SRI International

A Reasoning-based Framework for the Computation of Technical Emergence

GTM 2013-Atlanta, GA

25 September 2013

Jeffrey Alexander
Associate Director for Research & Analytics
Center for Science, Technology & Economic Development

Acknowledgement

- *All research presented was supported by the Intelligence Advanced Research Projects Activity via the Department of the Interior National Business Center under contracts D11PC20152 and D11PC20155.*

- *The views and conclusions contained herein are those of the authors, and should not be interpreted and necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.*

Key Topics
The challenges of computational emergence detection

- Defining Technical Emergence
- Indicator Development
- Indicator Evaluation & Presentation
- Research Directions

Defining Technical Emergence

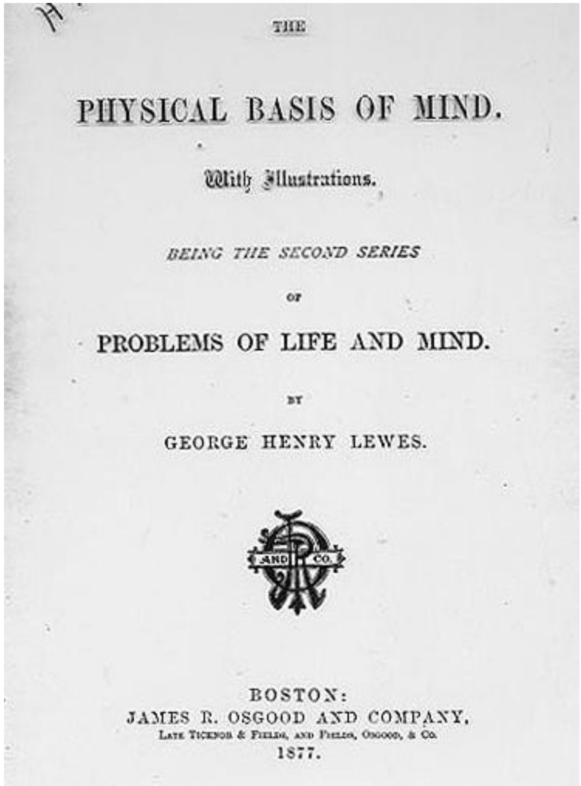
Starting Point: Definitions

> "There's no point in being precise if you don't even know what you're talking about."

John von Neumann
Pioneer in computer science research

Emergence of the concept of emergence

- “Emergence” coined as a response to Darwinian biology
  - As structures evolve to a level of sufficient complexity, they manifest qualitatively unique characteristics and behaviors
- “Emergence” as a characteristic of complex adaptive systems (CAS) (Waldrop, 1992)
  - Agents in large, unordered agglomerations exhibit behaviors which appear to be ordered and yet are largely unpredictable
  - Applied to stock markets (bubbles), ecological systems (birds flocking), materials science (crystallization)


THE

PHYSICAL BASIS OF MIND.

With Illustrations.

BEING THE SECOND SERIES

OF

PROBLEMS OF LIFE AND MIND.

BY

GEORGE HENRY LEWES.

BOSTON:
JAMES R. OSGOOD AND COMPANY,
LATE TICKNOR & FIELDS, AND FIELDS, OSGOOD, & CO.
1877.

How does emergence apply to the global S&T system?

- Complexity in the global science system
  - Increasing scale (volume of activity continues to rise)
  - Increasing scope (spread of science system to encompass new geographies & knowledge domains)
  - Increasing interdependence (knowledge flows are diverse and lack predictability) (Bonaccorsi & Vargas, 2010)

- Complexity in the global technological system
  - Rise of technical systems requiring broad, multi-dimensional expertise (Rycroft & Kash, 1999)
  - Complex technical systems are developed and managed by organizational structures which are also complex (Graf, 2011; Powell et al, 2005)
  - “Breakthroughs” arise in many different contexts (large firms vs. small, developed nations vs. developing, convergence of multiple technologies) (Arthur, 2009)

- The global S&T system has evolved to a level of complexity consistent with the phenomenon of “emergence”

Working Definition and Propositions

- Technical emergence can be defined as:
  - The phase during which a concept or construct is adopted and iterated by an expert community of practice, resulting in a fundamental change in (or significant extension of) human understanding or capability

