Trends in Topics at SE Conferences (1993-2013)

George Mathew
Computer Science, NCSU
Raleigh, North Carolina
george.meg91@gmail.com

Amritanshu Agrawal
Computer Science, NCSU
Raleigh, North Carolina
aagrawa8@ncsu.edu

Tim Menzies
Computer Science, NCSU
Raleigh, North Carolina
tim@menzies.us

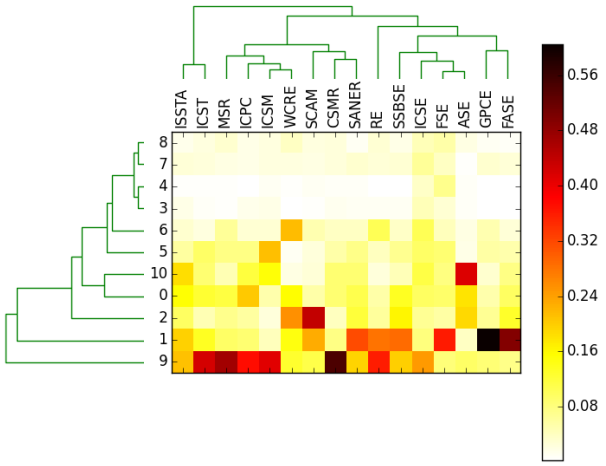

Fig. 1: Figure shows the hierarchical clustering heatmap of Topics and Conferences between the years 2009-2016.

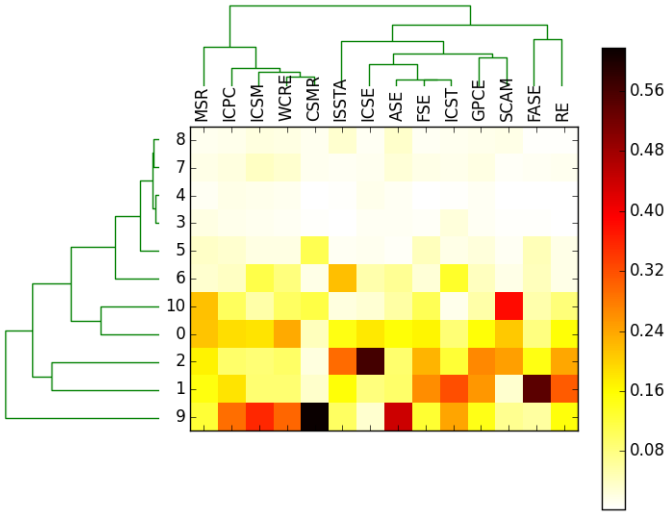

Fig. 2: Hierarchical Clustering Heatmap Results 2001-2008: Topics vs Conferences. Similar to Figure 1.

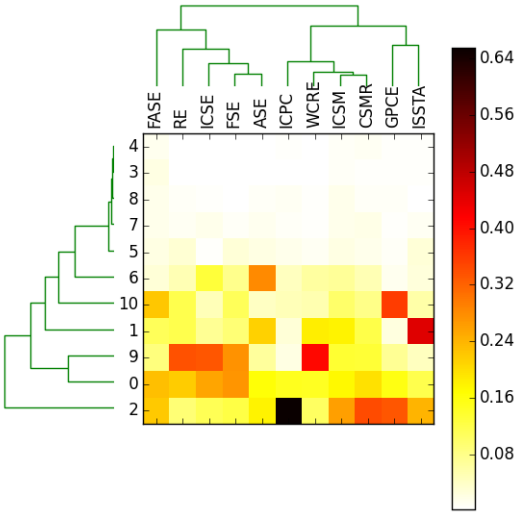

Fig. 3: Hierarchical Clustering Heatmap Results 1993-2000: Topics vs Conferences. Similar to Figure 1.

