Cognitive Distance vs. Research Output in Doctoral Computing Education: A Case-Study

Damian A. Tamburri, *Member, IEEE,*
Giuliano Casale, *Member, IEEE,*

***Abstract*—*Contribution.* The current organisational structures for doctoral education often lead to a stagnation of cross-disciplinary collaboration, reducing the potential impact of academic research.**
***Background.* We argue the need for research into the social and organisational variables that regulate educational "silos" where doctoral education is carried out. In particular, we believe that, from an organisational and social dimension, the lack of cognitive distance across educational communities – a measure of the background, cultural, and expertise difference – hinders the quality and variety of research outputs.**
***Research Questions.* TTo evaluate our research hypothesis we elaborate on the following key research question:"Does collaboration with similarly-expert researchers yield better research?"**
***Methodology.* We test our conjecture in a mixed-methods case study involving 123 researchers in a large computing research institute in Europe, ranked among the first 50 in the world.**
***Findings.* We observe that an increase in expertise overlaps across sub-fields of computing are, in fact, strong predictors for further collaboration (quantity), but *research impact (quality) decreases with larger overlaps*. We argue that this reveals an educational "silo" effect in doctoral computing education and, consequently, a flaw in the connected research output. Also, we observe that there is no single and uniquely acknowledged way to evaluate research impact across sub-fields, which further hinders cross-departmental collaboration among doctoral students.**

***Index Terms*—Education Organisational Structures, Social Adaptivity, User Perspective, organisational Decision-Making**

I. Introduction

A rapid look into the proceedings of top computing conferences, e.g., the Education and Training track within the International Conference on Software Engineering[1], reveals that computing education research is focusing mainly on technical education, typically through courses that cover foundations and industrial-strength computer science and engineering topics [1]. Although extremely valuable, this research says little on *how to educate the educators*. This subject is broad, but includes the important objective of defining collaboration best practices, leading to good research as part of a healthy, collaborative, and sustainable research community.

Many researchers in computing have touched on this topic, most prominently D. Parnas [2] who, about 10 years ago surveyed the production and productivity races typically expected of non-tenured and student-level computing researchers. Parnas comes to argue that these races do in fact produce more research, but *not better* research. Parnas concludes that the computing research community must strive to stop the numbers game, aiming for (more appraisable) quality impact rather than quantity of outcomes.

As a first step towards understanding the above mentioned "numbers game", this article offers insights from a mixed-methods study aimed at understanding the influence of cognitive distance [3] and expertise overlap [4] onto the quality and quantity of collaborative research output, with a focus on young doctoral students. We set out to answer the following research question: "Does collaboration with similarly-expert computing researchers yield better research?". We answer this question with a focus on doctoral students since they represent the future of the computing discipline as we know it.

The objectives of this study are threefold: (1) understand how collaboration across diversely-skilled peers influences output quantity and quality; (2) understand if and how research across different areas may be evaluated and compared equally in terms of quality; (3) understand if Social-Network Analysis (SNA) of collaboration patterns across the same department reveals cohesive groups rather than fragmented ones.

To address the above objectives, we quantitatively and qualitatively investigated a large computing research department from a high-standing European (EU) computing research institution ranked among the first 50 in the world. Our study collects answers from 123 professionals. Although we have focused on this dataset, our scope of study is not limited to the targeted department; rather, we looked at their joint publications with research practitioners *outside* of their department (albeit in the same organization), to evaluate any external expertise overlap, gaps as well as these variables' role in quality and quantity of research output.

Perhaps unsurprisingly, our data shows that increased expertise overlaps across research professionals, especially doctoral students, are strong predictors for further continued collaboration and for the *quantity* of research outputs . Remarkably, however, the *research impact* of resulting collaboration becomes *exponentially lower* with respect to research conducted jointly with collaborators that have little or no overlapping expertise. This becomes an even major problem for doctoral students whose output relevance, according to our data, becomes negligible.

