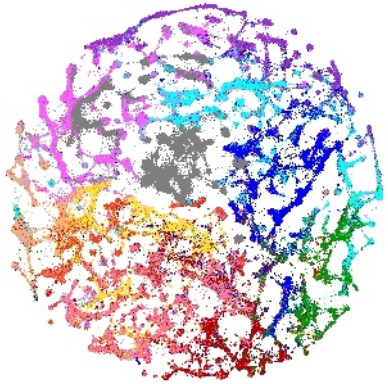
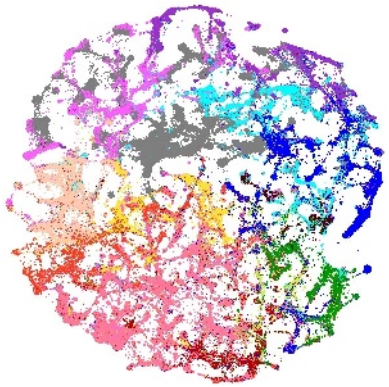


SCITECH STRATEGIES

Better Maps • Better Solutions



Physics | Chemistry | Engineering | Biology | Disease | Medicine | Brain | Health | Social | Computer | Patents

Validating emerging topics in science and technology

GTM Workshop 2013
Atlanta, GA
September 25, 2013

Henry Small, Kevin W. Boyack, Richard Klavans
SciTech Strategies, Inc.

ACKNOWLEDGMENT

- This research is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D11PC20152. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.
- Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