***Abstract*—We report a topic modeling analysis of the the abstracts and titles from 9291 papers published in 11 top-ranked SE conferences between 1993 to 2013. Seven topics are identified as the dominant themes in modern software engineering. We show that these topics are not static; rather, some of them are becoming decidedly less prominent over time (modeling) while others are become very prominent indeed (defect analysis).**

Also, by clustering conferences according to the topics they usually publish, we can see that SE conferences fall into four large groups. For example, for the last 20 years, ASE, FSE and ICSE have been publishing mostly the same work (exceptions: FSE publishes somewhat more work on program analysis compared to ASE or ICSE).

Our results highlight the enormous impact of a program committee (PC) on a research conference. Even though PC members comprise less than a quarter of the authors with published papers, the topics of their papers matches almost exactly with the topics seen in all published papers.

Using these results, we offer numerous recommendations including how to plan an individual's research program; how to intelligently make new conferences or merge old ones; how to better recruit program committees; and how to encourage a broader range of topics at SE conferences.

***Keywords*-Software Engineering; Bibliometrics; Topic Modelling; Text Mining**

I. Introduction

Short	Name	Start	Group
MSR	Working Conference on Mining Software Repositories	2004	A
SANER	IEEE International Conference on Software Analysis, Evolution and Re-engineering	2014	A
WCRE	Working Conference on Reverse Engineering	1995	B
ICPC	IEEE International Conference on Program Comprehension	1997	B
ICSM	IEEE International Conference on Software Maintenance	1994	B
CSMR	European Conference on Software Maintenance and Re-engineering	1997	B
FASE	International Conference on Fundamental Approaches to Software Engineering	1998	C
RE	IEEE International Requirements Engineering Conference	1993	C
SSBSE	International Symposium on Search Based Software Engineering	2011	D
ISSTA	International Symposium on Software Testing and Analysis	1989	D
ICST	IEEE International Conference on Software Testing, Verification and Validation	2008	D
SCAM	International Working Conference on Source Code Analysis & Manipulation	2001	E
GPCE	Generative Programming and Component Engineering	2000	E
ICSE	International Conference on Software Engineering	1994	F
FSE	ACM SIGSOFT Symposium on the Foundations of Software Engineering	1993	F
ASE	IEEE/ACM International Conference on Automated Software Engineering	1994	F

TABLE I: Corpus of conferences studied in this paper. The "Group" column shows conferences that publish "very similar" topics (where "very similar" is computed via a cluster analysis shown later in this paper).

#	Top 7 terms	Name
0	systems, architecture, development, paper, applications, component, tools	Design
1	test, testing, program, analysis, programs, execution, techniques	Testing
2	model, requirements, approach, specification, design, language, formal	Modelling
3	workshop, android, international, apps, conference, app, reviews	Mobile
4	energy, consumption, objectoriented, spreadsheet, poster, green, power	Energy
5	bug, fault, models, defect, prediction, defects, localization	Defects
6	code, language, programming, security, java, program, type	SourceCode
7	web, api, application, gui, user, client, interfaces	Web Apps
8	product, kernel, line, feature, configuration, variability, products	Configuration
9	code, source, developers, information, changes, results, study	Developer
10	repositories, project, study, research, methods, opensource, case	Mining

TABLE II: The top 7 terms in each topic of SE as seen in our corpus.

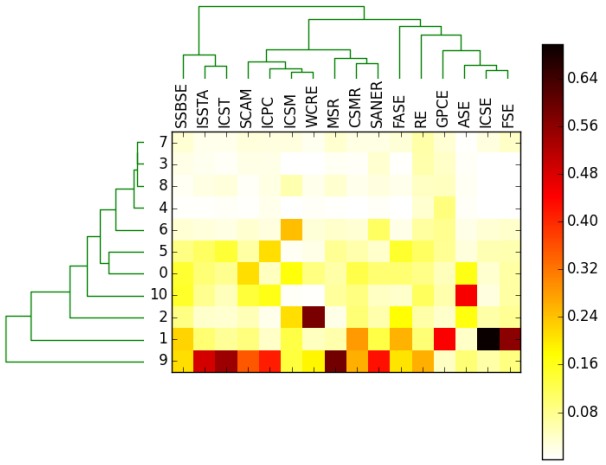

Fig. 4: Hierarchical Clustering Heatmap showing association between Topics and Conferences from the Program Committee members between 2009-2016