Also, we observe that it is very difficult to find a homogeneous metric for research impact across sub-fields of computing research, let alone across related disciplines. Our qualitative evidence confirms that this lack of homogeneous impact metric hinders further collaboration of both tenured

D. A. Tamburri is with Politecnico di Milano - Italy
E-mail: [$damianandrew.tamburri$]@polimi.it
G. Casale is with Imperial College London.
Manuscript Submitted January 2017.
[1]http://icse2017.gatech.edu/seet

and non-tenured personnel across computing sub-disciplines as well as other disciplines. What is more, doctoral students as well as non-tenured personnel tend to ignore any possibility of collaboration or even compete against each other. For example, the data shows that doctoral students have the highest power distance across them, as well as the lowest betweenness centrality in future-intended collaborations reported across the considered department.

Finally, from a SNA perspective, we observe that scientists in different sub-disciplines of computing create closed sub-communities. These silos of researchers build their own organisational and social culture as well as research standards. Moreover, these silos span well beyond the original organisational boundary, i.e., the established collaboration and organisational culture remains even if researchers leave the organization. Although this process may seem quite natural, it should be guided carefully since the resulting organisational-social structure [5] further reinforces silo stagnation, hindering collaboration between computing researchers and other disciplines.

The rest of this paper elaborates on the analysis framework and study that led to the observations above (see Section III and IV). Further on, the paper elaborates on the research design behind the study itself (see Section IV). The paper concludes with an outline of simple strategies to overcome the shortcomings outlined above (see Section VI and VII).

II. Background and Motivation

The state of the art in computer science education provides plenty of examples on how computing research and doctoral education are often conducted in many detached sub-communities where multi-disciplinary interactions are often latent if not absent.

On the one hand, several studies report on the critical centricity covered by computer science in pretty much all disciplines; for example, the report on multidisciplinary and interdisciplinary research by InnovationToday [6] elaborates on the need for further cross-fertilisation among the practical areas (Computer Science, Engineering, etc.) of science and their theoretical/empirical counterparts (e.g., Chemistry, Physics, Biology, etc.) both in terms of method and joint educational programmes.

On the other hand, beyond the anecdotal arguments, much literature discusses the limitations that currently affect computer science alone [7], [8], [9]. For example, Ni and Dongarra [8] chair a symposium to discuss the very same limitations that we are addressing in the present manuscript. The authors conclude that a dense and overly-varied proliferation of sub-sectors in computer science has made it difficult to establish a clear-cut community at present. Also, authors prominently call for joint restructuring of educational programmes in computer science such that cross-fertilisation is indeed utilised as a driver for innovative educational content at all levels, including the doctoral education we aim to address. Although the work does provide several ideas on how to provide for said restructuring, it remains to provide methods to quantify, track, and evaluate the status of the cross-fertilisation and quantify the impact of the proposed restructuring. We set out to fill these gaps so that policy-makers and educational leaders may refine their educational structures using quantitative empirical evidence.

Similarly to Ni and Dongarra [8], Roberts [10] highlights that the cognitive gap between practitioners with different expertise is one of the key reasons why multidisciplinary research is difficult in computer science, let alone with other disciplines like biology or economic sciences. From this background research, we inherit the idea and conjecture that cognitive distance may indeed play a role in anticipating quantity and quality of impact across computer science and other disciplines. We seek the goal of making the phenomenon more explicit and encouraging its further investigation from an educational research perspective.

From a slightly different perspective, we seek to understand what are the variables that dictate likely successful (i.e., impactful) collaboration between computer science and other disciplines. Previous research by Dryburgh et al. highlights that a very strong gender-gap is clearly predominant in computer science [11]; Dryburgh et al., also highlight that women collaborating with others do provide a more straightforward collaboration, possibly leading to better quantity and quality of results. In our study, we seek to evaluate this conjecture quantitatively, understanding if and how there is in fact a gender bias between male- and female-oriented collaboration across computer science as a whole. If so, then educational institutions may need to restructure their educational programmes by encouraging more homogeneous collaboration across any identified gender gaps.