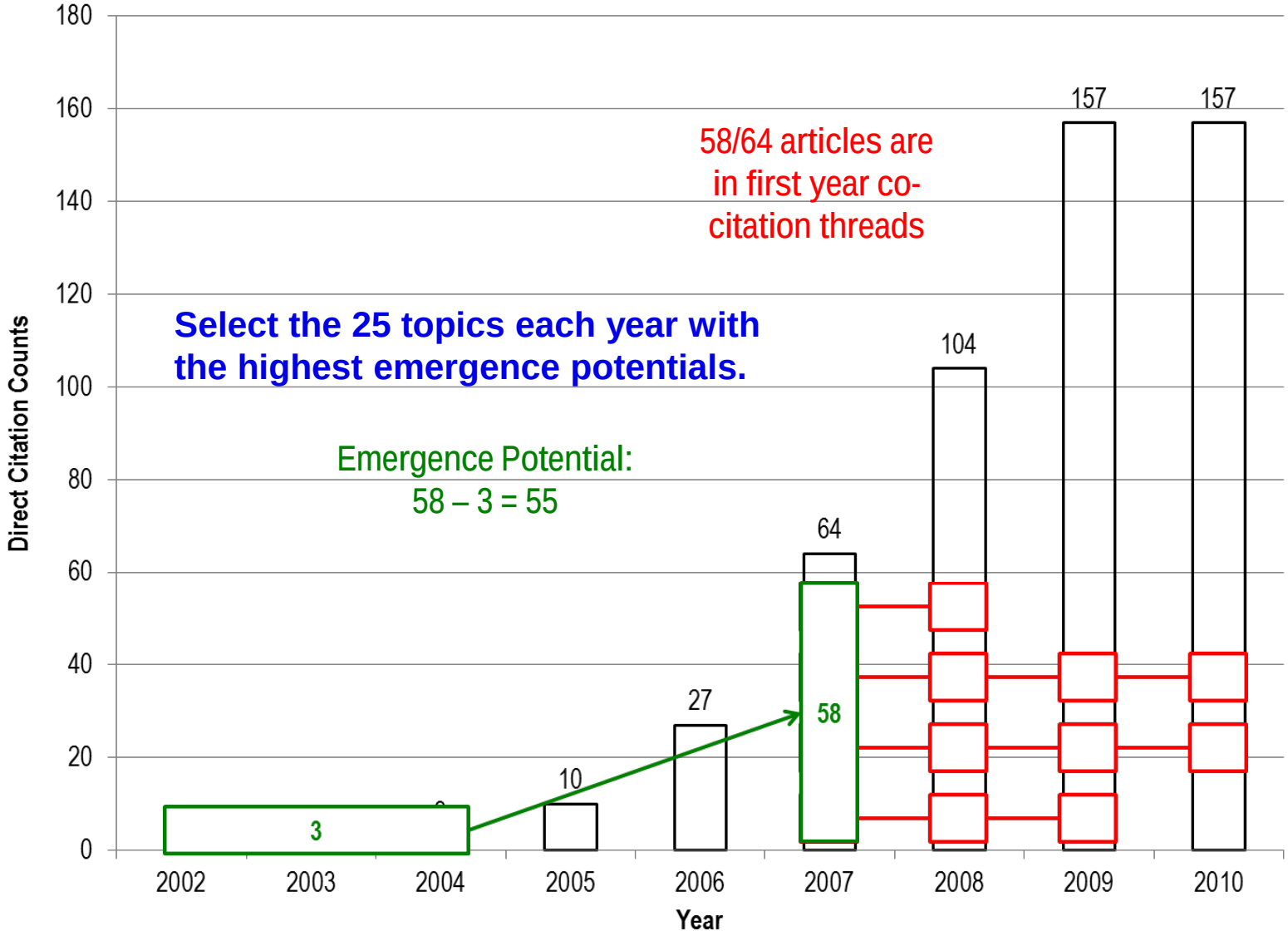
OVERVIEW

- Recent work to identify emerging topics
  - Approach combines two global models
  - Approach used to nominate 25 emerging topics for each year from 2007-2010
  - 71 topics in total were characterized – emergence date, event, description, etc.
- Evidence of importance
  - Awards to authors of most cited papers; recognition in the science press
  - False negatives were not investigated (emergent topics that weren't on our list)
- New work: Reverse validation
  - Start with lists of award winners
  - Find related topics in the map and calculate emergence metrics
  - Compare with control group

