OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

Research Challenges in Forecasting Technical Emergence

LEADING INTELLIGENCE INTEGRATION

Dewey Murdick, IARPA
25 September 2013

"Invests in high-risk/high-payoff research programs that have the potential to provide our nation with an overwhelming intelligence advantage over our future adversaries"
http://www.iarpa.gov/

A Few Interesting Research Problems

- Scan for technical emergence
  - Move beyond search
  - Reliably query for indicative "patterns of technical emergence" without starting with a known, named subject
- Analyze diverse and large data streams across disciplines, cultures, and languages
  - Support strategic investment
  - Facilitate discovery and innovation
- Forecast scientific, technical, application, and market events
  - Quantitatively event forecasts
  - Improve accuracy and early event event detection

Foresight and Understanding from Scientific Exposition (FUSE) Program

Reduce "technical surprise" via reliable & validated, early detection of emerging scientific and technical capabilities across disciplines and languages found within the full-text content of scientific, technical, and patent literature

Special focus from the outset on multiple languages, Phase 2 focus on **English** and **Chinese**

Novelty	→ Discover <u>patterns</u> of emergence and <u>connections</u> between technical concepts at a speed, scale, and comprehensiveness that exceeds human capacity
Usage	→ <u>Alert analyst</u> of emerging technical areas with sufficient explanatory evidence to support further exploration

What is technical emergence?

Hypotheses from Phase 1

- A concept has emerged if it has been accepted by others within and beyond one's community. **~Columbia**

- A concept is emerging when its "actant network" is increasing in robustness. **~BAE**

- A concept has emerged when evidence has appeared that the concept is new and unexpected, noticeable and growing. **~Raytheon BBN**

- A concept is emerging when it is identifiable by its own practitioners, enables a capability that was not achievable previously, and persists. **~SRI**
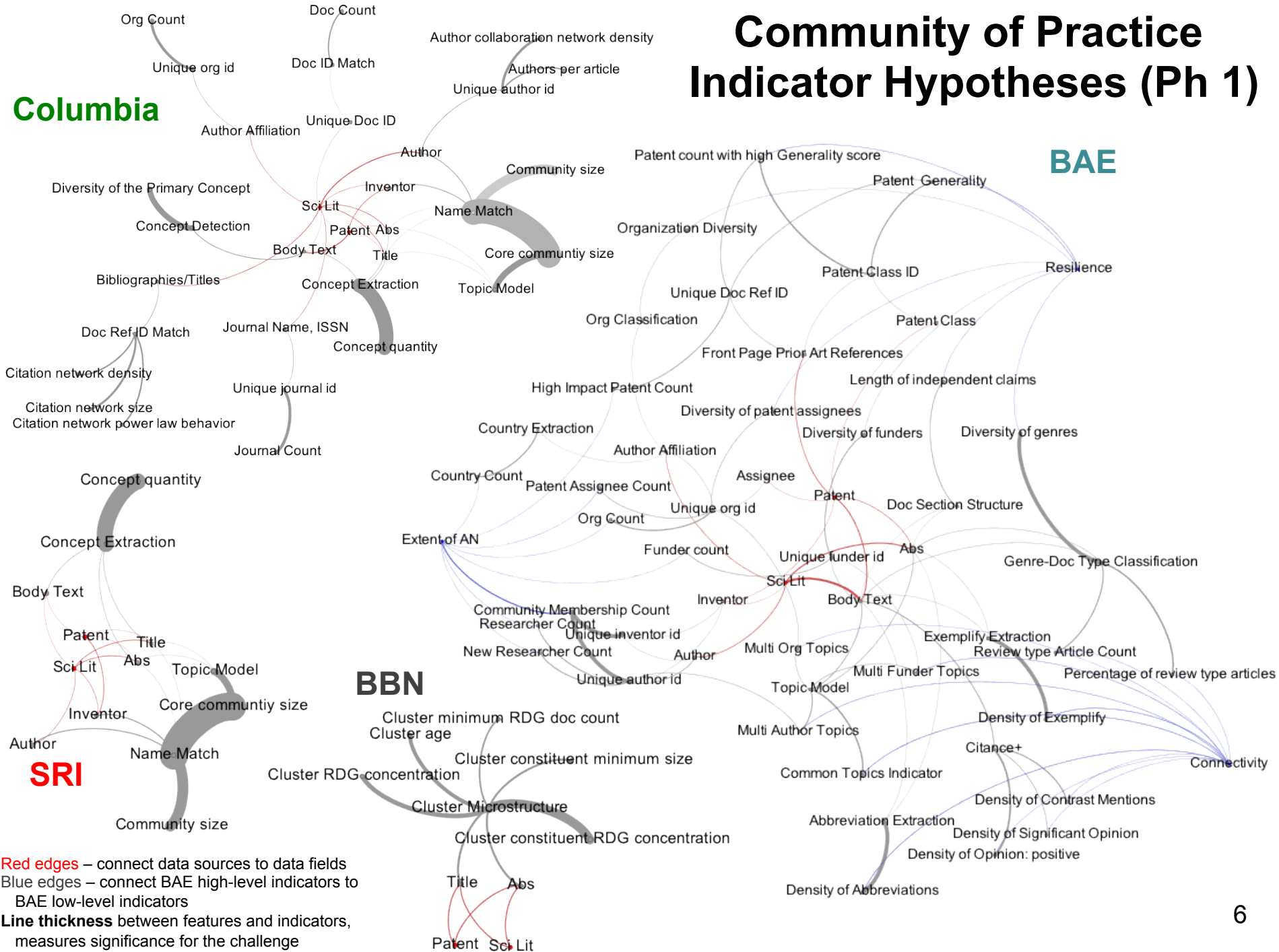
Many ways to probe technical emergence
- Community of Practice
- Practical Application
- Debates
- Alternative
- Acceptance
- Interdisciplinarity
- Attention (Citation) Prediction
- Dominant sub-topic within set
- Commercial Application
- Infrastructure