III. Theoretical Framework

For the scope of this study we aim at understanding the conditions wherefore, according to D. Parnas et al. , there is in fact a heavy tendency to produce more, rather than better, with a focus on doctoral scholarly research. As stated, we wish to examin wether collaboration with similarly-expert researchers yields better research. Our conjectures with respect to this question is that cognitive distance, as measured by expertise similarity, has a measurable effect in quantity and quality of research output. More in particular, our study yields the following qualitative claims:

- **Claim 1. Computing Research Area is a Strong Predictor of Publication Frequency -** We claim that the area of expertise in which doctoral education is conducted, may bias students (or even researchers at all levels) towards producing more results, rather than groundbreaking research. This conjecture assumes that the areas in question have a very fast research cycle, e.g., since they are completely novel and still in an aggressive expansion phase. Finding out insights into this conjecture may provide valuable quantitative means for each research area to address their speed, e.g., focusing doctoral education paths towards more regimented forms.
- **Claim 2. Expertise Non-Overlap Leads to Higher-Impact Research -** We also claim that a sensible overlap of different expertise may measurably mitigate the

EXPERTISE	OPEN CODES	AXIAL CODING	GENERALI-ZATION	DEPARTMENT
automata (on infinite objects)	AUTOMT		FRML-LNG	Theoretical Software Engineering
automated component configuration	AUCOCONF	SOFT-DES	SOFT-ENG	Distributed Systems
Automated Content Analysis	AUCONANA	CNT-ANL	CNT-ANL	Artificial Intelligence

Fig. 1: Expertise Coding, an example.

above speed-up effect. This assumes that it may be more difficult, and hence slower, for doctoral students and researchers to collaborate with diversely-expert colleagues, even though the multi- and inter-disciplinary nature of the collaboration may yield highly innovative and hence high-impact result. Finding out insights into this claim may provide valuable indications and collaboration patterns which are healthier from a doctoral education perspective, and hence are to be used in structuring scholarly doctoral education and research.

The above two claims led us to formulating two distinct research questions in the following form :

1) **RQ1** "Does same-topic collaboration reflect higher quantity of research output with respect to a cross-topic collaboration?"
2) **RQ2** "Does cross-topic collaboration reflect higher quality of research output with respect to a same-topic collaboration?"

As a side study to the above research questions, we also set out to understand if there exists a way to classify and compare collaborations using an impact measurement which is homogeneous, that is, it can be used as a single common denominator of impact factors across expertise and research areas which are very distant from each other. Our conjecture here is that, without such a mechanism, researchers of all rank and status are not encouraged to embark on time-consuming collaborations outside their publication confort zone. Investigating further this shortcoming may reveal valuable insights into the strengths and limitations of the current academic output evaluation schema, such that healthy and impactful collaboration can be effectively encouraged.

IV. Research Design

This paper draws results from an analysis of collaboration structures emerging across 123 scientists active in various disciplines of computing from business web and media engineering to distributed systems to theoretical computing.

A. Evaluating Expertise Areas

Expertise overlap is a measurement of "the quantity of overlaps in self-reported expertise items across co-workers" [12].

The results in this study were obtained combining qualitative and quantitative empirical research, featuring surveys, interviews, qualitative, and quantitative modelling and analysis. The study started with a department-wide survey respondents were asked to complete a survey in which they first reported on their own areas of expertise (up to 5, self-selected and self-ranked in order of expertise level) and their formal and informal relationships with other respondents in the same department, reporting on: (a) previous studies location, topic and years of study; (b) supervisor-supervisee relationships with other department respondents; other organizational relations outside the department and planned collaborations[2].

Approximately 400 entries provided by the respondents as self-reported expertise areas were coded and classified applying principles of Grounded-Theory [13][3].

GT is generally split into three phases:

1) *Open Coding:* during this phase the primary sources of data (usually text or qualitative data of some sort) are read or viewed and each portion of the text is constantly compared to "codes", i.e. labels for portions of text that represent the concepts expressed. If the elements in the text express concepts that cannot be mapped to an existing code, then an additional code is created. Moreover, the researcher applying the method is constantly noting down observations in the form of "memos", i.e. brief textual notes describing an idea or observation made on the data so far or on the portion of text just analysed. The end result of open coding is a set of *categories*, that is, an inferred taxonomy of codes.