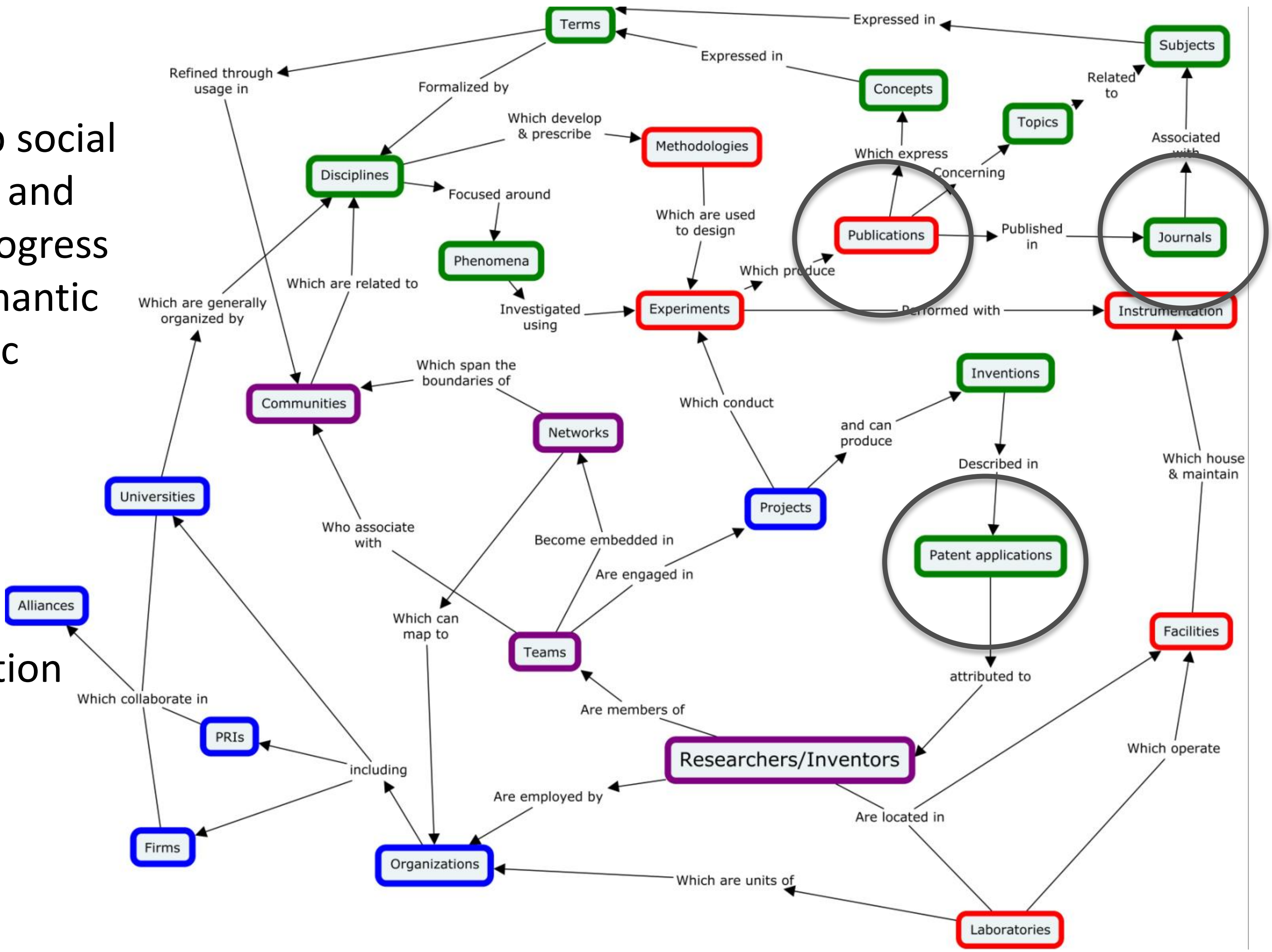
- Sample propositions
  - P1: A technical concept can emerge only if a **relevant community** forms around that concept to fuel its **elaboration and development**
  - P2: A technical concept will emerge from the activity of a **particular subset of a broader community**, and the **dynamics at the broader level** are an important influence on the emergence at the level of the technical concept.
  - P3: Emergence of a technical concept will be detectable as a function of the dynamics of the community and system in which the concept is embedded. **Change is the critical signal of emergence**, rather than absolute levels of activity or organization.