Community of Practice Indicator Hypotheses (Ph 1)

Columbia

BAE

SRI

BBN

Red edges – connect data sources to data fields
Blue edges – connect BAE high-level indicators to BAE low-level indicators
Line thickness between features and indicators, measures significance for the challenge

6

Evaluation Attempt #1: Case Studies

- Drawn from diverse areas of scientific inquiry & application:
    - Biological Sciences / Biotechnology
    - Computer Science / Information Science; Engineering
    - Mathematics / Statistics
    - Physical Sciences; Earth Science
    - Medical / Clinical / Infectious Disease / Health Services;
    - Social Sciences; …

- Technical emergence measured from "real world" view point, but connected to literature

- Multiple case studies to be produced; some are held back for evaluation
    - Case studies are representative but not comprehensive
    - Insufficient to train technical emergence classifiers
    - Limited examples of emergence & non-emergence (10s planned)
    - Reference baseline has limited temporal resolution (~5 year blocks)

Phase 2 Evaluation: Nomination Test
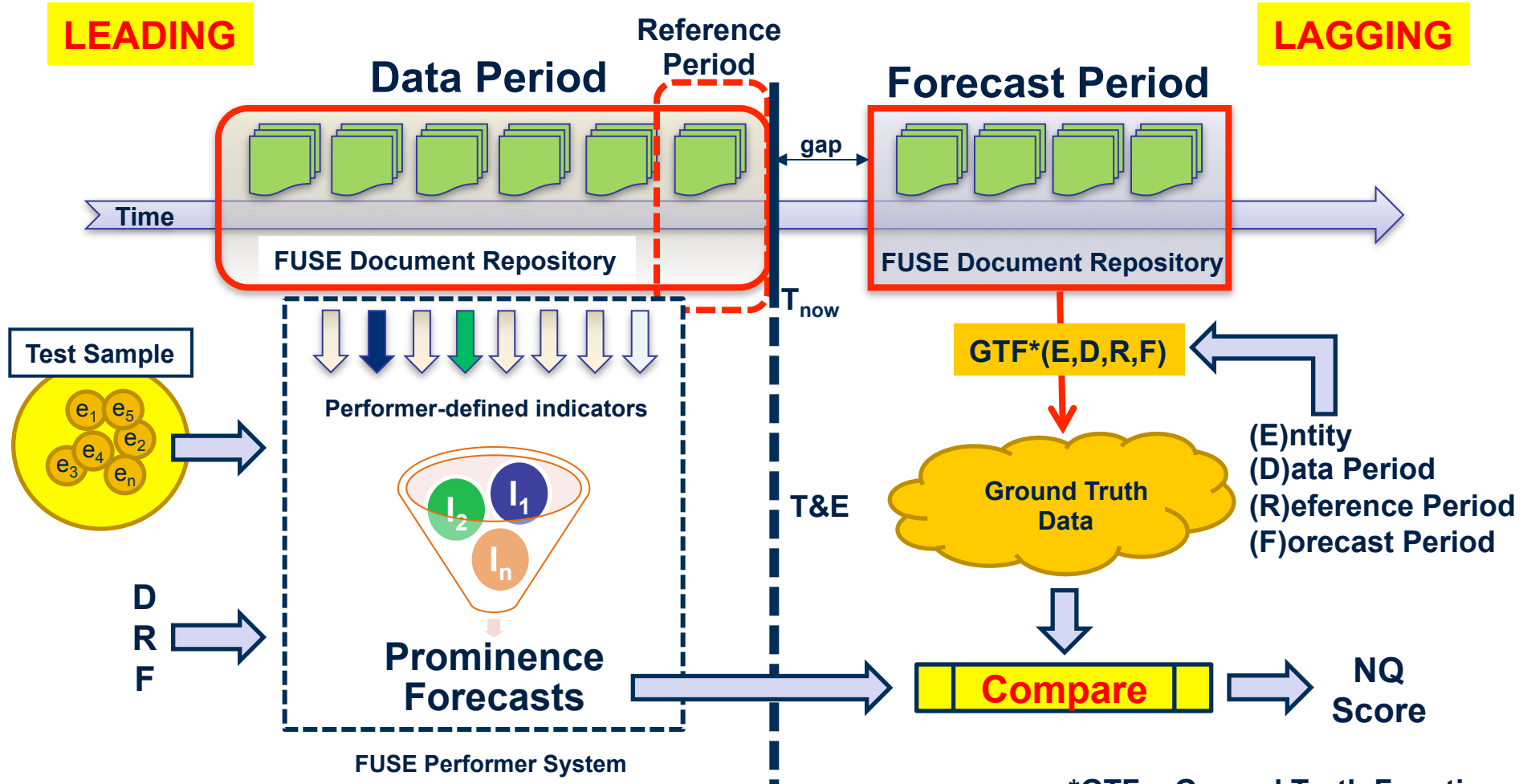
Indicator Development and Testing Underway