2) *Selective Coding:* during this phase the codes are constantly compared with each other (rather than with other pieces of data) until the categories are saturated into clusters of core-categories (this step essentially provided us with clusters of self-reported expertise areas).

3) *Theoretical Coding:* during this phase the codes and core-categories are sorted and tentatively modelled (a process sometimes known as Theoretical Modelling). A Grounded Theory is generated by letting it emerge from one (or more) theory views or available data models, usually by observation, e.g., through theory-inference focus groups. For example, a focus-group of expert code reviewers may visually inspect, confirm, and refine all the relations inferred by researchers through open- and selective-coding.

One the one hand, the resulting dataset consisted of a two-mode social network [14] containing the respondent names mapped to 121 fine-grained expertise areas in computing.

On the other hand, we obtained presentable computing research sectors featured in our sample — Figure 1 shows an example of said coding exercise featuring Grounded-Theory and the resulting coding of expertise areas. Also, Figures 2 to 5 give an overview of the population in our sample. First, Figure 2 highlights population arrangement in the five computer science and engineering areas in the department under study, ordered by decreasing population size: (1) Artificial Intelligence (AI) and Internet of Things (IoT); (2) Software Systems and Networks Engineering; (3) Applied computing; (4) theoretical computing; (5) information systems and management; (6) Bioengineering; in this respect, to allow proper assessment of expertise overlap ratio, Figure 3 plots

[2]The actual survey questionnaire is obscured to hide the target of our study. For the sake of clarity and replicability we link the standard Canadian Organizations questionnaire for work practices improvement — this was used as basis for our study: http://tinyurl.com/h6g65n2

[3]the entire coded dataset can be found here: http://tinyurl.com/hac5nhn

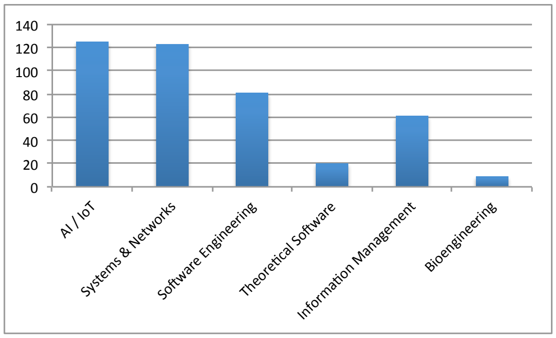

Fig. 2: Population expertise distribution in our sample.

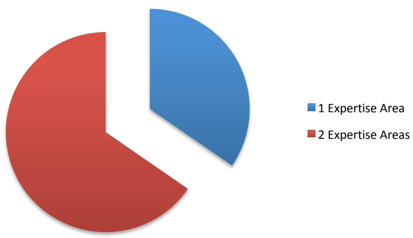

Fig. 3: Ratio between overlapping & non-overlapping expertise areas.

the ratio between researchers expert in 1 or 2 areas of the previous list. Second, Figure 4 elaborates on the tenure status arrangement across our sample, showing an almost one-to-one ratio between tenured (i.e., assistant or associate professors) and junior (i.e., ph.d. students and post-docs) scientists as well as an almost one-to-three ratio between tenured and full professors. Fourth, finally, Figure 5 shows citations averages for full (11500 circa), tenured (3100 circa) and junior (134 circa) scientists. Finally, in terms of gender imbalance, our data highlights that the arrangement of gender across our sample tends towards male scientists (82%) with respect to their female colleagues (18%).

To proceed with sample analysis, the dataset was transformed with direct projection and matrix transposition to compute overlap in expertise at the dyadic level [15] – that is, overlaps between the expertise of pairs of individuals – these overlaps are strongly predictive of current and intended future collaboration as reported by the same survey respondents [16]. Figure 6 outlines the conceptual overview of the expertise overlap as obtained in our study. Finally, we empirically investigated the relationship between expertise diversity/overlap as a means to evaluate cognitive distance (i.e., a measure of the difference in background, knowledge and reasoning schemas [17]) against research output.

