A shared understanding between humans and autonomous cyber collectives

Josef Schaff,
Naval Air Systems Command

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Situational Awareness

- May have smart sensors, System-of-Systems(SoS).
  - May be distributed SoS over large areas.
  - May have thousands of smart entities (Battlespace).
  - May produce large volumes of data.
- Need to share information with humans.
  - Big Data problem = need to semantically integrate the voluminous data.
  - Aggregate data needs to be summarized to relevant details, in an actionable (i.e. short) timeframe.
- Need to create: sparse yet relevant data stream for shared “world model”.

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Problems in Achieving Data Rendering

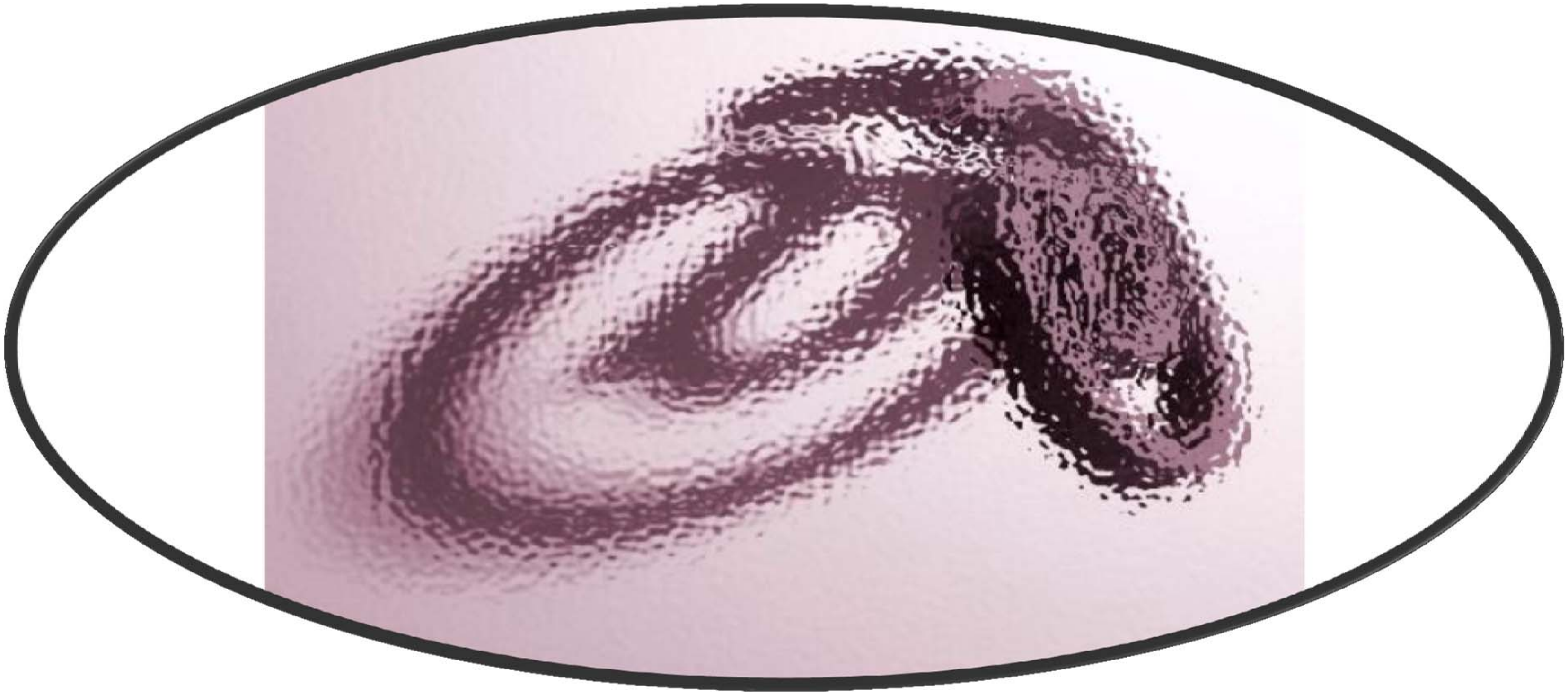
- Many methods are computationally intensive.
- Want to render Big Data to sparse data without losing relevant information.
  - Nature example: Stigmergy = collective behaviors.
  - How to do this in our synthetic autonomy?
- Leverage off topology for distributed systems.
  - Map topological relationships, each sensor's "perspective" can overlap to derive permutation of shared information = renders to sparse info.