An emergent concept is detected in the context of its fundamental knowledge domains, as represented by the community of practice which adopts and iterates that concept.

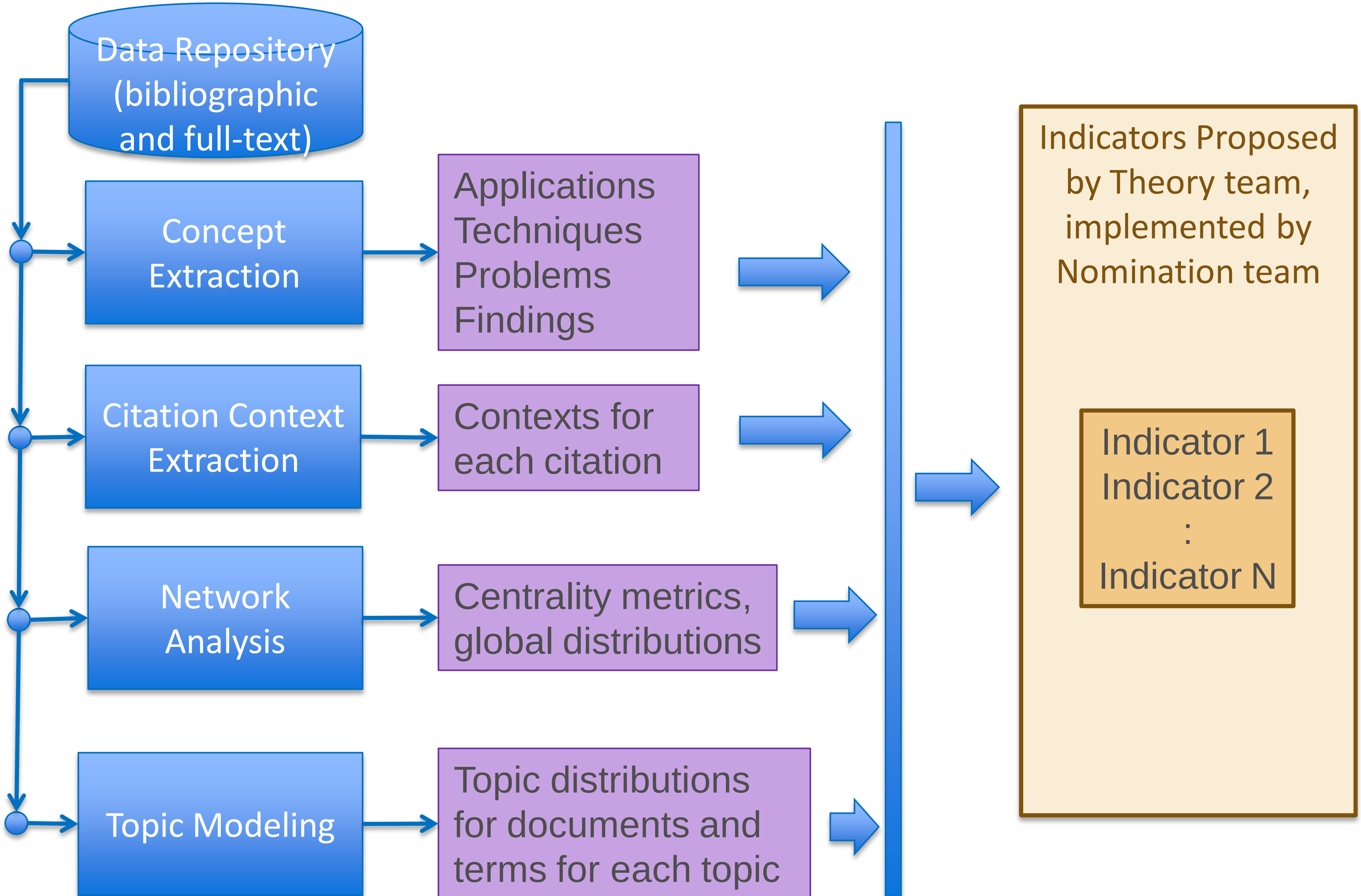
Indicator Development

