

2018-2019 Contest & Conference Track

Measuring Tech Emergence

www.vpinstitute.org/wordpress/academic-portal/tech-emergence-contest/

Detecting and forecasting emerging R&D topics can help focus research policy, program management, and individual research agendas. The VPInstitute is sponsoring a **Measuring Tech Emergence “Contest”** and paper track at the 9th Global TechMining (GTM2019) “*Conference*” to stimulate thinking about emergence and tech mining [text analyses of Science & Technology (S&T) information resources] to generate novel and viable indicators.¹

Who may participate? The “contest” is open to individuals or teams associated with academic, non-profit, or governmental organizations, and eligible to attend the conference in Atlanta, GA on October 17, 2019 (i.e., if you are not from the U.S., you should be eligible for a visa). We especially encourage student and/or academic research groups. A submission to the *Measuring Tech Emergence* track at the GTM2019 “conference” is open to anyone (i.e. contest and non-contest participants).

Key dates:

- **October-December 2018** – [CONTEST PRE-REGISTRATION](#). Submit your intent to participate by completing the on-line pre-registration form with name(s) of likely participant(s) as well as your initial ideas and interests. Pre-registrants will proceed to contest registration in January by completing “terms of agreement” and, thereby, gaining access to the practice datasets.
- **January 2019** – CONTEST REGISTRATION and DISTRIBUTION OF PRACTICE DATASETS. Three practice WoS datasets will be provided (free) to each registrant (official registration form will be sent to each pre-registrant). Each dataset will contain WoS search results on a given domain for 12 years (e.g., ~10,000 abstract records on synthetic biology to give you 10 years of data to analyze so as to best predict activity in the following 2 years).
- **April 2019** – CONTEST. We will provide the contest dataset, requiring you to send back your list of emerging topics in 10 days, with a brief description of your algorithm/process.
- **February – April 2019** – CONFERENCE SUBMISSION PERIOD. GTM2019 will be accepting abstract submissions for the *Measuring Tech Emergence* track.
- **October 17, 2019** – CONTEST AWARD. The winner and second prize will be awarded at the 9th Global Tech Mining (GTM2019) Conference in Atlanta, GA.

Overview

Many exciting approaches are blossoming to tackle facets of tech emergence and related concepts. To give the flavor, two conferences held in Leiden, The Netherlands, in September, 2018, touched on many pertinent aspects, including:

- [Global Tech Mining \(GTM\) Conference](#) – theoretical roots of emergence (Burmaoglu et al.), visualizing cross-topic keyword aggregation (Li), detecting hotspots (Li et al.), technological convergence as antecedent of speciation (Caferoglu and Moehrle), science-technology interactions (Winnink; Qi et al.), determining technology fronts (Garechana et al.), network

¹ This work is supported by the US National Science Foundation (Award #1759960 – “Indicators of Technological Emergence”). The findings and observations contained herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

analyses re: technological evolution (Boelman et al.), crowdfunding text mining of consumer oriented innovations (Boye et al.), and evaluating tech emergence (Burmaoglu and Saritas)²

- [Science & Technology Indicators \(STI\) Conference](#) – measuring scientific novelty, as tagged by initial journal co-citations (Mairesse and Pezzoni; Carayol et al.); applying stochastic citation time series analyses to measure topical emergence (Förster et al.); and emerging technology forecasting (Garechana et al.)³

Measuring tech emergence touches on choices concerning:

- what scale?
- what data?
- what analytical approaches?
- what outputs, toward what ends?

An essential scale choice presents between “macro” – i.e., which research domains are ascending? -- and “micro” -- within a domain, what topics are accelerating? Data possibilities range widely – e.g., R&D publication abstracts, patent full texts, altmetrics, combining with auxiliary sources (Wikipedia), etc. Approaches include various types of text mining, bibliographic coupling (of cited topics, journals, fields, authors), social network analyses, etc. Choice of approach affects what parts of the data one treats (various topical text fields, keywords, authors/affiliations, journals, citations). End targets diverge too -- one could imagine measuring emergence to study scientific evolution; offer individual science or technology or innovation indicators; contribute to composite indicators (e.g., dashboards) for science policy or technology management; provide competitive technical intelligence by identifying key players in particular frontier topics; and so on. These measures share a common thread of empirical analyses, but also offer potential value in tapping experts’ knowledge.

We see rich synergies among analyses treating these different aspects relating to tech emergence. BUT, we also confront practicalities in arranging a fun, learning experience without undue burden on participants or contest administrators. We thus adopt this 2-part approach:

- 1) A tightly delimited “contest” (some might call it a game)
- 2) A “Measuring Tech Emergence” track at the 9th Global TechMining Conference consisting of “contest” and “non-contest” (i.e., approaches outside contest constraints) submissions.

The Contest

This Contest challenges you to *devise a repeatable procedure to identify emerging R&D topics within a designated S&T domain* (e.g., “synthetic biology”). Topics can be terms, or term-based themes; they must appear in Web of Science (WoS) abstract records. The data resource to be mined is an R&D publication dataset that will be provided for you, on a designated science or technology domain, drawn from WoS. A key criterion is: *who best predicts topics that are notably active in the following two years of research?*

Contest Parameters:

- Scale: Focus within a given science/technology domain
- Data: We provide WoS abstract records for an S&T domain, including abstracts, keywords, funding acknowledgements, times cited counts, and cited references. The query for the test dataset search will not be provided to you.
- Analytical approaches: Open -- but recognize that you must submit explicit emergent terms or topics. By “topics,” we mean sub-technologies or other subject matter (methods, concepts, applications, etc.) addressed within the research abstract records.