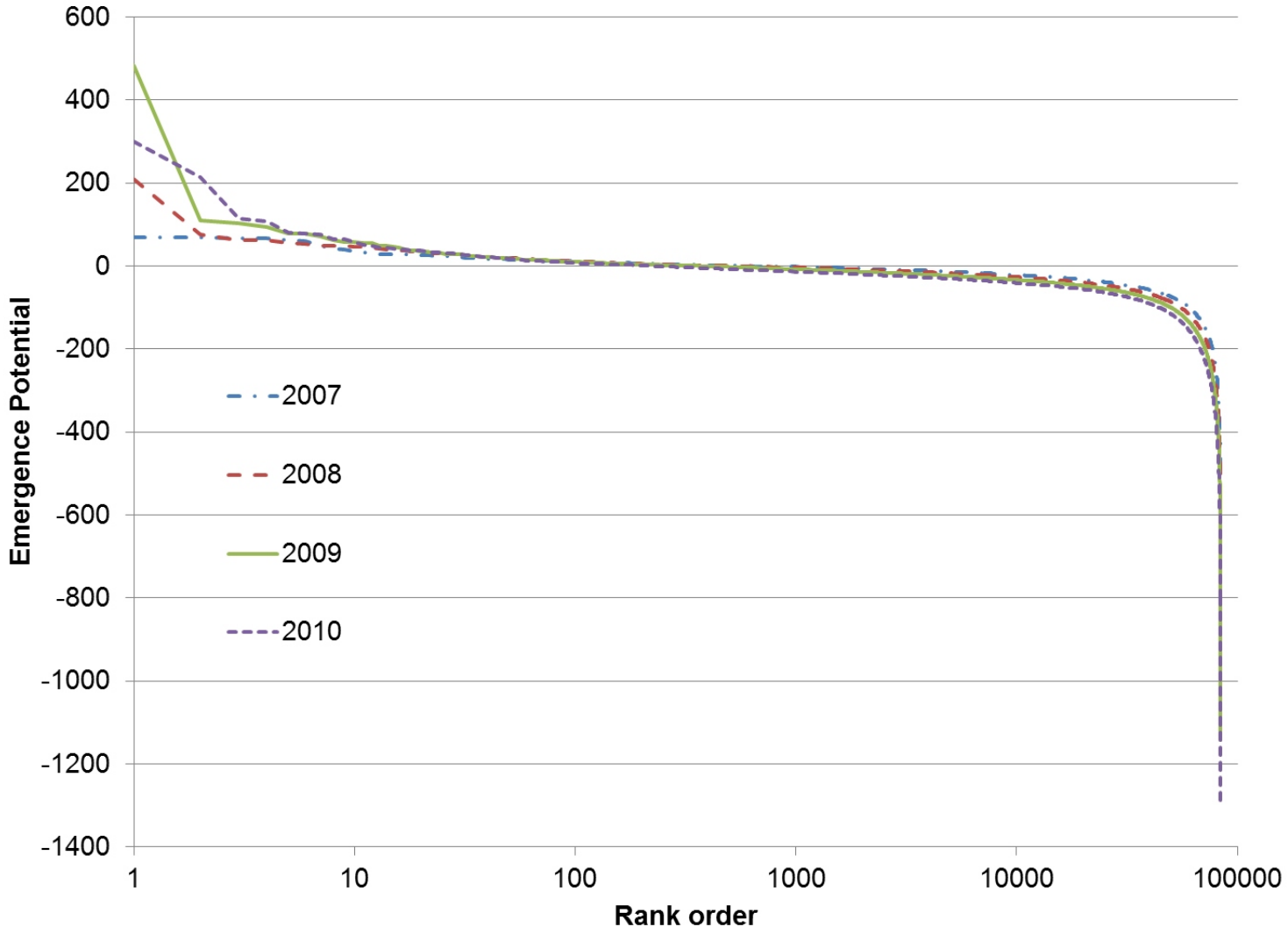
IDENTIFYING EMERGING TOPICS

- Technique based on using two models of science
  - Premise is that two votes for emergence (from two different models) are better than one
- Direct citation model
  - Waltman, L., & Van Eck, N. J. (2012). JASIST 63(12), 2378-2392.
  - Scopus data, 18M documents, 1996-2010, 84,000+ clusters
  - Most clusters last all 15 years
  - Clusters that are born are a strong indication of “something new”
- Co-citation model
  - Boyack, K. W., & Klavans, R. (2013). JASIST, in press.
  - Same data, annual clusters, clusters linked into threads
  - Most threads are short; lots of births and deaths

