Engenharia de Confiabilidade de Software: Um mapeamento sistemático da última década

† Joicymara Xavier, †Autran Macêdo, †Rivalino Matias, ‡Lúcio Borges
† Faculdade de Computação, ‡ Faculdade de Matemática
Universidade Federal de Uberlândia
Uberlândia, Brasil
joicymara@mestrado.ufu.br, autran@fc.ufu.br, rivalino@fc.ufu.br, lucio@famat.ufu.br

***Resumo*—Este artigo apresenta os resultados de um mapeamento sistemático sobre confiabilidade de software. Este mapeamento considerou os últimos 12 anos de publicações em 92 veículos (revistas e conferências internacionais), totalizando uma amostra de 300 artigos. Foram utilizadas técnicas estatísticas para agrupar artigos similares e identificar tópicos de pesquisa em confiabilidade de software. Os resultados mostram que trabalhos que adotam uma abordagem puramente teórica predominam sobre aqueles que adotam abordagem experimental. Contudo, nos últimos anos, percebe-se que a abordagem teórico-experimental tem se tornado mais acentuada. Além disso, observou-se que o fator "colaboração externa" apresenta-se como um dos mais evidentes nos centros de pesquisa influentes. Finalmente, entre os tópicos de pesquisa abordados neste trabalho, os resultados indicam uma demanda por pesquisas teórico-experimentais em testes de software para os próximos anos.**

Palavras-chave — confiabilidade de software; mapeamento sistemático; agrupamentos

I. INTRODUÇÃO

Os sistemas computacionais estão presentes em praticamente todos os principais processos da sociedade moderna. Esta alta dependência faz com que a falha destes sistemas tenha impacto significativo na vida das pessoas, em especial quando se trata de sistemas críticos. Com base na literatura (ex. [1], [2]), sabe-se que a principal causa de falhas em sistemas computacionais são os componentes de software. Essas falhas são objeto de estudo da Engenharia de Confiabilidade de Software (SRE - *Software Reliability Engineering*) [2].

A SRE é uma área multidisciplinar que integra, principalmente, probabilidade, estatística e computação. São variados os campos de pesquisa em SRE, tais como modelagem de confiabilidade, testes de software, modelagem de crescimento de confiabilidade, análise de tempos de falha, modelos estocásticos e Markovianos, árvore de falhas, dentre outros. Diante de tal diversidade, algumas perguntas podem ser formuladas, por exemplo: Quais campos de pesquisa em SRE têm sido mais investigados recentemente? Quais são os atuais grupos de pesquisa em SRE mais ativos em nível internacional? Quais problemas são abordados pelos trabalhos mais citados em SRE? Quais são os tópicos de pesquisa emergentes em SRE? Encontrar respostas para essas e outras questões correlatas é essencial para pesquisadores que desejam direcionar melhor seus esforços nessa área de pesquisa.

Surpreendentemente, ao procurar trabalhos que respondessem essas perguntas, apenas um artigo recente foi encontrado [4]. Esse artigo apresenta uma revisão sistemática de 141 trabalhos publicados em 34 revistas. Sabe-se que em SRE muitas pesquisas de qualidade são publicadas não somente em revistas, mas também em conferências especializadas de qualidade, as quais não foram consideradas em [4]. Desse modo, a quantidade limitada de estudos recentes discutindo as tendências em SRE nos motivou a realizar este trabalho. Neste artigo, são apresentados os resultados obtidos com a realização de um mapeamento sistemático [5] em SRE, cobrindo 300 artigos e 92 veículos internacionais de publicação nos últimos doze anos. Técnicas estatísticas para agrupar os trabalhos similares no conjunto de dados foram utilizadas. O restante desse artigo está organizado como segue. Na seção II são apresentados os materiais e métodos usados nesse trabalho. Na seção III são discutidos os resultados e na Seção IV são apresentadas as conclusões.

II. METODOLOGIA

A. Material

O período de amostragem considerou artigos publicados de 2002 até 2013. Pesquisou-se pelos termos "*software reliability*" nos metadados de artigos das seguintes bibliotecas digitais: ACM Digital Library [6], IEEE Explore [7] e Science Direct [8]. Inicialmente, foram coletados 535 artigos. Analisando o título e *abstract* desses artigos, foram excluídos aqueles que não tinham plena aderência com áreas da SRE. Assim, a amostra resultante para este estudo foi composta por 300 artigos de 92 veículos internacionais de publicação, entre revistas e conferências. A lista completa desses veículos é apresentada em [9].

B. Método

A SRE envolve diversos tópicos de pesquisa distintos, resultado da multidisciplinaridade existente nessa área. Diante disso, julgou-se necessário agrupar os artigos conforme a similaridade entre eles. Para isso, o método Ward [10] foi utilizado, a fim de encontrar agrupamentos de artigos que tratam de temas semelhantes. Esse método segue uma abordagem iterativa composta de dois passos principais: primeiramente, calcula-se a distância Euclidiana, *d*, entre os artigos, e, em seguida, eles são agrupados de acordo com essa distância. Esses dois passos são repetidos até que se obtenha um único agrupamento. A Fig. 1, denominada Dendograma, apresenta o resultado da aplicação do método Ward aos artigos da amostra analisada. O eixo *x* representa os 300 artigos e o eixo *y* a similaridade entre eles, sendo que quanto maior o valor no eixo *y*, menor é a semelhança entre os artigos.

