/asi.20732.
Li, B., Wang, K., Zhang, X.-x., 2012. Semantic computation of medical image annotation based on semantic tree. In: In: Qian, Z., Cao, L., Su, W., Wang, T., Yang, H. (Eds.), Recent Advances in Computer Science and Information Engineering, vol. 129. Springer Berlin Heidelberg, pp. 37–44.
Li, L., Zhou, C., He, J., Wang, J., Li, X., Wu, X., 2017. Collective semantic behavior extraction in social networks. Journal of Computational Science. https://doi.org/10.1016/j.jocs.2017.11.003.
Loughran, T.I.M., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Finance 66 (1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.
Martindale, C., 1975. Romantic Progression: the Psychology of Literary History. Hemisphere Pub. Corp, Washington, New York distributed solely by Halsted Press).
Martindale, C., 1981. Cognition and Consciousness. Dorsey Press, Homewood, Ill.
Martindale, C., 1986. Aesthetic evolution. Poetics 15, 439–473.
Martindale, C., 1999. Biological bases of creativity. In: Sternberg, R.J. (Ed.), Handbook of Creativity (pp. 137-152). Cambridge University Press, Cambridge, MA.
Martínez-García, M., Valls, A., Moreno, A., Aldea, A., 2018. A semantic multi-criteria approach to evaluate different types of energy generation technologies. Environ. Model. Software. https://doi.org/10.1016/j.envsoft.2018.04.003.
Nepomniaschaya, A., 2006. Associative parallel algorithm for dynamic update of a minimum spanning tree after addition of a new node to a graph. Cybern. Syst. Anal. 42 (1), 18–27.
Newman, M.E.J., 2001. Clustering and Preferential Attachment in Growing Networks. Santa Fe Institute, pp. 13.
Oliver, D.M., Fish, R.D., Winter, M., Hodgson, C.J., Heathwaite, A.L., Chadwick, D.R., 2012. Valuing local knowledge as a source of expert data: farmer engagement and the design of decision support systems. Environ. Model. Software 36, 76–85. https://doi.org/10.1016/j.envsoft.2011.09.013.
Omidi, S., Masoudi-Nejad, A., 2010. Network evolution: theory and mechanisms. In: Abraham, A., Hassanien, A.-E., Snasel, V. (Eds.), Computational Social Network Analysis: Trends, Tools and Research Advances (pp. 191-240). Springer, London.
Paulsen, J.S., Romero, R., Chan, A., Davis, A.V., Heaton, R.K., Jeste, D.V., 1996. Impairment of the semantic network in schizophrenia. Psychiatr. Res. 63 (2–3), 109–121.

Piantadosi, S.T., Tily, H., Gibson, E., 2011. Word lengths are optimized for efficient communication. Proc. Natl. Acad. Sci. Unit. States Am. 108 (9), 3526–3529. https://doi.org/10.1073/pnas.1012551108.
Pinkas, G., 1995. Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge. Artif. Intell. 77 (2), 203–247.
Polhill, J.G., Filatova, T., Schlüter, M., Voinov, A., 2016. Modelling systemic change in coupled socio-environmental systems. Environ. Model. Software 75, 318–332. https://doi.org/10.1016/j.envsoft.2015.10.017.
Ravasz, E., Barabasi, A., 2003. Hierarchical organization in complex networks. Phys. Rev. 67 (026112), 7.
Reagan, A.J., Danforth, C.M., Tivnan, B., Williams, J.R., Dodds, P.S., 2017. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. EPJ Data Science 6 (1), 28. https://doi.org/10.1140/epjds/s13688-017-0121-9.
Real, R., Vargas, J.M., 1996. The probabilistic basis of jaccard's index of similarity. Syst. Biol. 45 (3), 380–385.
Redington, M., Chater, N., 1997. Probabilistic and distributional approaches to language acquisition. Trends Cognit. Sci. 1 (7), 273–281.
Reichardt, J., White, D.R., 2007. Role models for complex networks. The European Physical Journal B 60 (2), 217–224.
Rivera, S.J., Minsker, B.S., Work, D.B., Roth, D., 2014. A text mining framework for advancing sustainability indicators. Environ. Model. Software 62, 128–138. https://doi.org/10.1016/j.envsoft.2014.08.016.
Robertson, E.K., Köhler, S., 2007. Insights from child development on the relationship between episodic and semantic memory. Neuropsychologia 45 (14), 3178–3189.