Selecting emergent clusters

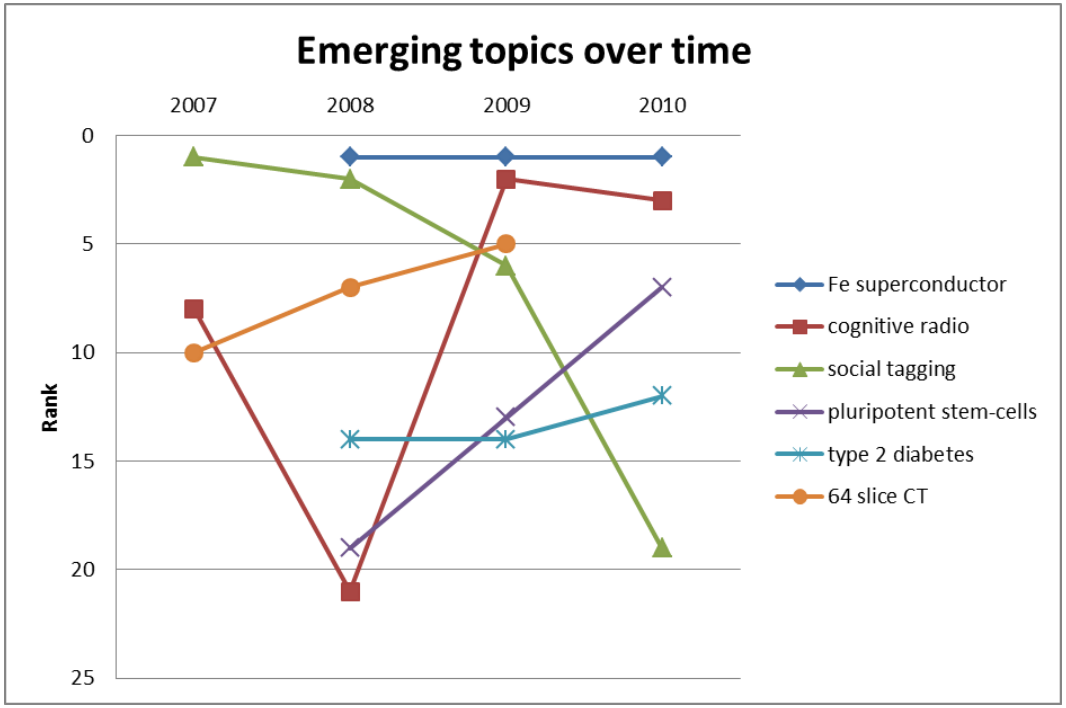


EMERGENCE POTENTIAL (EP)



SELECTION

- For each year, the emergence potential was computed for each direct citation cluster
- The top 25 clusters for each year (2007-2010) were selected for further analysis
- Some clusters were listed in the top 25 in multiple years
- 71 clusters in total were nominated
- Each cluster was characterized



CHARACTERIZATION

- Based on lots of reading and analysis of citation patterns and counts
- Emerging topics were characterized in several ways:
  - Broad fields: 40% medicine/life sciences; 34% engineering and computer science; 26% physical sciences.
- Motive force for emergence
  - Scientific discovery – a new or unexpected finding is made or fundamental knowledge is gained (~60%)
  - Technological innovation – existing science or technology is used to create new devices or capabilities that serve specific purposes (~40%)
  - Exogenous event – (~56%) includes government actions, software releases, technical standards, clinical trials, health threats, product introductions (factor in 44% of discovery and 75% of innovation cases).
  - Internal event/discovery – topic emergence was direct result of publication of specific findings (~40%) The discovery paper is highly cited.

EMERGING TOPICS (20 of 71)