To What and Where Do We Look?

- Big Data starts out as a “high dimensional” problem.
  - All or most information is there, but too complex or time-consuming to decipher.
  - One solution = reduce dimensionality from higher to lower, preserving relevant information.
- Look at methods to reduce dimensionality from theoretical mathematics, physics, to stigmergy and multivariate analysis.
  - Most methods use a variety of tools from topology.

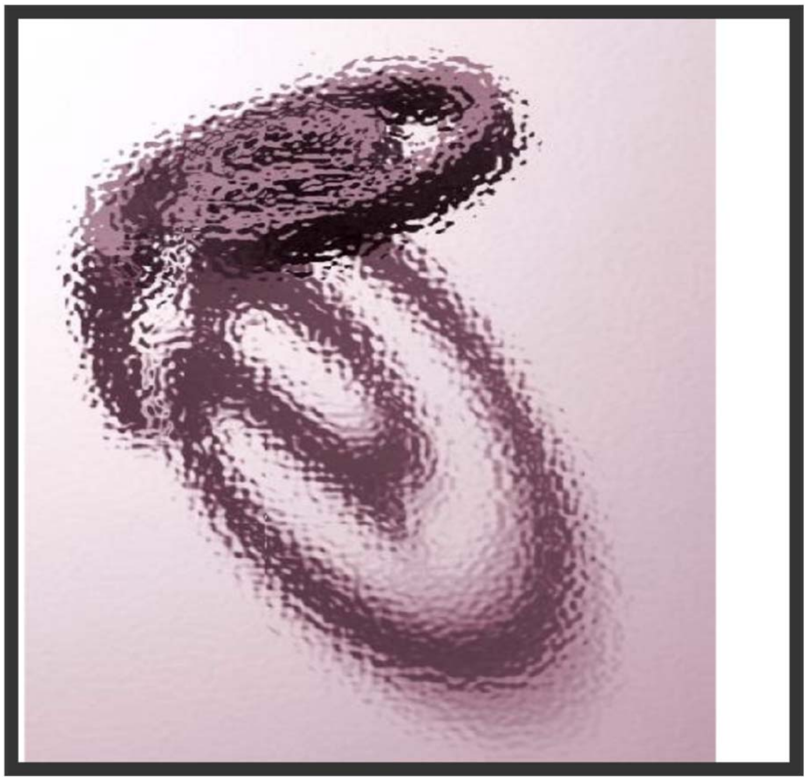
NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Different Perspectives and reflections



A better understanding

From this:



Transposed to this:

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

But How?

- Information is related through multidisciplinary approaches.
- Different areas to look for a common thread:
  - Communication theory
  - Network traffic patterns
  - Complexity science & Chaos theory
  - Topology:
    - braid / knot theory relationships
    - Topological relationships to multi-agent control
    - Clusters and dimensional reduction methods
    - Quantum entanglement
  - Stigmergy (collective self-organization in nature)
  - Sparse data effects in ecosystems

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Highlights of References (1)

1 Paxson and Floyd: Large-scale network ***packet traffic does not follow*** the anticipated model of ***a Poisson distribution***, ***but*** instead has ***a fractal*** quality in the way that the packets self-organize into clusters, producing'burstiness'.

2 Mandelbrot: Observed scaled [self-] similarity (currently called fractal topology) pattern in communication error clusters for signals on telephony transmission lines.