² See: <http://www.vpinstitute.org/wordpress/global-techmining-conference/2018-global-techmining-conference/>

³ Available at: <http://sti2018.cwts.nl/proceedings>

[Pursuing document-based clustering to get at emergence would likely NOT generate such terms, so would not work for the contest.]

- Output: 10 (+/-3) emerging terms or up to 10 topics that you predict will be highly active in the subsequent 2 years (i.e., appearance in WoS abstract records) compared to their frequency in the most recent 2 years of the dataset. Rank these from most to less emergent.

Prizes? The Contest Winner will receive up to \$1,500 in travel support as well as complimentary registration to present and receive the award at the 9th GTM Conference to be held October 17, 2019, in Atlanta, GA. Second prize will also receive complimentary conference registration to receive that award.

Interested? Complete the pre-registration form at <http://www.vpinstitute.org/wordpress/academic-portal/tech-emergence-contest/> by December 31, 2018. We will provide a separate registration form with Terms of Agreement for you to sign when we then share the practice datasets.

Contest Details:

1. **DATA:** As noted, we will provide a set of WoS abstract records, including Cited References, on a research area. These will be provided in XML format. Your analytical approach can derive information from any fields contained therein [keywords, single-word and/or multi-word abstract phrases elicited via Natural Language Processing, authors, citations, etc.]. If you opt to augment the records with additional information, keep in mind the contest constraints.

2. **OUTPUTS** from you:

The contest submission will entail three elements.

- a. Description of your analytical approach (suitable for sharing openly; not in detail, such as computer code).
- b. Your result – Ranked, Top 10 (+/-3) emergent terms (ETs), or up to 10 topics, for the test technical domain. If your process generates more ETs, submit the Top 10 or so reflecting a systematic, reproducible selection process.
The ETs need be in a form for which we can readily search in a set of abstract records. If you present topics (e.g., themes, composite factors), you can include up to 10 terms/topic, for which we would search in the abstract record sets. We require discrete terms – single word or multi-word phrases. So if you generate topics not clearly countable in WoS records, submit a short list of accompanying “n-grams” for each topic.
- c. Your identity – to enable blind judging, we’ll separate this from your description and results.

3. **JUDGING:** We will count the # of records in which each of your nominated ETs appear in the last 2 years of the 10-year dataset that we provide you, and in the subsequent 2 years (that we generate for testing). We’ll normalize both counts by dividing by the size of the respective 2-year datasets.

We anticipate some awkwardness in gauging emergence of topics and terms. We know there are issues in term cleaning and consolidation. Therefore, we will constitute a small judging team to consider empirical results (#2b) and augment those with human perspective on what constitutes *meaningful, interesting “emerging” terms/topics potentially valuable in deciding on R&D priorities*. We hope all treat the contest as a fun, learning experience (no appeals on the judging).

Background:

We have been devising one such approach to generate tech emergence indicators. We set up this contest to “reach beyond that box.” Not to constrain your thinking, but to offer one illustrative approach, we note our text analytics process to identify emerging research topics within a science and technology

domain.^{4 5} We extract terms from titles & abstracts and filter them based on 1) novelty, 2) persistence, 3) a research community, and, especially, 4) rapid growth in research activity.

Can you devise a better way? You might treat words and phrases differently, or combine multiple WoS fields' content (e.g., Web of Science Categories with new authors). You might exploit other data attributes like author social networks, breakout citation patterns, and/or funding trends – it's up to you!

The Conference

The 9th Global TechMining Conference (October 17, 2019) aims to ENGAGE cross-disciplinary networks of analysts, software specialists, researchers, policymakers, and managers to ADVANCE the use of textual information in multiple science, technology, and business development fields. We invite submission from both a) contest participants, describing their approach, and b) researchers who, after reading contest constraints, decide their "Measuring Tech Emergence" approach is outside stated boundaries, but advances use of data, tools, and outputs. The conference call for papers will be distributed January 2019. Past conference proceedings can be found at <http://www.vpinstitute.org/wordpress/global-techmining-conference/>

⁴ Drawing on: U.S. Intelligence Advanced Research Projects Activity (IARPA) Program on Foresight and Understanding from Scientific Exposition (FUSE) [<http://www.iarpa.gov/index.php/research-programs/fuse>]; and Rotolo, D. D., Hicks, D., and Martin, B.R. (2015). What is an emerging technology? *Research Policy* 44, 1827-1843.

⁵ Carley, S.F., Newman, N.C., Porter, A.L., and Garner, J. (2018). An indicator of technical emergence, *Scientometrics*, 115 (1), 35-49; <http://link.springer.com/article/10.1007/s11192-018-2654-5>.

Porter, A.L., Garner, J., Carley, S.F., and Newman, N.C. (to appear). Emergence scoring to identify frontier R&D topics and key players, *Technological Forecasting and Social Change*