O tema de um artigo pode ser indicado por suas palavras-chave. Porém, diferentes palavras-chave podem estar associadas a um mesmo tema. Por exemplo, "detecção de faltas", "minimização de faltas", "correção de faltas" e "tolerância à faltas" são termos que podem ser representados por um único termo mais genérico, tal como "faltas de software". Então, as 76 palavras-chave coletadas dos artigos da

amostra foram analisadas quanto à sua similaridade. Essa análise resultou na definição de 40 termos genéricos diferentes.

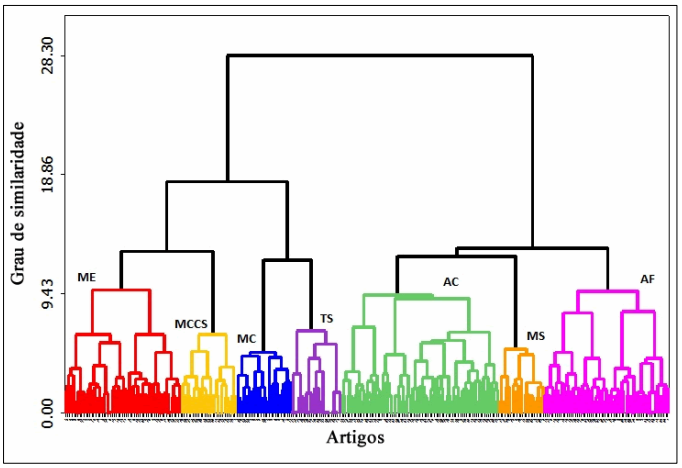

Fig. 1. Dendograma para os 300 trabalhos selecionados.

Como um termo genérico compreende várias palavras-chave, então esse termo foi nomeado como "grupo de palavras-chave" (*keyword group* – KG). Em seguida, cada artigo foi associado a um ou mais KGs. Um artigo é associado ao KG_t, tal que $1 \leq t \leq 40$, se pelo menos uma palavra-chave do artigo for uma daquelas que compõe o KG_t. Assim, o cálculo da distância Euclidiana foi realizado tendo como objetos os KGs. A distância Euclidiana entre os artigos *i* e *m* é dada por (1):

$$d(i,m) = \sqrt{\sum_{j=1}^{p}(X_{ij} - X_{mj})^2}\,, \qquad (1)$$

na qual *p* é o número de palavras-chave presentes em ambos os artigos; X_{ij} representa a presença ou ausência da palavra-chave *j* no artigo *i*; e X_{mj} representa a presença ou ausência da palavra-chave *j* no artigo *m*. X_{ij} e X_{mj} são valores discretos, 0 (ausência) ou 1 (presença). O algoritmo Ward foi adotado para agrupar os artigos da forma mais homogênea possível. A métrica de homogeneidade utilizada é baseada na distribuição da soma dos quadrados totais (*SST*) da técnica Análise de Variância (ANOVA) [9]. Essa distribuição é obtida a partir do cálculo de homogeneidade dos artigos que compõem um KG, e de heterogeneidade de um KG específico com relação aos demais KGs. O resultado dessa distribuição é armazenado em um vetor *X* de *t* posições; uma posição representa um KG. Por exemplo, (2) apresenta o cálculo do *SST* para o KG representado pela posição 1 de *X* (expresso por X_1).

$$SST(X_1) = SSW(X_1) + SSB(X_1), \qquad (2)$$

$$\sum_{j=1}^{k}\sum_{i\in G_q}(X_{i1} - \bar{X}_1)^2 = \sum_{j=1}^{k} n_q(\bar{X}_{j1} - \bar{X}_1)^2 + \sum_{j=1}^{k}\sum_{i\in G_q}(X_{i1} - \bar{X}_{j1})^2,$$

tal que, $SSW(X_1)$ representa o grau de homogeneidade interna de *k* agrupamentos com relação a X_1; $SSB(X_1)$ representa a heterogeneidade entre os agrupamentos com relação a X_1; G_q corresponde ao conjunto de artigos do agrupamento *q*; n_q é o número de elementos de *q*, $\bar{X}_1$é a média de X_1 a cada iteração e $\bar{X}_{q1}$é a média do KG de posição 1 no agrupamento *q* a cada iteração. Desse modo, uma boa partição para a variável X_1 seria aquela que minimizasse $SSW(X_1)$ e maximizasse $SSB(X_1)$. Considerando todas as *p* variáveis (KG_1 até KG_{40}), a soma dos quadrados entre grupos, *SSWP*, é dada por (3):