Rogers, S.L., Friedman, R.B., 2008. The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia. Neuropsychologia 46 (1), 12–21.
Rott, H., 2001. Change, Choice and Inference: a Study of Belief Revision and Nonmonotonic Reasoning. Clarendon, Oxford.
Sailer, L.D., 1978. Structural equivalence: meaning and definition, computation and application. Soc. Network. 1, 73–90.
Sato, T., 2014. Integrated local environmental knowledge supporting adaptive governance of local communities. In: Alvares, C. (Ed.), Multicultural Knowledge and the University (pp. 268-273). Multiversity India, Mapusa, India.
Schvaneveldt, R.W., 1990. Pathfinder Associative Networks: Studies in Knowledge Organization. Ablex Pub. Corp, Norwood, N.J.
Scott, J., 2000. In: Social Network Analysis: a Handbook, second ed. SAGE Publications, London; Thousands Oaks, Calif.
SivaKumar, A.P., Premchand, P., Govardhan, A., 2011. Application of latent semantic indexing for Hindi-English CLIR irrespective of context similarity. In: In: Wyld, D.C., Wozniak, M., Chaki, N., Meghanathan, N., Nagamalai, D. (Eds.), Trends in Network and Communications, vol. 197. Springer Berlin Heidelberg, pp. 711–720.
Smith, S.M., Ward, T.B., Finke, R.A., 1995. The Creative Cognition Approach. MIT Press.
Stone, P.J., 1997. Thematic text analysis: new agendas for analyzing text content. In: Roberts, C. (Ed.), Text Analysis for the Social Sciences. Lawrence Erlbaum Associates Publishers.
Tàbara, J.D., Chabay, I., 2013. Coupling Human Information and Knowledge Systems with social–ecological systems change: reframing research, education, and policy for sustainability. Environ. Sci. Pol. 28 (0), 71–81. https://doi.org/10.1016/j.envsci.2012.11.005.
Takase, S., Okazaki, N., Inui, K., 2016. Modeling semantic compositionality of relational patterns. Eng. Appl. Artif. Intell. 50, 256–264. https://doi.org/10.1016/j.engappai.2016.01.027.
Tausczik, Y.R., Pennebaker, J.W., 2009. The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol. 29 (1), 24–54. https://doi.org/10.1177/0261927x09351676.
Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. Environ. Model. Software 24 (5), 577–587. https://doi.org/10.1016/j.envsoft.2008.09.009.
Voorbraak, F., 2004. A nonmonotonic observation logic. Artif. Intell. 157 (1–2), 281–302.
Watkins, D.S., 2004. Fundamental of Matrix Computations, second ed. John Wiley & Sons, New York.
Watts, D.J., 1999. Small Worlds: the Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton, N.J.
Webb, A.W., 2013. Community Perspectives on Sustainability and Resilience in a Social-ecological Paradigm. Masters of Science MSc. University of the Virgin Islands, St. Thomas, US Virgin Islands.
Wesche, S., Armitage, D.R., 2010. From the inside out: a multi-scale analysis of adaptive capacity in a northern community and the governance implications. In: Armitage, D., Plummer, R. (Eds.), Adaptive Capacity and Environmental Governance (pp. 107-132). Springer Berlin Heidelberg.
White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. Am. J. Sociol. 81 (4), 730–780.
Williams, M., 2009. Social objects, causality and contingent realism. J. Theor. Soc. Behav. 39 (1), 1–18.
WordStat, 2016. (Version 7.1): Provalis Research.
Xagi, M., Guerbas, A., Kianmehr, K., Karampelas, P., Ridley, M., Alhajj, R., Rokne, J., 2010. Employing social network construction and analysis in web structure optimization. In: Memon, N., Alhajj, R. (Eds.), From Sociology to Computing in Social Networks (pp. 13-34). Springer Vienna.
Xu, Z., Tresp, V., Rettinger, A., Kersting, K., 2010. Social network mining with nonparametric relational models. In: In: Giles, L., Smith, M., Yen, J., Zhang, H. (Eds.), Advances in Social Network Mining and Analysis, vol. 5498. Springer Berlin/Heidelberg, pp. 77–96.
Zhang, M., 2010. Social network analysis: history, concepts, and research. In: Furht, B. (Ed.), Handbook of Social Network Technologies and Applications (pp. 3-21). Springer US.