B. Assessing Impact of Interdisciplinary Collaboration

To assess the impact of interdisciplinary collaboration using relevant outcomes, we collected historical data on academic output both previous and subsequent to the starting date of our study (early 2013) for the sampled group of 123 software scientists. Our goal was to derive a complete overview of contributions stemming from research collaboration prior and subsequent to the starting date of our study. To obtain said research output, several academic databases and search engines

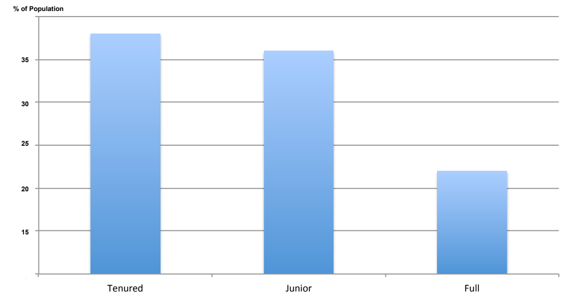

Fig. 4: tenure status arrangement across in our sample.

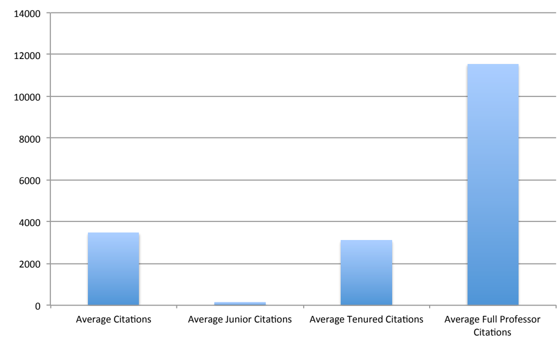

Fig. 5: citations average arrangement across in our sample.

were used. More specifically, the Microsoft Academic Search, Web of Knowledge and Google Scholar, DBLP, Pubmed and Scopus databases and search engines were combined to derive a complete overview of the contributions and collaborations of the 123 scientists in our dataset.

To complement said publications after extensively reviewing the databases above, the publication registration system of the European research institute under examination was contacted. The research information database system enables individual researchers as well as universities and research organizations to register information about their research and make this information available in a multitude of ways. This registration system provided a comprehensive list of research output of our 123 samples. More specifically, using database queries, we obtained 1.058 research outputs for the period from 2000 to 2004, the year in which the survey was conducted. Furthermore, for the period 2005 till now 3.570 research outputs

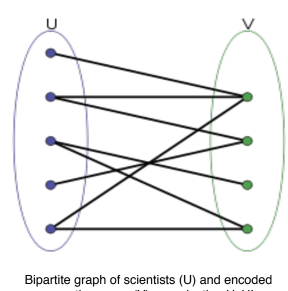

Fig. 6: Expertise Overlap, a conceptual overview, a bipartite graph of scientists (U) and expertise areas can be projected (operator "->") by multiplying it by its transposed (U') - UxU' is a new matrix ("scientists" x "scientists") where each cell contains scientists' expertise overlaps.

were generated, including book chapters, books, journal publications, patents, posters and conference proceedings.

At this point, we already reported a lack of a standardized impact measure for conference proceedings, as well as patents and posters - bibliometrics for these outputs were analysed manually and an H-index was generated. The focus around using H-index for conference proceedings, posters and patents is that there was no homogeneous measurement scale available to compare with impact scores derived for other scientific output outside of H-index.

Our complete dataset counts as follows: (a) 1.070 journal publications; (b) 2348 conference proceedings papers and posters+abstract; (c) 67 books; (d) 248 book chapters.