$$SSWP = \sum_{j=1}^{p} SSW(KG_i), \qquad (3)$$

em que $SSW(KG_i)$ é a soma dos quadrados dentro dos grupos da variável *i*. Por fim, o dendograma (ver Fig. 1) é obtido como saída do algoritmo Ward, dando o arranjo hierárquico final dos agrupamentos. A quantidade de agrupamentos é um parâmetro do método Ward. Foram definidos sete agrupamentos tal qual sugerido em [11]. Em seguida, nomeou-se cada agrupamento de acordo com as palavras-chave mais frequentes. O resultado da nomeação (e as cores correspondentes) foi: vermelho = Modelagem Estocástica (ME); amarelo= Modelagem de Crescimento de Confiabilidade (MCCS); azul = Modelagem de Confiabilidade (MC); roxo = Testes de Software (TS); verde = Avaliação de Confiabilidade (AC); laranja = Manutenção de Software (MS); lilás = Análise de Faltas (AF). A análise detalhada do resultado dos agrupamentos é apresentada na próxima seção.

III. Resultados

A. *Análise Quantitativa*

Os 300 artigos selecionados cobriram 304 autores de 92 afiliações diferentes, distribuídos entre 15 países. A Tabela I mostra a distribuição dos trabalhos entre os agrupamentos. Vale lembrar que os grupos foram nomeados de acordo com as palavras-chave mais frequentes, porém, isto não quer dizer que "Modelagem Estocástica", por exemplo, seja um assunto exclusivo desse grupo e sim que é o assunto nele predominante. Cada artigo é mapeado para apenas um grupo, então, é possível que um trabalho trate, por exemplo, de "Avaliação de Confiabilidade" e também de "Modelagem Estocástica". O método Ward buscará mapeá-lo para um dos dois tópicos, prevalecendo aquele que é mais recorrente no artigo.

Tabela I. Distribuição dos Trabalhos por Agrupamento

Agrupamentos	Qtde. Artigos
Avaliação de Confiabilidade	26%
Análise de Faltas	21%
Modelagem Estocástica	20%
Modelagem de Confiabilidade	9%
Modelagem de Crescimento de Confiabilidade	9%
Testes de Software	8%
Manutenção de Software	7%

Tabela II. Afiliação dos Dez Autores Mais Produtivos

Afiliação	País	NTP	Média
Hiroshima University	JPN	56	6,80
Duke University	USA	49	34,04
National Tsing Hua University	TWN	48	22,02
Hiroshima University	JPN	34	6,79
Chinese University of Hong Kong	HKG	27	33,90
University of Connecticut	USA	25	25,70
Tottori University	JPN	25	9,08
Indian Inst. of Tech. of Kanpur	IND	18	15,44
North Carolina State University	USA	18	46,83
Beihang University	CHN	16	2,00

A Tabela II apresenta a afiliação dos dez autores mais produtivos em termos de quantidade de publicação. A primeira coluna, Afiliação, se refere ao local ao qual o autor está afiliado. A segunda coluna indica o país de origem da afiliação. A terceira coluna, NTP, indica o número de trabalhos publicados pelo autor no período pesquisado (2002-2013), e a quarta coluna apresenta o número médio de citações desses trabalhos. Isto é, esse número corresponde à soma do número de citações dos trabalhos do autor, dividido pelo número de trabalhos publicados por ele. Percebe-se que valores altos de NTP nem sempre estão associados a altos valores de média de citações. A explicação para esse fato pode estar relacionada à "idade" de publicação de alguns artigos (quanto maior a idade de um artigo, maior é a possibilidade de esse artigo ter um maior número de citações). Duas linhas dessa tabela mostram a mesma afiliação (Hiroshima University). Isto significa que há dois autores afiliados à mesma instituição.

A Tabela III apresenta a afiliação dos dez autores mais citados no conjunto de dados analisado, aqui considerados como os mais influentes. A descrição da Tabela III é semelhante à da Tabela II, exceto pela quarta coluna, NC. Esta corresponde à soma das citações de todos os trabalhos publicados pelo autor (coluna NTP), de acordo com o Google Scholar [12]. Por exemplo, a soma das citações para os 49 trabalhos do autor da Duke University é 1668. Sendo assim, essa tabela nos permite identificar quem são os pesquisadores mais influentes em SRE no período pesquisado.

Percebem-se, pelas Tabelas II e III, que alguns dos dez autores mais produtivos não estão entre os dez autores mais influentes. Adicionalmente, pode-se concluir que, considerando a quantidade de trabalhos publicados e citações, a Duke University e a National Tsing Hua University têm tido significativa importância nesta área na última década.

TABELA III. AFILIAÇÃO DOS DEZ AUTORES MAIS CITADOS

Afiliação	País	NTP	NC
Duke University	USA	49	1668
National Tsing Hua University	TWN	48	1057
Chinese University of Hong Kong	HKG	27	916
North Carolina State University	USA	18	843
University of Connecticut	USA	25	643
Microsoft Research	USA	4	436
Duke University	USA	5	436
National Tsing Hua University	TWN	10	423
Hiroshima University	JPN	56	381
Duke University	USA	3	297

B. Tipos de Pesquisa em SRE

Entende-se por tipo de pesquisa a abordagem utilizada pelos autores na realização dos trabalhos. Foram consideradas três abordagens: (*i*) teórica – aquela em que o autor apresenta um corpo conceitual, sem a realização de experimentos; (*ii*) experimental – aquela em que o autor realiza experimentos sobre um corpo conceitual; e (*iii*) teórica-experimental – aquela em que o autor apresenta um corpo conceitual e realiza experimentos para avaliar tal conceito. Observa-se pela Tabela IV que os trabalhos experimentais e teórico-experimentais (coluna Ambos) totalizam 31%. Enquanto os trabalhos puramente teóricos representam 69% dos artigos.