Regular analysis and evaluation of each team's features (e.g., scientific noun phrases, topic models) and their portfolio of indicators (i.e., quantitatively measured aspects / patterns of technical emergence)

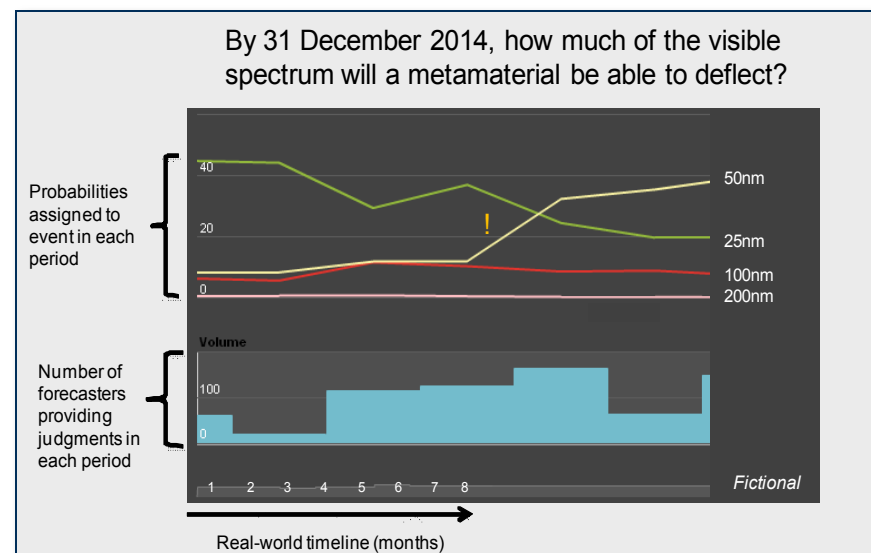| Promising Midterm Indicator Types	Fundamental Research
• Citation, Author Networks (All) • Topic Diversity (SRI) • Citation Context and Sentiment (SRI) • Technology and application concept type evolution (SRI) • Patent classification dynamics (SRI, BAE) • Emerging cluster / hot patent status (BAE) • Patent originality (BAE) • Corporate, Academic patent authorship (BAE) • Topic modeling across time, thread dynamics (BBN) • Research levels (BBN) • Time series analysis, extensive portfolio (COL) • Temporal pattern classification, time-series clustering (COL)	• Argumentative Zoning (SRI, COL) • Time-dependent term co-occurrence (SRI) • Author-topic modeling (SRI) • Operations on annotated graphs, e.g., scientific concepts, terms (SRI) • Chinese patent indicators (BAE, BBN) • Fine-grained topic models (BBN) • Causality modeling framework (BBN) • Primary concept mentions (COL) • Citation sentiment (COL)

Now Developing a Market for Scientific and Technical Forecasting

- **Goal:** Generate precise, testable forecasts for S&T developments

- **Approach:** Build world's largest prediction market for S&T events
  - Thousands of subject matter experts in dozens of countries will make nuanced conditional forecasts for around one thousand S&T events
  - Data-driven (i.e., scientific and patent literatures) indicators will be used to generate questions and adjust forecasts

- **Evaluation:** Forecasts will be scored against actual events, as they occur
- **Potential impact:** Dramatically improve S&T foresight with actionable information
- **Schedule:** June 2013–June 2015

By 31 December 2014, how much of the visible spectrum will a metamaterial be able to deflect?

Probabilities assigned to event in each period

Number of forecasters providing judgments in each period

Fictional

Real-world timeline (months)

Teams will Generate Questions

- What is the probability of a 10cm carbon nanotube being fabricated before 31 Dec 2014?

- Will the number of accepted articles for the 2015 International Conference on Machine Learning (ICML) conference that contain the term 'deep learning' in the title/abstract exceed those that contain the term 'support vector machine(s)' in the title/abstract?

- How many unique assignees will have at least two USPTO patent applications published using the term 'Type III Secretion System' in its title/abstract/background/claims between 1 Oct 2013 and 30 Sep 2014?

- By 31 Dec 2017, how many FDA-approved products will be based on RNA interference?

- Will there be reported shortages of technetium-99m in the US in 2015?

Discussion & Questions

Dewey Murdick, Ph.D.
Program Manager, IARPA
dewey.murdick@iarpa.gov