With respect to journal publications we refer to article influence scores derived from the website eigenfactor.org as an objective comparison means for different journals across any number of scientific segments. This characteristic is unique to Eigenfactor, paraphrasing from the definition: “The modified Eigenfactor centrality algorithm used to rank journals at Eigenfactor.org expands upon a thirty-year tradition of using iterative methods to quantify the influence of scholarly publications" [18]. The central role in the Eigenfactor metric is the number of citations, providing a direct estimate of how often certain journals are used. Ranking systems account for prestige levels of citing journals where citations are compared to citations from third-tier journals with narrower readership. In addition, “the Eigenfactor score also adjusts for differences in citation patterns among disciplines" [18]. In the end this leads to an article influence percentile that is a measure of the average influence of each of the articles in a given journal over the first five years after publication. For each of the 1.070 journal publications in the dataset the article influence percentile was derived manually. For some publications the year in which it was published was not provided on Eigenfactor. For these publications we established an average score based on averages for preceding and subsequent years provided. A similar approach was used to infer Eigenfactors automatically for all conference proceedings, based on citations and conference venue acceptance rate, available through ERA Core Rankings[4].

Finally, to include books and book chapters we had to establish a homogeneous measure of impact. To do so we referred to a ranking of book publishers provided by the research institution under examination. Book publishers are ranked according to an A- or B-rank where publishers that are not in the list are considered to be C-rank publications. As books are more concise and well-elaborated pieces of work they are generally accepted to be more renown and influential publications. Therefore entire books are considered to have higher influence percentiles opposed to book chapters. A book published by an A-rank publisher derives a percentile of 0.95 opposed to 0.45 for A-rank published book chapters. For B-rank publishers awarded percentiles are 0.7 and 0.15 where C-rank publishers are worth 0.4 and 0.15 respectively. Several control variables that might influence hypothesized relationships between expertise diversity, cognitive distance and quantity as well as impact of subsequent output were included in our work.

[4]portal.core.edu.au

C. Factors Control

Several control variables were checked to mitigate their influence and effect on the dataset and results.

First, we controlled on supervisors since the de-facto process of academic PhD studentship envisions students and supervisors to likely co-author several studies together. In addition, supervisors often invite PhD students to collaborate with other PhD students, thereby influencing the dataset.

Second, we controlled gender, assuming that individuals similar in gender might be able to collaborate more efficiently, as suggested by [19].

Third, we controlled department, to assess differences among departments. For example, collaborations between bioinformatics and artificial intelligence may yield significantly better or worse results than other collaborating departments.

Fourth, we controlled Position (e.g., difference in career status), referring to the function a researcher has, as described by the research institute. It is plausible that two scientists with the same position might compete with each other. Such competition might impede collaborative efforts.

Finally, we controlled tenure, referred to as years of employment at the research institute. It is likely that two scientists with very different levels of tenure experience miscommunication in their collaborative efforts due to different levels of experience.

In addition to the above five control variables we also controlled for differences between collaborations from the year 2000 up to 2004 and 2005 till the end of the study (i.e., Oct. 2014).

D. Data Analysis

For the purpose of analysing our data, we used standard SNA metrics from previous literature [14], basic clustering according to the BNP clustering algorithm as tuned by cognitive distance [20] as well as mean-centered and non mean-centered regression modelling [21].

Concerning SNA, we used boundary-spanning and centrality metrics to understand the organizational characteristics of the social network reflected in our dataset (see Section 3). One the one hand, boundary-spanning metrics allow a researcher to find the most highly-connected “in-betweeners" in a social-network, that is, those individuals who are most frequently between the communication or relations across other members. On the other hand, centrality metrics in social-networks analysis allow a researcher to understand the most influential person(s) across a social-network — in our case, we adopted Freeman’s betweenness centrality [22], defined as the number of shortest paths which pass through the given vertex. We applied both these metrics to cross-reference which type of population (tenured, non-tenured, supervisor or non-supervisor) across our organizational social-network was most influential with respect to high-impact research for young doctoral students.

TABLE I: Average Impact Output after year 2005, an overview.

Variable	Non Mean-Centered	Mean Centered
Supervisor's Role	0.273**	0.273**
Gender	0.002	0.002
Department	0.023**	0.023**
Position	-0.013	-0.001
Tenure	0.002	0.002
Avg. Impact <2005	1.196**	1.196**
Expertise Overlap	-0.022**	-0.022**

TABLE II: Exact Output Count after year 2005, an overview.