TABELA IV. PORCENTAGEM POR TIPOS DE PESQUISA

Agrupamento	Teórico	Experimental	Ambos
Modelagem Estocástica	15%	1%	3%
Avaliação de Confiabilidade	18%	3%	5%
Modelagem de Confiabilidade	5%	1%	3%
Manutenção de Software	6%	0%	1%
Análise de Faltas	14%	4%	3%
Testes de Software	4%	1%	3%
Modelagem de Crescimento de Confiabilidade	6%	1%	2%

A Tabela V apresenta a porcentagem de cada tipo de abordagem de pesquisa em cada agrupamento. Por exemplo, dos 58 artigos presentes em "Modelagem Estocástica", 79% são teóricos, 7% são experimentais e 14% são teórico-experimentais. Os agrupamentos "Análise de Faltas" e "Testes de Software" congregam a maioria dos trabalhos experimentais. Em contrapartida, o grupo "Manutenção de Software" contém o menor número.

TABELA V. PORCENTAGEM DE TIPOS DE PESQUISA POR AGRUPAMENTO

Agrupamento	Teórico	Experimental	Ambos
Modelagem Estocástica	79%	7%	14%
Avaliação de Confiabilidade	68%	12%	20%
Modelagem de Confiabilidade	55%	15%	30%
Manutenção de Software	86%	5%	9%
Análise de Faltas	68%	18%	14%
Testes de Software	48%	16%	36%
Modelagem de Crescimento de Confiabilidade	70%	8%	22%

A Fig. 2 apresenta a distribuição dos tipos de pesquisa ao longo do tempo. O eixo *x* representa o período de publicação em anos e o eixo *y* a quantidade de artigos. Claramente, trabalhos teóricos prevalecem em todo o período. Observa-se também que, gradativamente, o tipo teórico-experimental (ambos) tem ganhado mais espaço. Em geral, os principais tópicos encontrados nos trabalhos teóricos são relacionados à modelagem de confiabilidade. Esses trabalhos tratam, principalmente, de modelagem de crescimento de confiabilidade e modelos preditivos. Grande parte dos trabalhos experimentais abrangem sistemas embarcados, diagnóstico de faltas, tolerância à faltas, processos de *Poisson* não homogêneos e cobertura de código em testes. Por outro lado, os trabalhos teórico-experimentais foram mais relacionados com processos de *Poisson* não homogêneos e teste de software.

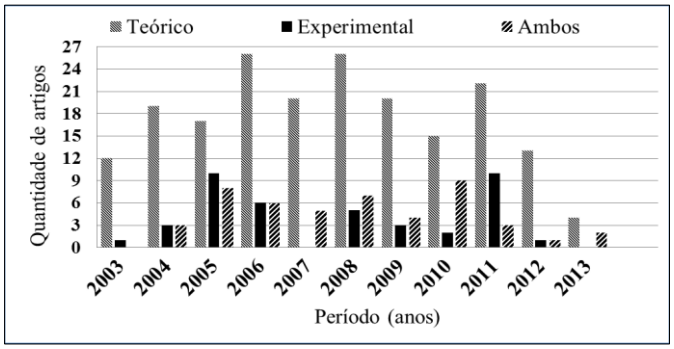

Fig. 2. Número de trabalhos teóricos, experimentais e ambos.

A Fig. 3 mostra as publicações anuais por agrupamento. É possível ver que as pesquisas em "Avaliação de Confiabilidade" tiveram um significativo progresso desde 2002, especialmente durante os últimos três anos. Um ligeiro incremento pode ser observado nas pesquisas em "Testes de Software". Por outro lado, trabalhos em "Modelagem de Confiabilidade" e "Análise de Faltas" têm reduzido gradualmente. Já as publicações em "Modelagem Estocástica" parecem ser as mais consistentes nos últimos anos.

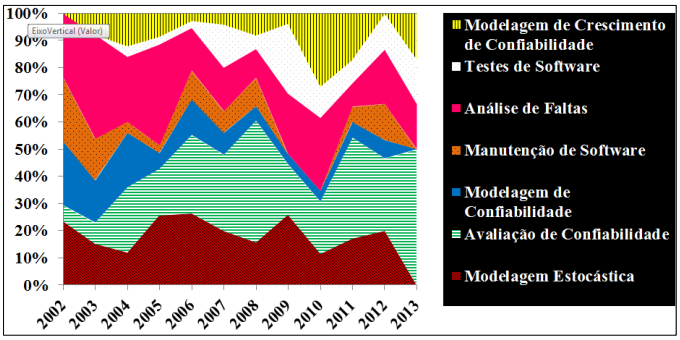

Fig. 3. Publicações anuais por agrupamento.