Emergence is Inherently Social

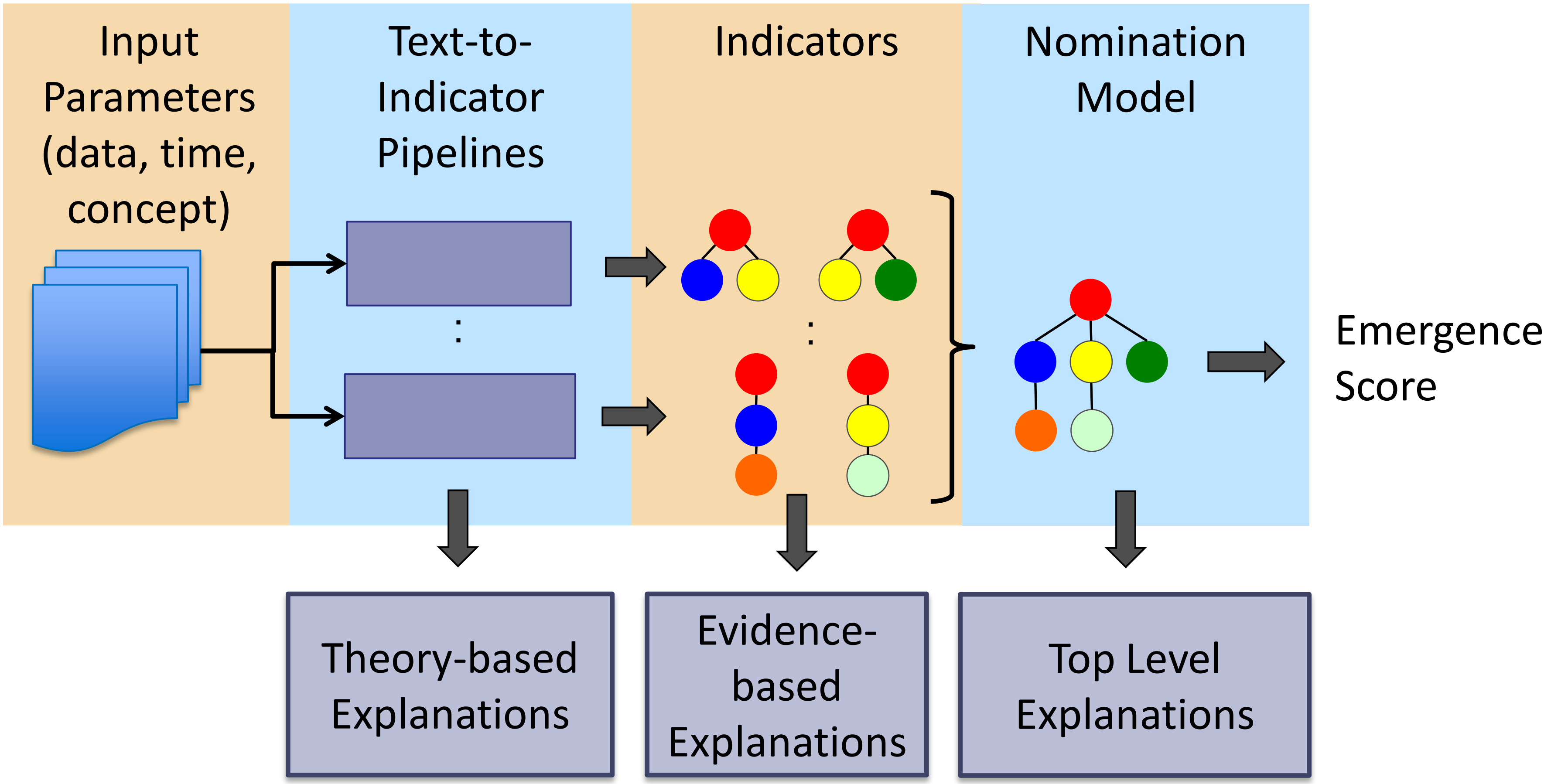
- We can map social interactions and technical progress through semantic and linguistic patterns in documents used in scientific & technical communication