id	rank	label	years	type	Ev year	HC year	Em year	Ev to HC	Ev to Em	HC to Em	H	cons
1	1.0	iron-based superconductors	2008-2010	dis	2008	2008	2008	0	0	0	42.0	80.8
2	2.0	JAK2 mutation in myeloproliferative disorders	2007-2007	dis	2005	2005	2006	0	1	1	28.0	63.5
3	2.5	swine flu pandemic	2009-2010	dis/exo	2009	2009	2009	0	0	0	17.0	32.1
4	3.0	drug treatment of type-2 diabetes	2007-2007	dis	2004	2004	2007	0	3	3	26.0	38.3
5	4.0	heart failure and rosiglitazone	2007-2007	dis/exo	2007	2007	2006	0	-1	-1	15.0	47.4
6	4.0	wireless sensor networks	2008-2008	inn/exo	2000	2003	2006	3	6	3	15.0	30.7
7	4.0	graphene nanosheets and nanocomposites	2010-2010	dis	2007	2004	2010	-3	3	6	30.0	52.6
8	5.0	terpene alcohols as fragrance ingredients	2008-2008	dis/exo	2008	2008	2008	0	0	0	6.0	98.5
9	5.0	Horava-Lifshitz gravity	2010-2010	dis	2009	2009	2010	0	1	1	24.0	66.1
10	6.0	human papillomavirus vaccination	2007-2007	dis/exo	2006	2006	2007	0	1	1	19.0	40.5
11	6.0	graphene oxide nanosheets	2010-2010	dis	2008	2004	2010	-4	2	6	22.0	53.5
12	6.5	signal recovery from compressed sensing	2009-2010	inn	2006	2006	2009	0	3	3	24.5	63.2
13	7.0	diabetes type-2 gene TCF7L2	2007-2007	dis	2006	2006	2007	0	1	1	18.0	54.1
14	7.0	social tagging	2007-2010	inn/exo	2004	2006	2007	2	3	1	13.3	33.5
15	7.3	sixty-four slice spiral CT angiography	2007-2009	inn/exo	2005	2005	2007	0	2	2	18.0	40.9
16	8.0	personalized cancer management	2008-2009	dis/exo	2007	2007	2008	0	1	1	17.5	49.3
17	8.0	cloud computing & MapReduce	2010-2010	inn/exo	2007	2008	2010	1	3	2	13.0	34.5
18	8.3	spectrum sensing in cognitive radio	2007-2010	inn	2005	2005	2007	0	2	2	18.5	45.0
19	9.0	human bocavirus	2007-2007	dis	2005	2005	2007	0	2	2	19.0	73.8
20	9.0	predictive markers for colorectal cancer	2008-2008	dis/exo	2004	2004	2008	0	4	4	16.0	50.6

EXOGENOUS INFLUENCES

Type of influence	*# topics*
Government actions	11
Software releases	7
Technical standards	7
Clinical trials	6
Health threats	5
Product introductions	3
Publishing events	1

EVENT DESCRIPTIONS

id	label	event description	application/objective
1	iron-based superconductors	Kamihara, Y., Watanabe T., Hirano, M. & Hosono, H. (2008). Iron-based layered superconductor La[O1-xFx]FeAs (x= 0.05-0.12) with Tc = 26 K. *Journal of the American Chemical Society*, 130(11), 3296-3297.	new superconducting materials
2	JAK2 mutation in myeloproliferative disorders	Baxter, E. J., Scott, L. M., Campbell, P. J., East, C., et al. (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *The Lancet*, 365(9464), 1054-1061.	drug treatment for myeloproliferative disorders
3	swine flu pandemic	outbreak of pandemic	minimize global health impacts of flu
4	drug treatment of type-2 diabetes	Ahren, B., Landin-Olsson, M., Jansson, P. A., Svensson, et al. (2004). Inhibition of Dipeptidyl Peptidase-4 Reduces Glycemia, Sustains Insulin Levels, and Reduces Glucagon Levels in Type 2 Diabetes. *Journal of Clinical Endocrinology & Metabolism*, 89(5), 2078-2084.	new drug treatments of type-2 diabetes
5	heart failure and rosiglitazone	drug safety alert issued by FDA	safer drugs for diabetes
6	wireless sensor networks	DARPA funding	location of low cost sensors in environment
7	graphene nanosheets and nanocomposites	Stankovich, S., Dikin, D. A., Piner, R. D., Kohlhaas, K. A., et al. (2007). Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. *Carbon*, 45(7), 1558-1565.	new nanocomposite materials and devices
8	terpene alcohols as fragrance ingredients	single issue of a journal	safety of a fragrance material
9	Horava-Lifshitz gravity	Horava, P. (2009). Quantum gravity at a Lifshitz point. *Physical Review D*, 79(8), art num: 084008.	develop a new theory of quantum gravity