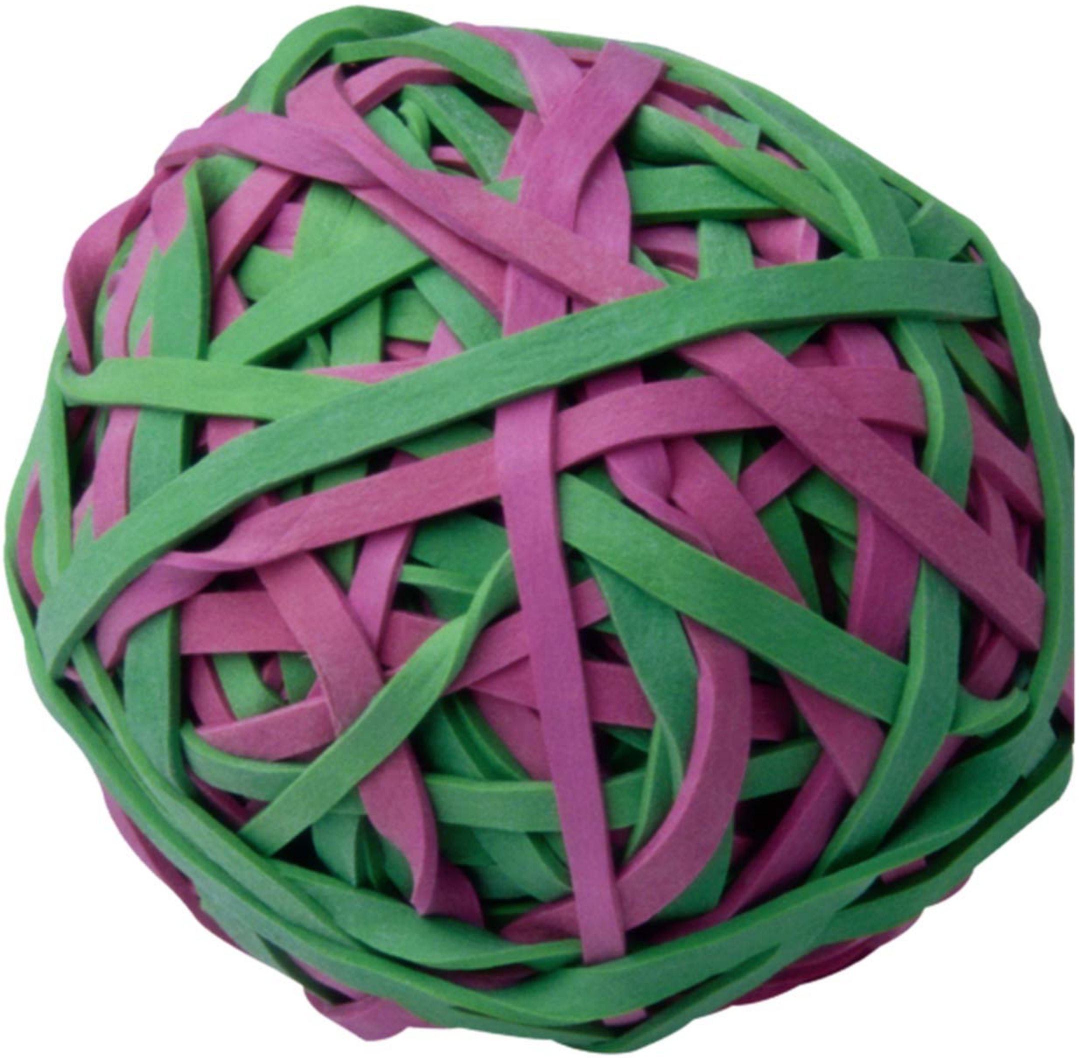
3 Kurlin: Algorithm using topological aspects of ***braid theory*** plans the ***motions of large numbers of robots***. His objective: robots to ***traverse*** a connected graph or '***map***' of connected paths, ***without collisions***.