C. Os Trabalhos Mais Citados por Agrupamento

Foram selecionados os três trabalhos mais citados de cada agrupamento, os quais estão listados na Tabela VI. Para cada agrupamento (primeira coluna), são referenciados os três trabalhos mais citados (segunda coluna), juntamente com seus respectivos números de citação (quarta coluna).

No grupo "Modelagem Estocástica" (ME), os dois primeiros trabalhos são teóricos e o terceiro é teórico-experimental. Em "Avaliação de Confiabilidade" (AC), um trabalho é teórico-experimental [13] e os outros dois [14] e [15] são puramente teóricos. No grupo "Modelagem de Confiabilidade" (MC), o primeiro e o terceiro são trabalhos experimentais e o segundo é teórico. Em "Manutenção de Software" (MS), o primeiro trabalho é teórico-experimental e os outros são somente teóricos. Dos três trabalhos mais citados do agrupamento "Análise de Faltas" (AF), o primeiro e o terceiro são teóricos e o segundo é experimental. No agrupamento "Testes de Software" (TS), o trabalho mais citado é experimental e o segundo e terceiro são teórico-experimentais. No agrupamento "Modelagem de Crescimento de Confiabilidade" (MCCS), o primeiro e o terceiro são ambos teórico-experimental, e o segundo trabalho mais citado é do tipo teórico.

Baseado nos resultados discutidos na Seção III.A, especialmente nos dados sumarizados nas Tabelas II e III, observa-se que as pesquisas em SRE nos Estados Unidos e em Taiwan têm tido importante influência nesta área na última década, considerando ambos os números de trabalhos publicados e citações. Estes dados corroboram os resultados apresentados nesta seção, já que os trabalhos mais citados de cada agrupamento vêm de ambos os países. Sendo que, Estados Unidos se relaciona mais com os grupos ME, AC, MC, MS, AF e TS e Taiwan com MCCS. Adicionalmente, nota-se que um autor da Duke University tem três trabalhos entre os mais citados.

TABELA VI. OS TRÊS TRABALHOS MAIS CITADOS POR AGRUPAMENTO

Grupo	Ref.	Venue	NC
ME	[16]	IEEE Trans. on DSC 2004	284
	[17]	IEEE Trans. on SE 2003	122
	[18]	IEEE Transactions on Reliability 2002	101
AC	[14]	ACM/IEE 27th ICSE 2005	282
	[13]	ACM/IEE 30th ICSE 2008	141
	[15]	IEEE Transactions on DSC 2007	140
MC	[19]	ACM Symp. on STA 2004	104
	[20]	IEEE Transactions on Reliability 2006	68
	[21]	IEEE Transactions on Software 2005	62
MS	[22]	Int'l. Symp. on Empirical SE 2002	121
	[23]	Symp. on Reliab. and Maintain. 2004	54
	[24]	IEEE Int'l.Workshop on Sw. Aging	42
AF	[25]	IEEE Transactions on DSC 2005	140
	[26]	Workshop on ASSISD 2006	115
	[27]	IEEE Int'l.Conf. on ASSAP 2004	58
TS	[28]	ESE Conf. + ACM Symp.on FSE 2011	36
	[29]	IEEE Trans. on SE 2010	22
	[30]	IEEE Transactions on Reliability 2010	20
MCCS	[31]	Journal of Systems and Software 2007	75
	[32]	Journal of Systems and Software 2006	38
	[33]	Journal of Systems and Software 2008	28

D. Rede de Colaboração

Outro achado importante refere-se à rede de colaboração dos dez autores mais citados (ver Tabela III). A Fig. 4 ajuda a identificar onde estão os pesquisadores mais citados (influentes) e com quais instituições eles cooperam. Neste gráfico, cada nó representa um centro de pesquisa (afiliação do pesquisador) e as arestas representam a coautoria nos trabalhos. Os nós são rotulados pelos acrônimos das respectivas afiliações. Foram adotados os acrônimos de nomes conhecidos mundialmente. O significado de todos os acrônimos usados na Fig. 4 pode ser obtido em [9].

Na Fig. 4, os nós sólidos (pretos) representam uma afiliação que contém um ou mais autores influentes, ou seja, estão entre os dez mais citados da Tabela III. O número acima do acrônimo da afiliação representa a quantidade de autores influentes dentro daquela instituição. Além disso, é importante notar que algumas afiliações têm mais de um autor influente, como é o caso de NTHU e DUKE.