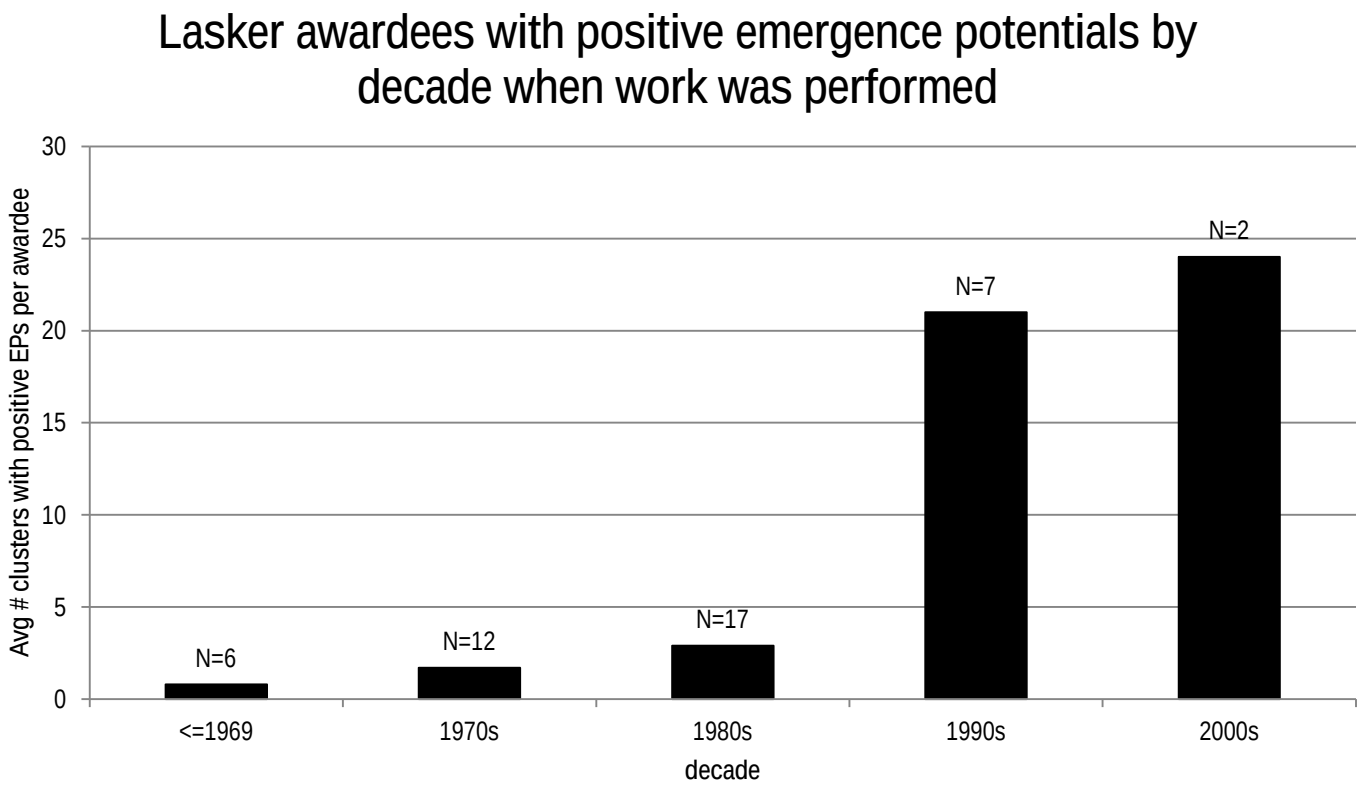
OBSERVATIONS

- Average time lag of 2.3 years from most highly cited paper or exogenous event to year if first emergence
- H-index higher for pure discovery / innovation cases than for those with exogenous co-factors
- Consensus within cluster on the most cited paper was higher if a specific paper caused emergence
- Validation: No way to prove this is the “best” set of emergent topics – there is no generally accepted list of emerging topics. Indirect approach: Look for evidence that topics and authors are significant.

EVIDENCE

seq	person/topic	Award	year
1	Hideo Hosono	Bernd T. Matthias Prize	2009
1	iron-based superconductivity	Science: breakthrough of the year runner-up	2008
3	swine flu virus	Science: virus of the year	2009
6	John Stankovic	IEEE Distributed Processing Annual Distinguished Achievement Award	2006
6	Tarek Abdelzaher	IEEE Outstanding Technical Achievement and Leadership Awards	2012
10	HPV	Nobel Prize to Haralid Zur Hausen	2008
12	David Dohono	IEEE Information Theory Society Paper Award	2008
13	David Altshuler	American Diabetes Association Outstanding Scientific Achievement Award	2012
18	Joseph Mitola	IEEE TCCN Recognition Award	2011
23	Metamaterials	Science: Insights of the decade	2010
23	John Pendry	Willis E. Lamb Award for Laser Science and Quantum Optics	2010
23	John Pendry	UNESCO-Niels Bohr gold medal	2009
28	Cathy Brown	American Association of Veterinary Laboratory Diagnosticians Best paper award	2008
31	Shinya Yamanaka	Nobel Prize	2012
31	cell reprogramming	Science: Insights of the decade	2010
31	Shinya Yamanaka	Breakthrough Prize in Life Sciences	2013
34	Ian Akyildiz	IEEE W. Wallace McDowell Award	2011
35	George Sheldrick	Gregori Aminoff Prize	2009
38	Rudolf Ahlswede	IEEE Claude Shannon award	2006
40	Paul G. Richardson	Warren Alpert Foundation Prize	2012
41	IDH1 & IDH2 mutations in cancer	Science: breakthrough of the year runner-up	2008