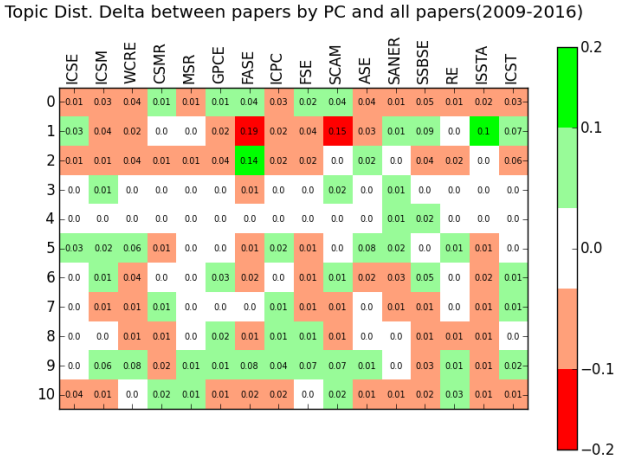

Fig. 5: Delta between the heatmaps of Figure 4 and Figure 1 showing association between Topics and Conferences from the Program Committee members between 2009-2016

Topic	Top Papers
Design	2005: CatchUp!: capturing and replaying refactorings to support API evolution; Henkel, J. and Diwan, A. 1995: A Component- and Message-Based Architectural Style for GUI Software; Taylor, R.N., Medvidovic, N., Anderson, K.M., Whitehead, E.J., Robbins, J.E., Nies, K.A., Oreizy, P. and Dubrow, D.L.
Testing	2005: CUTE: a concolic unit testing engine for C.; Sen, K., Marinov, D. and Agha, G. 2001: Dynamically discovering likely program invariants to support program evolution; Ernst, M.D., Cockrell, J., Griswold, W.G. and Notkin, D.
Modelling	2003: Model checking programs; Visser, W., Havelund, K., Brat, G., Park, S. and Lerda, F. 2002: Automatic extraction of object-oriented component interfaces; Whaley, J., Martin, M.C. and Lam, M.S.
Mobile	2009: Identification of extract method refactoring opportunities; Tsantalis, N. and Chatzigeorgiou, A. 2013: Asking for (and about) Permissions Used by Android Apps; Stevens, R., Ganz, J., Filkov, V., Devanbu, P. and Chen, H.
Energy	2004: Validating the unit correctness of spreadsheet programs; Antoniu, T., Steckler, P.A., Krishnamurthi, S., Neuwirth, E. and Felleisen, M. 2015: Witness validation and stepwise testification across software verifiers; Beyer, D., Dangl, M., Dietsch, D., Heizmann, M. and Stahlbauer, A.
Defects	2006: Who should fix this bug?;Anvik, J., Hiew, L. and Murphy, G.C. 2005: Use of relative code churn measures to predict system defect density.; Nagappan, N. and Ball, T.
Source Code	2001: Identifying similar code with program dependence graphs; Krinke, J. 2004: Static checking of dynamically generated queries in database applications; Gould, C., Su, Z. and Devanbu, P.
Web Apps	2008: Static detection of cross-site scripting vulnerabilities; Wassermann, G. and Su, Z. 2011: Auto-locating and fix-propagating for HTML validation errors to PHP server-side code; Nguyen, H.V., Nguyen, H.A., Nguyen, T.T. and Nguyen, T.N.
Configuration	2006: What's in a Name? A Study of Identifiers; Lawrie, D., Morrell, C., Feild, H. and Binkley, D. 2008: Configuration lifting: Verification meets software configuration.; Post, H. and Sinz, C.
Developer	2008: When do changes induce fixes?; liwerski, J., Zimmermann, T. and Zeller, A. 2005: PR-Miner: automatically extracting implicit programming rules and detecting violations in large software code.; Li, Z. and Zhou, Y.
Mining	2006: Mining email social networks.; Bird, C., Gourley, A., Devanbu, P., Gertz, M. and Swaminathan, A. 2010: Codebook: discovering and exploiting relationships in software repositories.; Begel, A., Khoo, Y.P. and Zimmermann, T.