Cada nó também possui um número abaixo do seu rótulo, o qual se refere à quantidade de pesquisadores afiliados que publicaram trabalhos em SRE pelo menos uma vez. As arestas representam a colaboração em coautoria nos trabalhos em SRE. Sejam A e B nós, não necessariamente distintos, estes nós são ligados por uma aresta se existir pelo menos uma cooperação (artigos publicados em conjunto) entre eles. Essa aresta é rotulada por três números: (*i*) o número mais próximo de *A*, representa a quantidade de trabalhos que os pesquisadores de *A* têm em coautoria com os autores influentes de *B*; (*ii*) o número no meio representa a quantidade de trabalhos que os autores influentes de *A* têm em coautoria com os autores influentes de *B*; (*iii*) o número mais próximo de *B* representa a quantidade de trabalhos que os pesquisadores de *B* têm em coautoria com um dos autores influentes de *A*. Por exemplo, o número 2, mais próximo de NEC, na aresta NEC-DUKE, significa que há dois

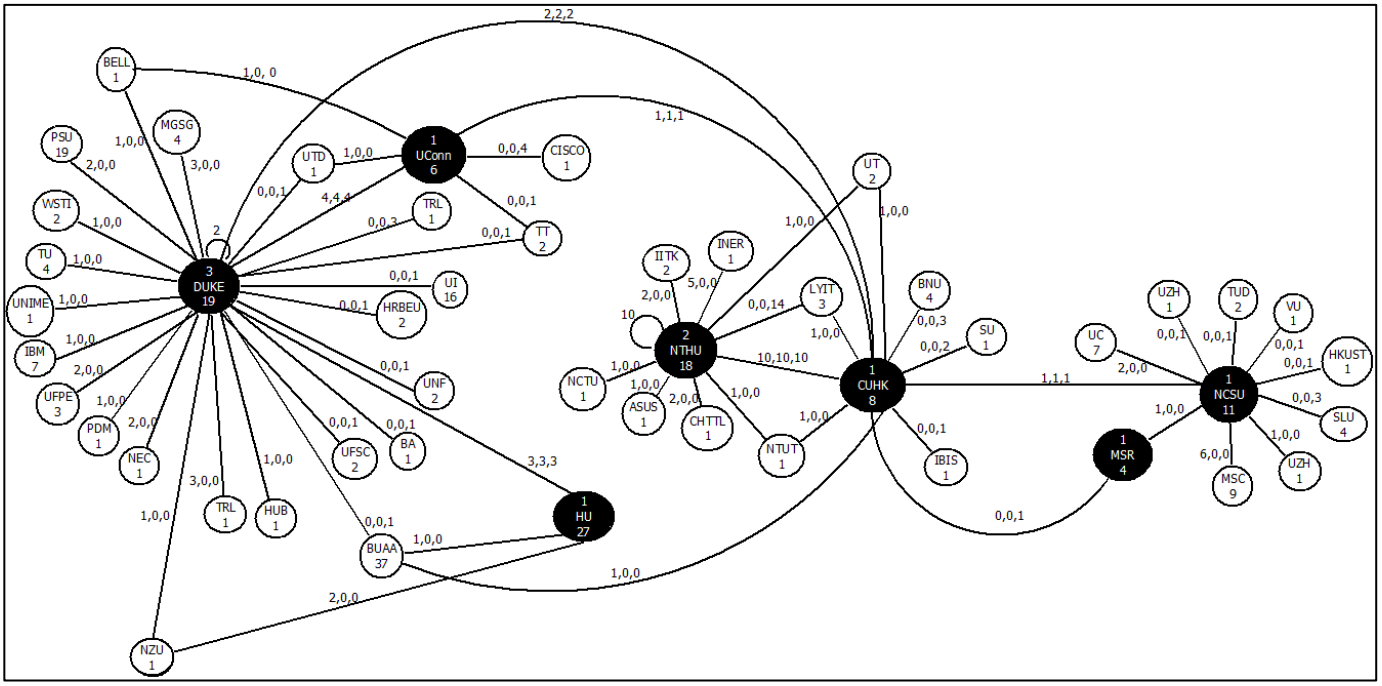

Fig. 4. Rede de colaboração entre os pesquisadores mais citados.

trabalhos em que pesquisadores da NEC colaboraram com um dos três pesquisadores influentes da DUKE. O autolaço em DUKE e NTHU representa a cooperação entre os autores influentes da mesma afiliação. No caso da DUKE, por exemplo, os autores influentes têm coautoria em dois trabalhos.

De acordo com a rede de colaboração apresentada na Fig. 4, observamos que a Duke University (DUKE) é o centro de pesquisa com mais cooperação em SRE, entre as instituições analisadas. Nesse estudo, ele tem 25 cooperações diretas diferentes que se propagam para vários países, como Itália, Japão, Canadá, Brasil e China, além dos Estados Unidos onde reside. Ele também tem colaboração direta com três centros de pesquisa influentes (University of Connecticut - UConn, Hiroshima University – HU e Chinese University of Hong Kong – CUHK). Em contrapartida, existem instituições que possuem um número expressivo de pesquisadores (maior do que 25) em SRE, porém têm um número limitado (menor do que três) de cooperações (ex. BUAA e HU). As evidências deste estudo sugerem que a cooperação entre centros de pesquisa parece ser mais importante do que a quantidade de pesquisadores afiliados.