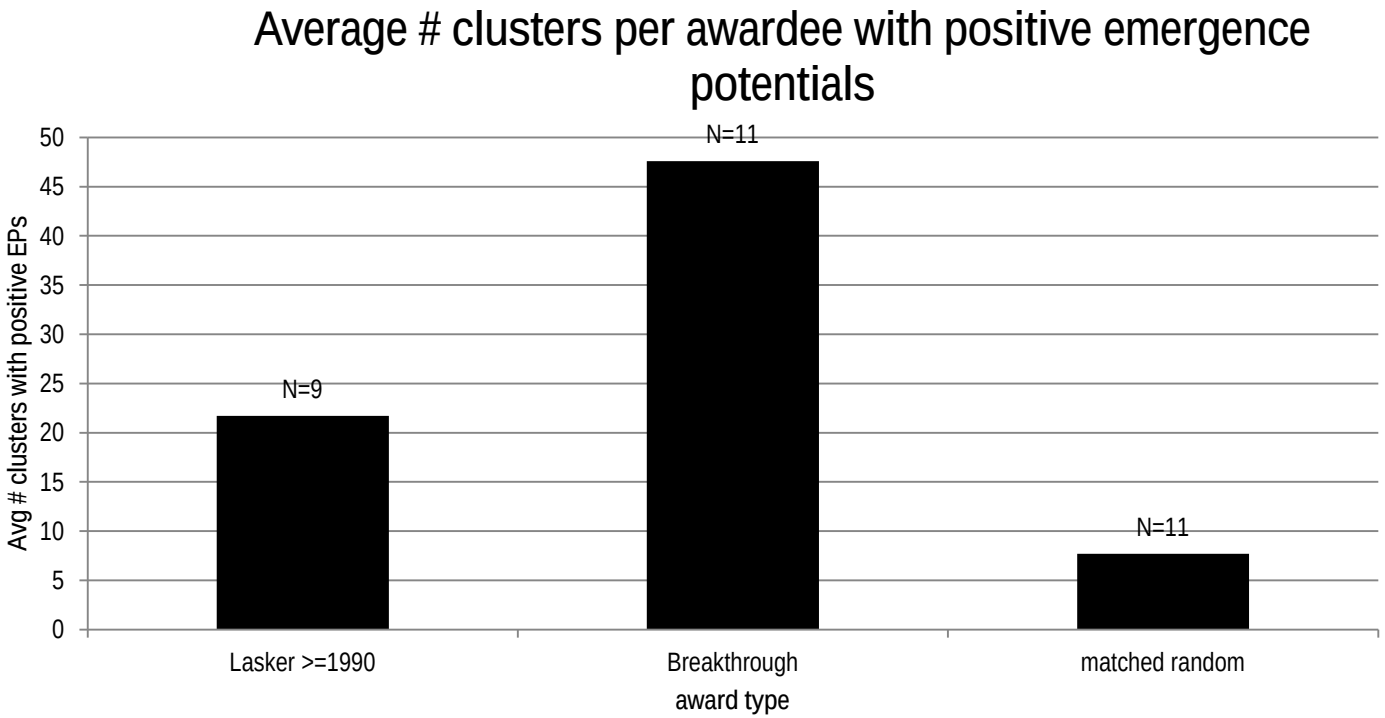
EP and AWARD RECENCY



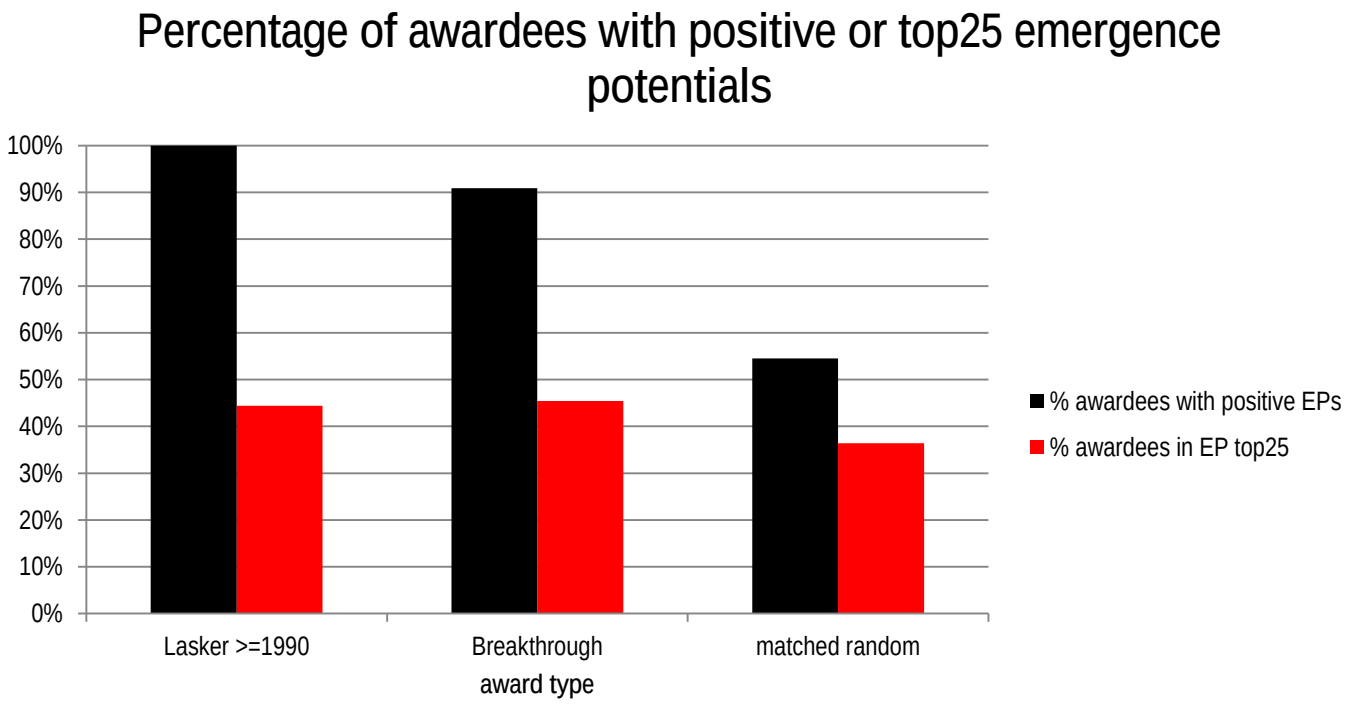
GOOGLE AWARDS

- Analysis of Life Sciences Breakthrough Prizes, the new award from the Google folks
  - 11 winners so far – the award is relatively current in focus
  - 10 out of 11 awardees matched with dc clusters having positive emergence potentials, and 5 of 11 of these were in the top 25 for years 2007-2010.
  - This compares to 25 out of 51 with positive EPs for Lasker winners, and 4 cases of 51 in the top 25.
  - There were 3 individuals who received both awards. They all got positive scores and two were in the top 25.