4 Huebschmann: ***Braid theory*** and its mapping as multiple strings which cross, which he refers to a "crossed module structure". He later describes what he calls a "permutahedron" as an overlay of a shape originating from a ***polyhedron*** that is ***mapped to higher dimensions***.

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Topological relationships

Knot or Braid Theory can show the shared points as intersections or overlaps between linear data sets. These points shared in common between multiple data sets ***might be*** a representation of the same observed object, or may indicate some sort of constraint to the systems as a whole.



NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Highlights of References (2)

5 Holden, Karlsen and Risebro: Initially starting from differing locations / trajectories, ***nonlinear weakly coupled*** degenerate parabolic ***systems*** eventually form an 'orbital' state around an attractor. Gives proof of ***random yet topologically constrained systems self-organizing towards an attractor***.

6 Stremler, Ross, Grover, and Kumar: Reciprocal effort to Holden, et al. - provide a method for the ***analysis of chaos using a topological approach*** - in this case, braid theory to bound the topological space.

7 Jordache: ***Nature can cause rules to be created from multi-agent systems interacting with their environment, by a consensus of information***, i.e. constraints and convergence for random data in general.

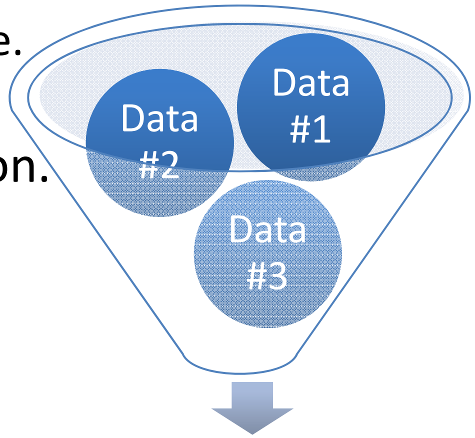
8 Larsen and Speckman: When clusters are not independent, or the clusters are ***not homogeneous*** relative to selected ***dependent variables***, this creates a ***need*** for ***multidimensional scaling***.

9 Tashkova, Šilc, Atanasova, and Džeroski: Models of ***ecosystems*** that are ***high dimensional and nonlinear with sparse data*** are tested, and it is shown that ***global methods are superior*** to localized methods in this type of case.

Distribution Statement A- "Approved for public release; distribution is unlimited"

The high dimensional data problem

- Heterogeneous data from many sources – each has its own perspective.
- Composite information is formed from data, but is intractable – why?
  - Because differing time / space observations & sensor resolutions creates multiple, different representations of any object.
  - Each representation may be reduced to a unique dimension, capturing all relevant information. Many representations = many dimensions.
  - Mathematically, n-dimensional matrices computationally hard.
  - N-dimensional topologies are even harder to conceptualize.
- Need to capture relevant details, yet reduce dimensions.
- Algebraically challenging, but possible topological solution.



Composite / aggregate data.

Highlights of References (3)

10 Mosseri and Ribeiro: High dimensional {Hilbert} space may include remotely linked systems or information, defines ***quantum entanglement***.

11 Solomon: Similar ***topological entanglement*** = vector cross products shows that entangled state of vector / topological spaces is linked to Braid theory.

12 Marseguerra and Zoia: ***Multiple inputs from various sensors create a high dimensional*** problem space for signal processing. In order to reduce the dimensionality so as to create a more tractable lower dimensional problem space, an autoassociative neural network is chosen.

13 Yuksel: ***Focused information sharing across distributed systems***, while avoiding some of the dimensionality issues. (Stochastic nestedness). Method uses a Markov chain of distributed information, and incorporates a 'belief' system to supplement missing information. This creates ***a recursive optimization pattern of minimum information***. Note that a this type of ***recursive pattern is computationally similar to a fractal***. Issue: open question as to whether the selected beliefs would prove valid and sufficient in all cases.

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Highlights of References (4)

14 Masoumi and Meybodi: For ***multi-agent systems*** of learning automata, a ***stigmergic learning algorithm*** with entropy based adaptation can be an improvement over general algorithms.