Framework for the Copernicus Approach



Copernicus Workflow



Linking theory development & indicator discovery

Indicator	Construct	Target	Explanation
Community Growth	"Critical mass" of research effort	Concept attracts researchers	Rapid uptake of a concept suggests critical mass will be reached
Core Community	Seminal works or authors	Concept as a focus of attention	Robust & growing core lends stability to the overall community
Topic Proliferation	Formation of "camps" within concept	Distinct clusters of research effort	"Camps" will appear as coherent subcommunities focused on their particular subtopics
Application Emergence	Diffusion of concept across multiple applied domains	Entropy of topics across publications	As a concept proves useful, researchers will see how it can be adapted to different problems

Indicator Evaluation

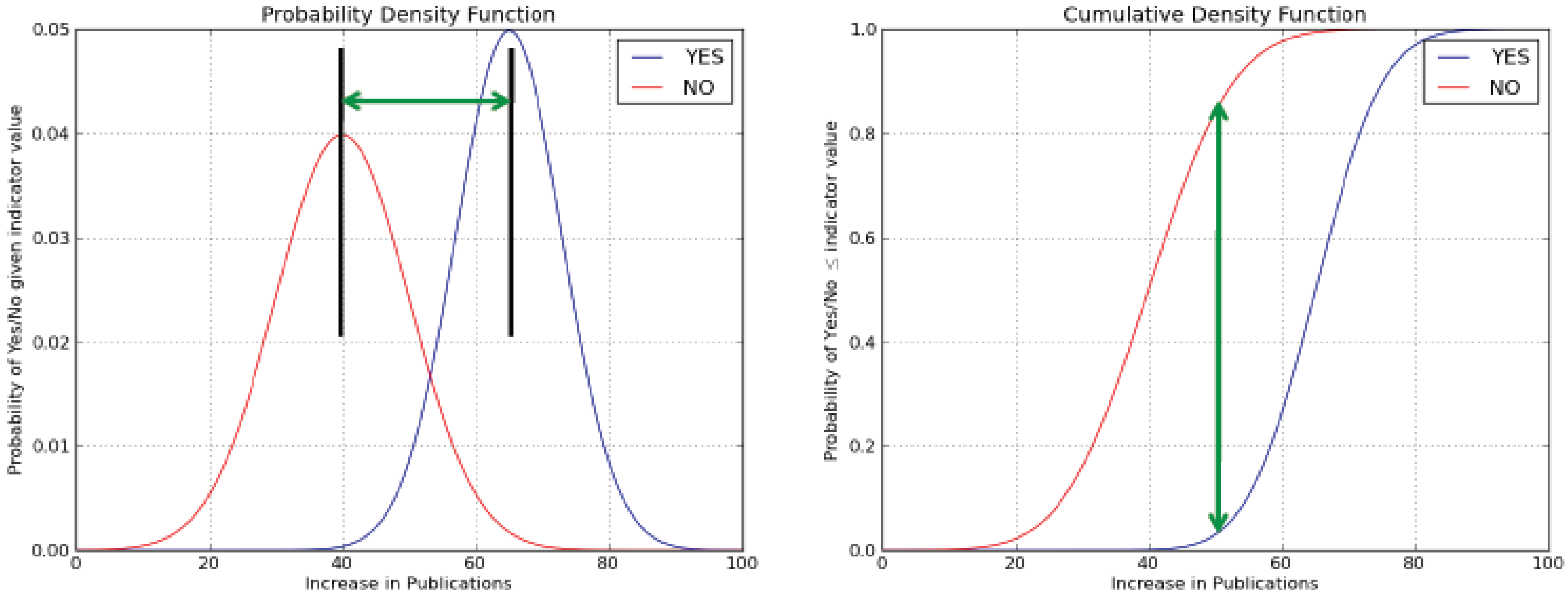
Key Attributes of Robust Indicators

- Initial challenge is essentially binary
  - "Is concept X emerging or not at this point in time?"