TABLE III: Most cited papers within the seven SE topics seen within our corpus 1993-2016

Topic	Top Papers
Design	2009: Model evolution by run-time parameter adaptation.;Epifani, I., Ghezzi, C., Mirandola, R. and Tamburrelli, G. 2009: Relax: Incorporating uncertainty into the specification of self-adaptive systems.; Whittle, J., Sawyer, P., Bencomo, N., Cheng, B.H. and Bruel, J.M.
Testing	2009: Weimer, W., Nguyen, T., Le Goues, C. and Forrest, S., 2009, May. Automatically finding patches using genetic programming. 2011: Are automated debugging techniques actually helping programmers?; Parnin, C. and Orso, A.
Modelling	2009: Learning operational requirements from goal models.; Alrajeh, D., Kramer, J., Russo, A. and Uchitel, S. 2009: FEATUREHOUSE: Language-independent, automated software composition.; Apel, S., Kastner, C. and Lengauer, C.
Mobile	2009: Identification of extract method refactoring opportunities; Tsantalis, N. and Chatzigeorgiou, A. 2013: Asking for (and about) Permissions Used by Android Apps; Stevens, R., Ganz, J., Filkov, V., Devanbu, P. and Chen, H.
Energy	2015: Witness validation and stepwise testification across software verifiers; Beyer, D., Dangl, M., Dietsch, D., Heizmann, M. and Stahlbauer, A. 2009: What's hot and what's not: Windowed developer topic analysis; Hindle, A., Godfrey, M.W. and Holt, R.C.
Defects	2009: Fair and balanced?: bias in bug-fix datasets; Bird, C., Bachmann, A., Aune, E., Duffy, J., Bernstein, A., Filkov, V. and Devanbu, P. 2011: How do fixes become bugs?; Yin, Z., Yuan, D., Zhou, Y., Pasupathy, S. and Bairavasundaram, L.
Source Code	2010: Static checking of dynamically generated queries in database applications; Gould, C., Su, Z. and Devanbu, P. 2012: On the naturalness of software.;Hindle, A., Barr, E.T., Su, Z., Gabel, M. and Devanbu, P.
Web Apps	2011: Auto-locating and fix-propagating for HTML validation errors to PHP server-side code; Nguyen, H.V., Nguyen, H.A., Nguyen, T.T. and Nguyen, T.N. 2009: Locating need-to-translate constant strings for software internationalization.; Wang, X., Zhang, L., Xie, T., Mei, H. and Sun, J.
Configuration	2010: Highly configurable and extensible code clone detection.; Biegel, B. and Diehl, S. 2011: Decomposing feature models: language, environment, and applications.; Acher, M., Collet, P., Lahire, P. and France, R.B.
Developer	2009: Semantics-based code search; Reiss, S.P. 2011: Portfolio: finding relevant functions and their usage; McMillan, C., Grechanik, M., Poshyvanyk, D., Xie, Q. and Fu, C.
Mining	2010: Codebook: discovering and exploiting relationships in software repositories.; Begel, A., Khoo, Y.P. and Zimmermann, T. 2009: Does distributed development affect software quality?: an empirical case study of Windows Vista.; Bird, C., Nagappan, N., Devanbu, P., Gall, H. and Murphy, B.