15 Sukumar, Jayakumar and Geetha: ***High-dimensional problem*** & methodology - Multiple web services compared, to select the best for certain tasks. Comparisons are made on multiple simultaneous parameters producing mappings that create clusters, each of which represents different criteria (high-dimensional problem).

Dimensional reduction is done using space-filling curves (SFC). The SFC can be thought of as a projection of some higher dimensional data set containing clustered subsets that map to portions of the curve (***Fractal structure***).

Choosing best SFC: Hilbert SFC shortcomings discussed - limited scalability and mapping irregularities, and failing to successfully map exact cluster patterns. Previous paper cited that mathematically proves the Peano SFC to be superior to the Hilbert SFC.

{Lays groundwork for investigation of novel / adaptive SFCs for better 'fit'.}

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Conclusions

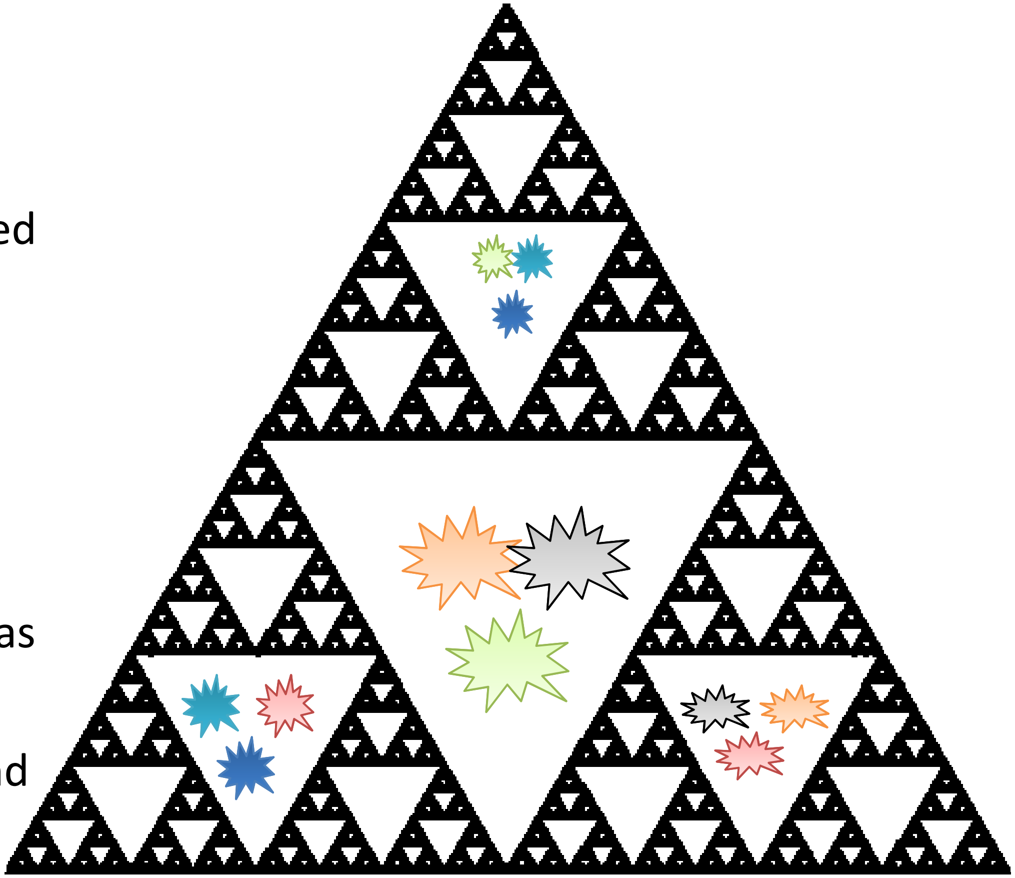
- Various approaches to large-scale distributed information.
- Topological approaches provide some mathematically tangible methods.
- Fractal curves both can be dimensional reduction of high-dimension space, and provide bounding conditions for multi-cluster relationships.
- Current fractals are statically limited, use an adaptive fractal method.