- Powerful
  - Clear distinction between values indicating "yes" versus "no"
  - Individual indicators address different aspects of dataset or phenomenon
  - Each indicator contributes substantively to predictive accuracy (lift)
  - Modular—indicators can be removed if appropriate data not available

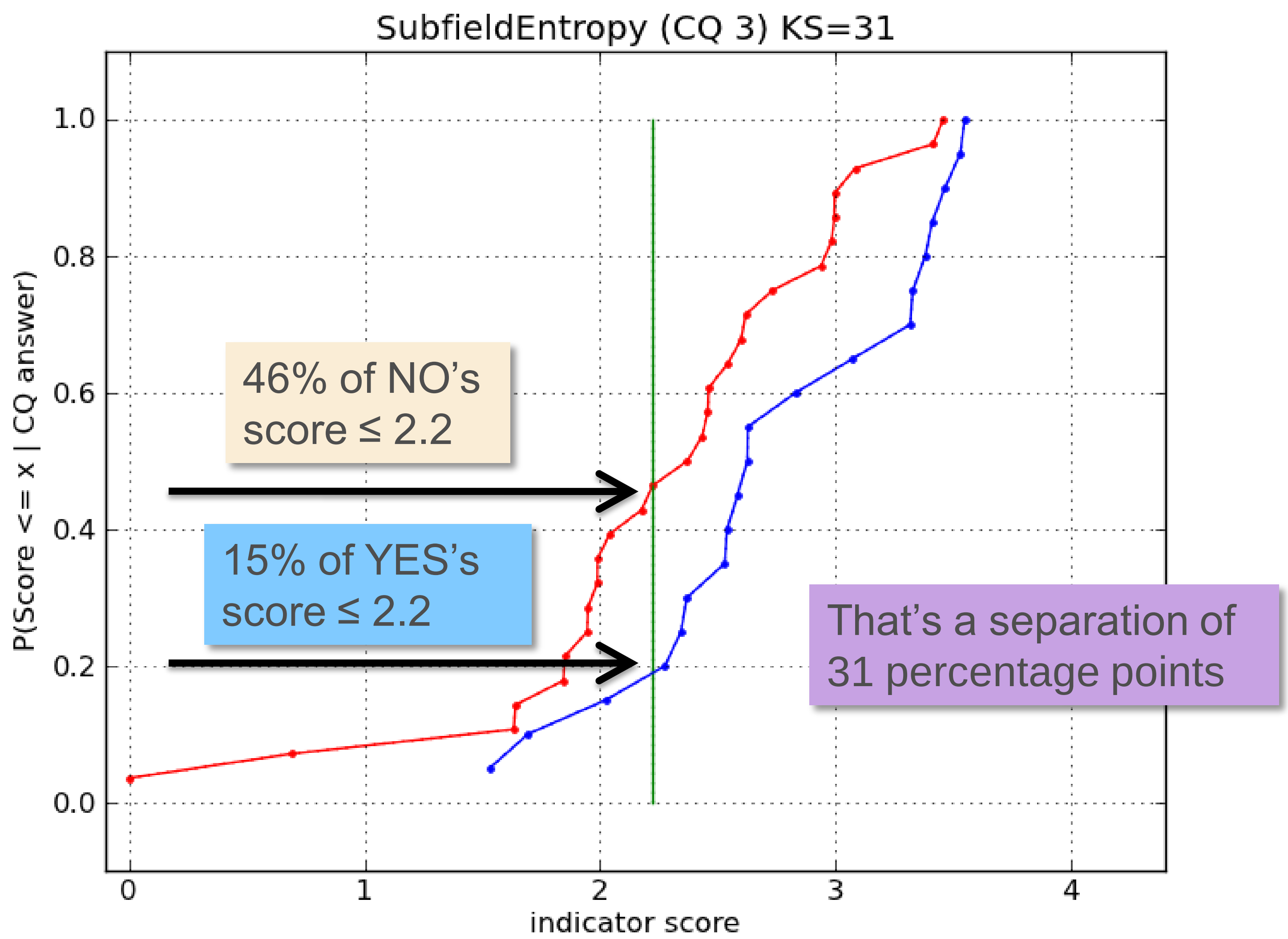
- Parsimonious
  - Interpretation should be intuitive
  - Solid theoretical grounding (we should know it indicates X because…)
  - Maximizes predictive/analytical value with minimum number of indicators