TABLE IV: Most cited papers within the seven SE topics seen within our corpus 2009-2016

Years	Authors
1993-2016	A. Zeller, M. D. Ernst, M.J.Harrold, T. Zimmermann, T. Xie, D. Marinov, A. Orso, P. T. Devanbu, N. Nagappan, S. Kim, G. C. Murphy, K. Sen, D. Notkin, S. Khurshid, C. Bird, Z. Su, G. Rothermal C. S. Pasareanu, W. Visser, A. E. Hassan, J. A. Jones, D. Lo, H. C. Gall, M. B. Dwyer, T. A. Ball
2009-2016	T. Xie, P. T. Devanbu, D. Lo, A. Zeller, C. Bird, A. E. Hassan, M. D. Ernst, A. Orso, T. N. Nguyen, S. Kim, D. Poshyvanyk, T. Zimmermann, S. Apel, H. A. Nguyen, T. T. Nguyen, M. Kim, N. Nagappan, L. Zhang, C. Kastner, B. Adams, H. Mei, M. Lanza, F. Tip, G. Fraser, N.Tillman

TABLE V: Top cited authors in SE seen within our corpus, at different time intervals. Listed in descending order of citations (so most cited appear first).

Conference	X = # papers accepted	Y = # paper with PC authors	100*Y/X Percent from PC
ICSE	1755	289	16%
ICSM	612	239	29%
WCRE	262	89	34%
CSMR	306	111	26%
MSR	393	171	44%
GPCE	136	31	23%
FASE	227	34	15%
ICPC	226	130	39%
FSE	556	83	15%
SCAM	173	29	17%
ASE	751	139	19%
SANER	231	93	40%
SSBSE	158	56	35%
RE	348	100	29%
ISSTA	264	61	23%
ICST	455	99	22%
all	6963	1751	25%

TABLE VI: Percentage of papers from members of the program committee, sorted in the descending order with respect to percentage. Data from 2009-2016

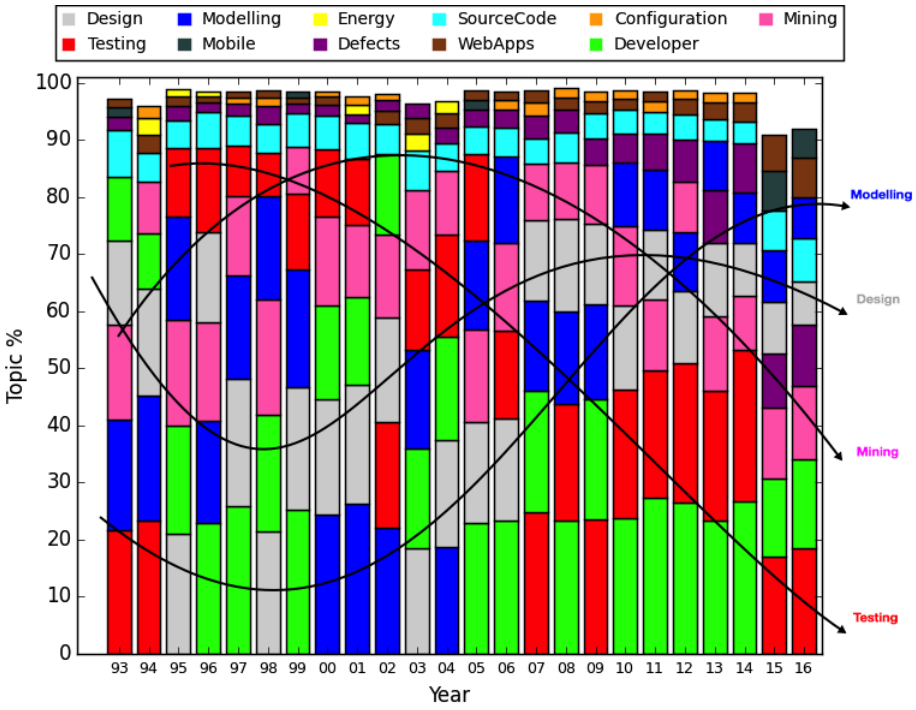

Fig. 6: This figure shows a stacked column chart of the contribution(in %) of each topic in a certain year between 1993-2013. The bars are stacked in ascending order;i.e. the lowest bar is most published that year and the highest bar is the least.