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

Non-overlapping topology relationships – multiple clusters

Capturing relationships for multiple cluster-of-cluster topologies, while reducing dimensionality:

- Non-stochastic solution is needed (preserve interrelationship details).
- Possible non-integer resulting dimensionality.
- Various self-similar curves have had success in the past – e.g. fractal models.
- Resulting curve / map emerges as the abstraction of complex collective interrelationships, dictated by world constraints and objectives or goals the autonomous systems .



NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"

References-1

1. Paxson, V. & Floyd, S. (1995). Wide area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking, 3*(3), 226-244.
2. Mandelbrot, B., (March 1965). Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Transactions on Communication Technology.* 71-90.
3. Kurlin, V., (2012). Computing braid groups of graphs with applications to robot motion planning. *Homology, Homotopy and Applications 14*(1), 159-180. doi: 10.4310/HHA.2012.v14.n1.a8
4. Huebschmann, J. (2012). Braids and crossed modules. *Journal of Group Theory 15,* 57-83. doi: 10.1515/JGT.2011.095
5. Holden, H., Karlsen, K. & Risebro, N. (2003). On uniqueness and existence of entropy solutions of weakly coupled systems of nonlinear degenerate parabolic equations. *Electronic Journal of Differential Equations, 2003*(46), 1-31. ISSN: 1072-6691.
6. Stremler, M., Ross, S., Grover, P. & Kumar, P. (2011). Topological Chaos and Periodic Braiding of Almost-Cyclic Sets. *Physical Review Letters 106*(11), 114101. doi: 10.1103/PhysRevLett.106.114101
7. Jordache, S., (2009). A framework for the study of the emergence of rules in multiagent systems. *Annals of DAAAM & Proceedings*, 1285-1286. ISSN 1726-9679, ISBN 978-3-901509-70-4.

References-2

8. Larsen, D. & Speckman, P. (2004). Multivariate Regression Trees for Analysis of Abundance Data. *Biometrics 60*(2), 543-549.
9. Tashkova, K., Šilc, J., Atanasova, N., & Džeroski, S. (2012). Parameter estimation in a nonlinear dynamic model of an aquatic ecosystem with meta-heuristic optimization. *Ecological Modelling*, *226*36-61. doi:10.1016/j.ecolmodel.2011.11.029
10. Mosseri, R. & Ribeiro, P. (2007). Entanglement and Hilbert space geometry for systems with a few qubits. *Mathematical Structures in Computer Science 17*, 1117-1132. doi:10.1017/S0960129507006329
11. Solomon, A. I., (2010). Bell, Group and Tangle. *Physics of Atomic Nuclei, 73*(3), 524-532. doi: 10.1134/S1063778810030166
12. Marseguerra, M. M., & Zoia, A. A. (2005). The autoassociative neural network in signal analysis: I. The data dimensionality reduction and its geometric interpretation. *Annals Of Nuclear Energy*, *32*(11), 1191-1206. doi:10.1016/j.anucene.2005.03.006
13. Yuksel, S., (2008). Stochastic nestedness and information analysis of tractability in decentralized control, *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, 1465-1472. doi: 10.1109/ALLERTON.2008.4797735
14. Masoumi, B., & Meybodi, M. R. (2011). Speeding up learning automata based multi agent systems using the concepts of stigmergy and entropy. *Expert Systems With Applications, 38*(7), 8105-8118. doi: 10.1016/j.ewsa.2010.12.152
15. Sukumar, S., Loganathan, J., & Geetha. T. (2012). Clustering web services based on multi-criteria service dominance relationship using peano space filling curve. *2012 International Conference on Data Science & Engineering (ICDSE),* IEEE.

Acronyms

- SoS – System of systems
- SFC – Space-filling curves

NAVAIR Public Release 2014-107
Distribution Statement A- "Approved for public release; distribution is unlimited"