Illustration of Indicator Distinctiveness

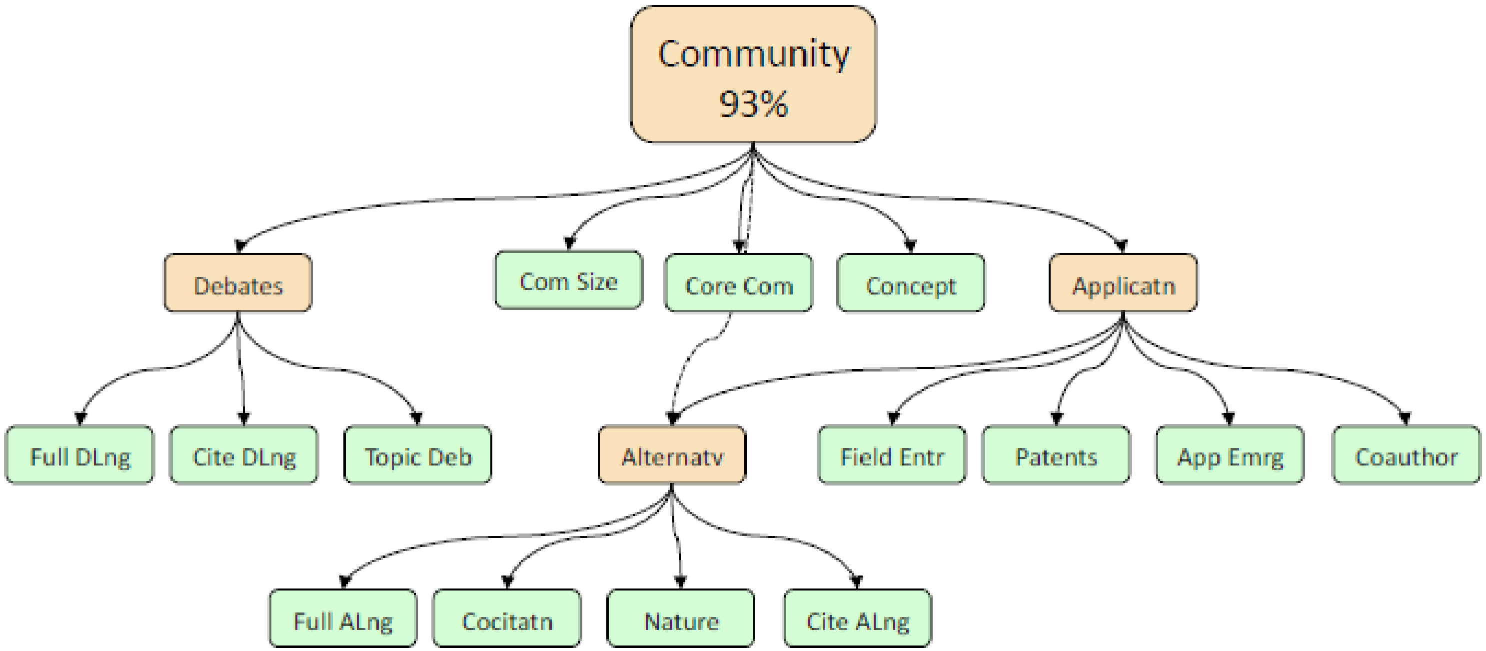


- Look at distribution of scores over known cases of emergence and non-emergence
- Examine distribution of scores between the two groups
- Represents scores as a cumulative distribution function (% of yes or no cases with a given indicator score or lower)