COMPARISON WITH CONTROL



COMPARISON WITH CONTROL



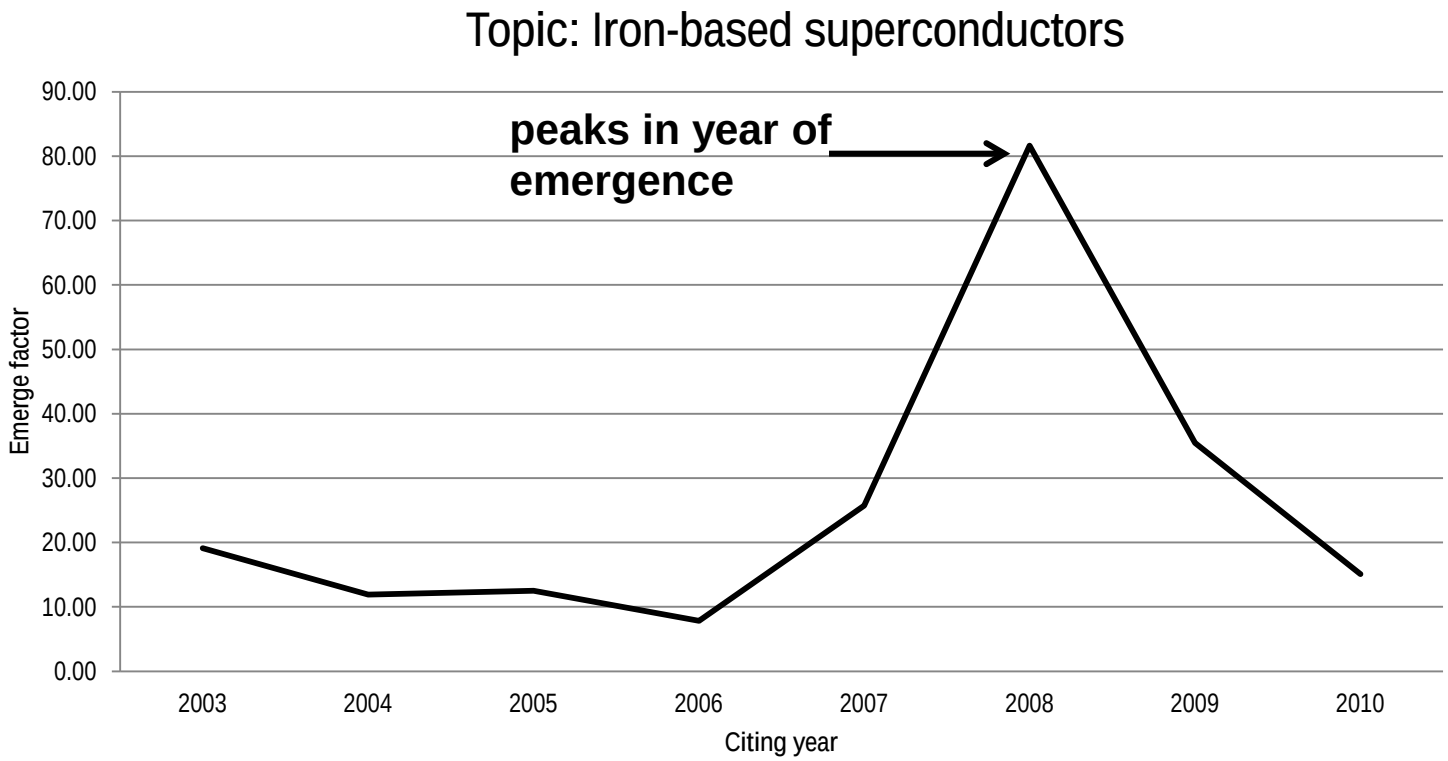
EMERGE FACTOR

- Topic emergence is often associated with discoveries or findings which are represented by highly cited papers
- These papers tend to be “new” or have low age at the time of topic emergence
- The papers also tend to continue to be cited and increase in citation in subsequent years, but are not cited in previous years
- The Emerge Factor formula takes these factors into account

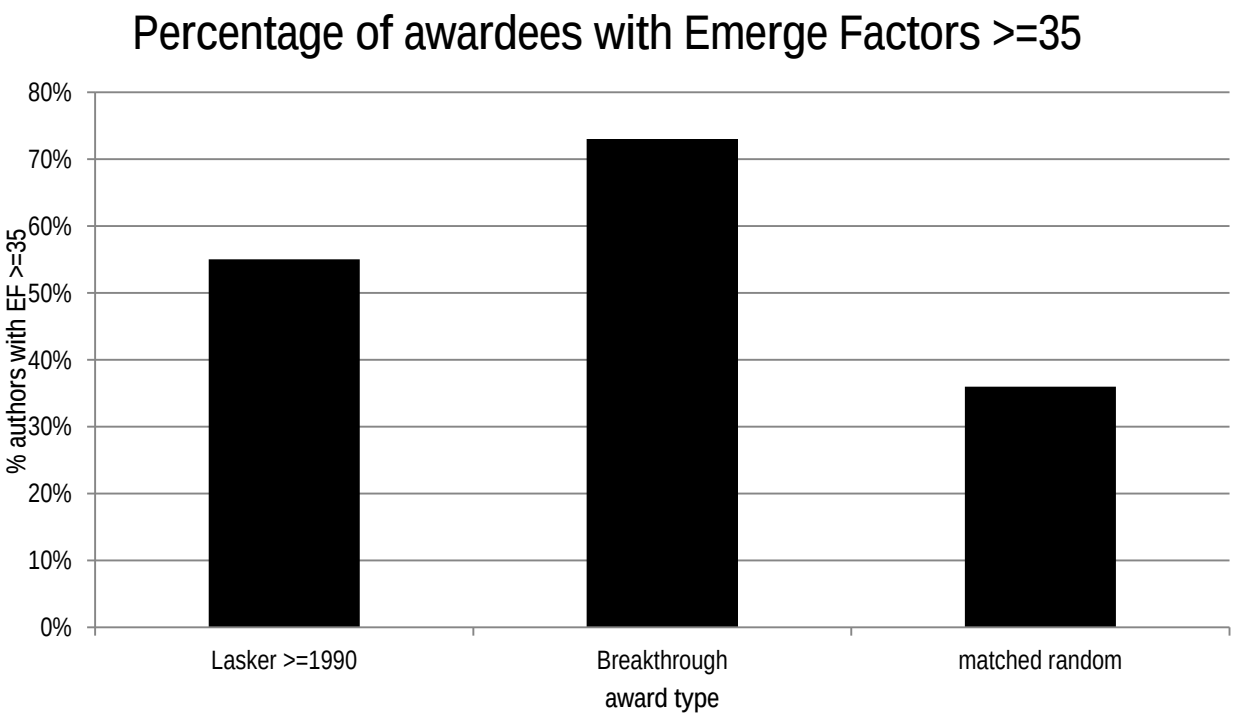
EMERGE FACTOR

$$EF = \frac{\sum((cites\_in\_year + cites\_in\_next\_year - cites\_in\_prior\_year)/(age + 1)^2)}{\sum(cites\_in\_year + cites\_in\_next\_year)}$$

Sum over all cited documents that continue from year to year



COMPARISON WITH CONTROL



CONCLUSIONS

- Recent awardees have a higher representation in emerging topics as measured by the Emergence Potential and the Emerge Factor than a matched sample of authors in the same topics and citedness range
- About half of the awardees participated in emerging topics
- Other factors that may contribute to whether an author participates in an emerging topic are the number of highly cited papers produced, and the number of clusters or topics an author is associated with
- Thus, reverse validation does not provide conclusive or statistically significant evidence that our emerging topic protocol works, but does not discount it either. The anecdotal strength of the results increases our confidence in the method.

QUESTIONS

Thank-you for your attention !