Entre os centros que possuem pesquisadores influentes (DUKE, UConn, HU, NTHU, CUKH, HU, MSR, NCSU), HU é aquele que possui a menor relação entre o número de pesquisadores influentes (PI) por número de pesquisadores afiliados (PA), a saber, 1/27. Observa-se que HU possui uma rede de colaboração pequena (apenas três), quando comparada contra os demais centros de pesquisa influentes. Com exceção do MSR, os demais centros de pesquisa influentes que possuem uma relação PI/PA melhor que HU, realizam um número significativamente maior de colaboração externa. Outra indicação da importância da colaboração externa é CUHK, que se apresenta mais influente do que centros com grande quantidade de pesquisadores afiliados, tal como BUAA.

IV. Conclusões

Neste trabalho, foram analisados 300 artigos na área de confiabilidade de software, provenientes de diferentes veículos internacionais de publicação. Observou-se que trabalhos puramente teóricos são predominantes, 69%. Trabalhos experimentais e teórico-experimentais, juntos, representam 31% do conjunto de dados analisado, dos quais somente 13,6% são puramente experimentais. Sugere-se a hipótese de que este pequeno número de trabalhos experimentais seja devido a duas razões principais: (*i*) dados experimentais em SRE são muito limitados; (*ii*) produzir dados de confiabilidade de software por meio de experimentos tipicamente requer muito tempo, que, por vezes, não é atrativo para o regime de publicação científica em vigor atualmente. Estas questões poderiam ajudar a explicar o pequeno número de trabalhos experimentais ao longo dos últimos anos, especialmente em campos tais como "Manutenção de Software" e "Modelagem Estocástica".

Especialmente no grupo "Análise de Faltas", observou-se que, desde 2005, os trabalhos relacionados a essa área têm diminuído consistentemente (ver Fig. 3). Uma análise detalhada indicou que autores que costumavam publicar em áreas relacionadas a "Análise de Faltas" têm migrado para outros tópicos de pesquisa (ex. "Modelagem Estocástica").

Baseando-se nos números de pesquisadores, trabalhos, citações, quantidade e qualidade de conexões de colaboração, pode-se concluir que, na comunidade de pesquisa em confiabilidade de software, o fator "colaboração externa" é o que mais se destaca. Parece existir uma forte correlação entre a quantidade e qualidade destas conexões, sendo mais importante que outros fatores como a colaboração interna, número de pesquisadores, ou mesmo, o número de trabalhos publicados.

Por fim, observou-se uma tendência nos últimos anos, em que pesquisadores em SRE estão migrando gradualmente de abordagens puramente teóricas para abordagens teórico-

experimentais. As evidências sugerem que trabalhos de pesquisa envolvendo "Testes de Software" não têm tido significativa atenção da comunidade SRE ao longo da última década, embora esta seja uma importante área para a confiabilidade de software. Portanto, verifica-se como oportunidade de pesquisa para os próximos anos, trabalhos teórico-experimentais relacionados a testes de software.

O presente estudo está sendo estendido em uma pesquisa em andamento, em que será realizada uma ampla revisão sistemática de literatura em SRE. A revisão sistemática utilizará a busca automática e manual [34] para incluir os artigos na amostra, enquanto este mapeamento sistemático utilizou somente a busca automática.

REFERÊNCIAS

[1] Z. Li, L. Tan,X. Wang, S. Lu, Y. Zhou e C. Zhai. "Have things changed now?: an empirical study of bug characteristics in modern open source software". *Proceedings of the 1st workshop on Architectural and system support for improving software dependability*,pp. 25-33. ACM. Outubro, 2006.

[2] M. Sullivan e C. Ram. "Software Defects and their Impact on System Availability: A Study of Field Failures in Operating Systems." *FTCS*. 1991.

[3] J. Radatz, A. Geraci e F. Katki. "IEEE standard glossary of software engineering terminology." *IEEE Std* 610121990 p. 121990, 1990.

[4] A. Singhal e A. Singhal, "A Systematic Review Of Software Reliability Studies". Software Engineering: An International Journal, vol. 1, n. 1, pp. 96-114, Setembro 2011.

[5] P. Kai, F. Robert, S. Mujtaba e M. Michael. "Systematic mapping studies in software engineering". In Proc. of the 12th international conference on Evaluation and Assessment in Software Engineering. British Computer Society, UK, 2008, pp. 68-77.

[6] ACM. ACM Digital Library. ACM, http://dl.acm.org.

[7] IEEE. IEEE Xplore - Home. IEEE, http://ieeexplore.ieee.org/Xplore/home.jsp.

[8] Elsevier. Science Direct. Elsevier B. V. http://www.sciencedirect.com.

[9] J. Xavier, A. Macedo, R. Matias Jr. e L. Borges. "Engenharia de Confiabilidade de Software: Um mapeamento sistemático da última década (Apêndice)". http://hpdcs.facom.ufu.br/apendice/. 2014

[10] R. A. Johnson e D. W. Wichern, "Applied Multivariate Statistical Analysis", Upper Sandler River, New Jersey, 2007.

[11] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information". Psychological Review, vol. 63, n. 2, pp. 81-97, Março 1956.

[12] Google. Google Scholar. Google, http://scholar.google.com.

[13] N. Nagappan, B. Murphy e V. R. Basili, "The influence of organizational structure on software quality: an empirical case study". In Proc. of the Int'l Conference on Software Engineering. ACM, NY, USA, 2008, pp. 521-530.