Sample Evaluation via Kolmogorov-Smirnov

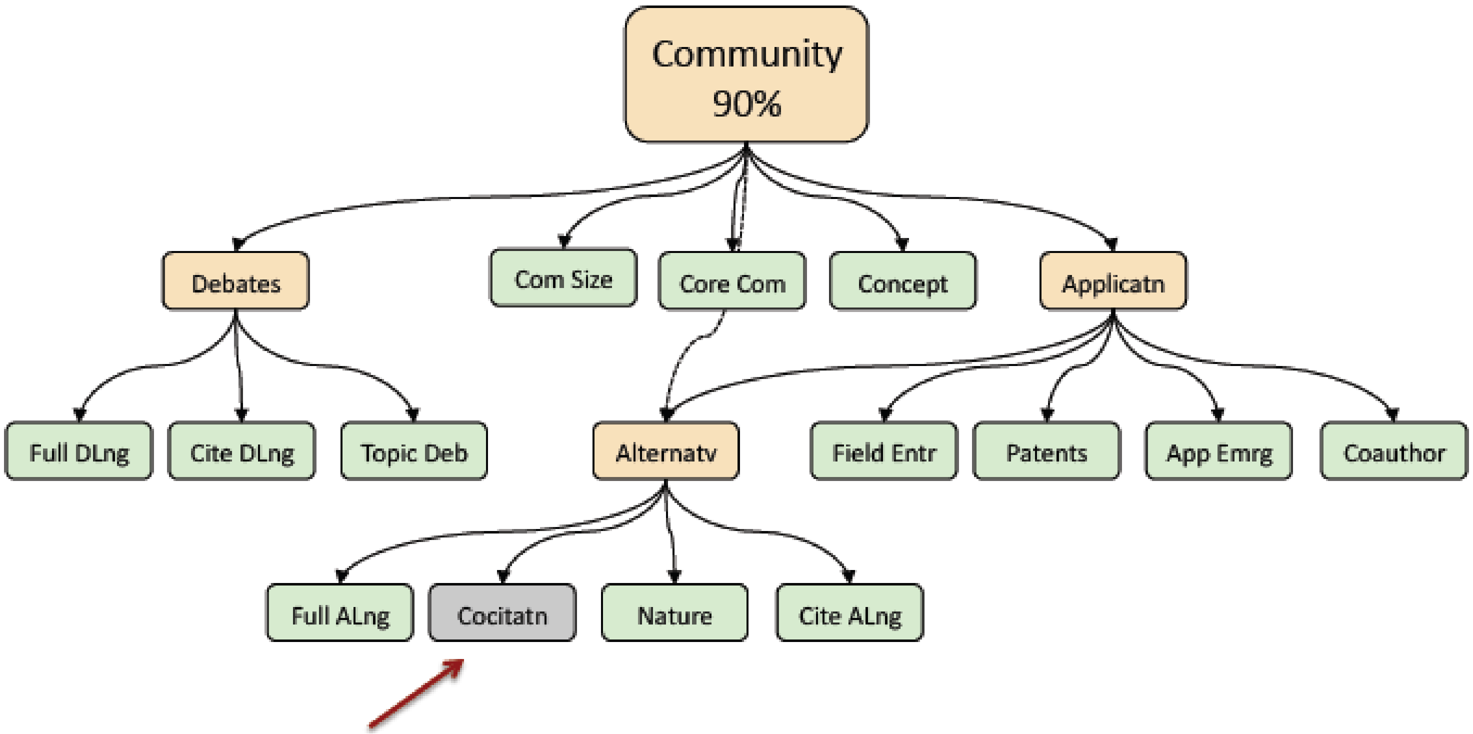


Model Development: Simple Bayesian Network



The model is a Bayesian Network. Each node has a Conditional Probability Table specifying the probability of any given state given the states of the parents.

Measuring Lift via Bayesian Model



We turn off an indicator to measure its impact on the score. In this example, the Community score dropped from 93% to 90%.

Research Directions

On-going Investigations

- **Methodological research & development**
  - Improve concept and context analysis
    - Can we identify entities such as theories, methods, tools, datasets, & assertions in the text?
    - Does citation context provide better clues about the dynamics of technical emergence?
  - Improve integration of methods
    - What can topic modeling reveal about the substructures of emergent concepts?
- **Research into the nature of emergence**
  - How easy is it to detect emergence, or pre-emergence, or post-emergence?
  - What types of indicators are most revealing about incipient emergence? Why do some concepts emerge later than expected?

Thank You

jeffrey.alexander@sri.com

john.byrnes@sri.com

Headquarters: Silicon Valley

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000

Washington, D.C.

SRI International
1100 Wilson Blvd., Suite 2800
Arlington, VA 22209-3915
703.524.2053

Princeton, New Jersey

SRI International Sarnoff
201 Washington Road
Princeton, NJ 08540
609.734.2553

Additional U.S. and international locations

www.sri.com