[14] N. Nagappan, e T. Ball, "Use of relative code churn measures to predict system defect density," In Proc. of the Int'l Conf. on Software Eng. USA, 2005, pp. 284 - 292.

[15] C.-Y. Huang e C.-T. Lin, "Software Reliability Analysis by Considering Fault Dependency and Debugging Time Lag," IEEE Transactions on Reliability, vol. 55, n. 3, pp. 436 – 450, Setembro 2006.

[16] D. M. Nicol, W. H. Sanders, e K. S. Trivedi, "Model-based evaluation: from dependability to security". IEEE Transactions on Dependable and Secure Computing, vol. 1, n. 1, pp. 48-65, Outubro 2004.

[17] C.-Y. Huang, M. R.-T. Lyu e S.-Y. Kuo, "A unified scheme of some Nonhomogenous Poisson process models for software reliability estimation," IEEE Transactions on Software Engineering, vol. 29, n. 3, pp. 261-269 ,Março 2003.

[18] C.-Y. Huang e S.-Y. Kuo, "Analysis of incorporating logistic testing-effort function into software reliability modeling," IEEE Transactions on Reliability, vol. 51, n. 3, pp. 261-270, Setembro 2002.

[19] D. Saff e M. D. Ernst, "An experimental evaluation of continuous testing during development". In Proc. of the 2004 ACM Int'l Symp. on Software Testing and Analysis. ACM, NY, USA, 2004, pp. 76-85.

[20] C.-Y. Huang e C.-T. Lin, "Software Reliability Analysis by Considering Fault Dependency and Debugging Time Lag," IEEE Transactions on Reliability, vol. 55, n. 3, pp. 436 – 450, Setembro 2006.

[21] S. Gokhale e M. R.-T. Lyu, "A simulation approach to structure-based software reliability analysis". IEEE Transactions on Software Engineering, vol. 31, n. 8, pp. 643 – 656, Agosto 2005.

[22] L. Li, K. Vaidyanathan e K. S. Trivedi, "An approach for estimation of software aging in a Web server". In Proc. of Int'l Symp. on Empirical Software Eng., Japão, 2002, pp. 91 - 100.

[23] S. V. Amari e L. Mclaughlin, "Optimal design of a condition-based maintenance model". Annual Symp. Reliability and Maintainability, EUA, 2004, pp. 528 - 533.

[24] M. Grottke, R. Matias Jr, e K. S. Trivedi, "The fundamentals of software aging," In IEEE Int'l Workshop on Software Aging and Rejuvenation, Seattle, WA, 2008, 1-6.

[25] K. Vaidyanathan, e K. S. Trivedi, "A comprehensive model for software rejuvenation". IEEE Trans. on Dependable and Secure Computing,vol. 2, n. 2, pp. 124 – 137, Junho 2005.

[26] Z. Li, L. Tan, S. Lu, Y. Zhou, e C. Zhai, "Have things changed now?: an empirical study of bug characteristics in modern open source software," Workshop on Architectural and System Support for Improving Software Dependability. ACM, NY, EUA, 2006, pp. 25-33.

[27] Y. Xie, L. Li, M. T. Kandemir, N. Vijaykrishnan e M. J. Irwin, "Reliability-aware co-synthesis for embedded systems". In Proc. of the IEEE Int'l Conference on Application-Specific Systems, Architectures and Processors, Galveston, Texas, EUA, 2004, pp. 41 - 50.

[28] Z. Yin, D. Yuan, Y. Zhou, S. Pasupathy, e L. Bairavasundaram, "How do fixes become bugs?" In Proc. of the ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering. ACM, NY, EUA, pp. 26-36.2011.

[29] R. Pietrantuono, S. Russo e K. S. Trivedi, "Software Reliability and Testing Time Allocation: An Architecture-Based Approach". IEEE Transactions on Software Engineering, vol. 36,n. 3, pp. 323-337, Janeiro 2010.

[30] R. Matias Jr., P. A. Barbetta, K. S. Trivedi e P. J. F. Filho, "Accelerated Degradation Tests Applied to Software Aging Experiments". IEEE Transactions on Reliability, vol. 59, n. 1, pp. 102 – 114, Março 2010.

[31] Y.-S. Su e C.-Y. Huang, "Neural-network-based approaches for software reliability estimation using dynamic weighted combinational models". Journal of Systems and Software, vol. 80, n. 4, pp. 606-615 Abril. 2007.

[32] J.-H. Lo e C.-Y. Huang, "An integration of fault detection and correction processes in software reliability analysis". Journal of Systems and Software, vol. 79, n. 9, pp. 1312–1323, Setembro 2006.

[33] C.-T. Lin e C.-Y. Huang, "Enhancing and measuring the predictive capabilities of testing-effort dependent software reliability models. Journal of Systems and Software", vol. 81, n. 6, pp. 1025–1038, Junho 2008.

[34] S. Keele, Guidelines for performing systematic literature reviews in software engineering, EBSE Technical Report EBSE-2007-01, 2007.