Variable	Non Mean-Centered	Mean Centered
Supervisor's Role	0.497**	0.497**
Gender	0.019	0.019
Department	0.072**	0.072**
Position	-0.033	-0.033
Tenure	0.002	0.002
Avg. Impact >2005	1.196**	1.196**
Expertise Overlap	-0.025**	-0.025**

In addition, we used the K-means clustering algorithm in R[5] to identify and further characterise clusters across our social-networks so as to make evident any peculiar nature of identified clusters.

Finally, concerning regression modelling, we sought to correlate the variables under our control with (a) the average quality resulting from collaboration across the social network and (b) quantity of output in the same network (see Sections 3 and 4).

V. Results Overview

This section outlines the results of our study using resulting tables, social-networks and graph-plots that served for the evaluation of the key findings stemming from our analysis[6]. Further on, Section VI outlines and discusses these results providing said findings and improvement actions for more sustainable scholarly education and its planning.

Table I shows correlation results featuring expertise overlap (i.e., number of overlapping self-reported expertise entries or skills - see Section IV) and average impact, as well as the significance of controlling variables (see Section IV). Also, Table II shows correlation results featuring, again, expertise overlap and quantity of research output. Figures 8 and 7 show the regression model reflected by correlation values in Table I and II. Graphs also account for positive and negative differentiations of standard deviations.

Social-Networks in Figures 9 and 10 reflect the collaboration structure inherent in our dataset (i.e., the social-network in 9) as represented by co-authorship data as well as the future collaboration intentions intended by department members subject to this study (i.e., the social-network in 10), as reflected by our own interview data. In both social networks squares represent non-tenured personnel, while circles represent tenured or fully-structured personnel while edges reflect collaborations. Multiple collaborations and higher expertise are used as parameters for edge-closeness and subsequent clustering. Colours across the social network identify considerable expertise-overlaps (3+ joint expertise). The dotted circles identify clusters found through K-means Machine-Learning — as evident, a single overlapping element across sub-communities is found.

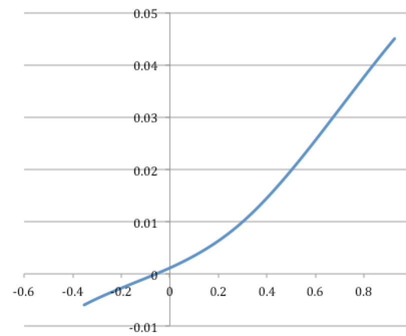

Fig. 7: Expertise Overlap Vs. Quantity of Output across a large computing research department, a Regression Model.

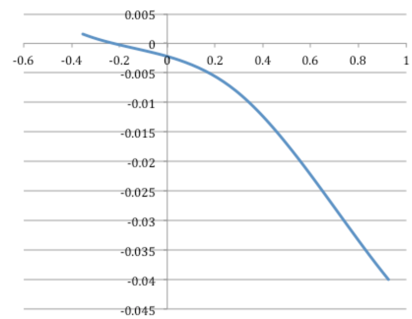

Fig. 8: Expertise Overlap Vs. Average Impact across a large computing research department, a Regression Model.

VI. Discussion

A first blatant observation is that our results indicate that similarity in expertise is extremely negatively related to average quality and impact of research output.

> **Finding 1.** there exists a negative relationship between expertise similarity and average quality and impact of research output - this relation is exponential, and similar by over 95% the second part of an inverted U shape according to graph-isomorphism analysis in R[7].

Furthermore, further analysing our corpus of data against control variables, we observed that the supervisor variable has the highest significance value in case of average impact of former and current doctoral students, which indicates that:

> **Finding 2.** doctoral students who have been supervised by supervisor X are strongly intent in collaborating further

[5]https://www.r-bloggers.com/k-means-clustering-in-r/

[6]for replication purposes only, we made available the coding scheme we refined and applied as well as all the analysis logs and available data online: http://tinyurl.com/mfobvrd

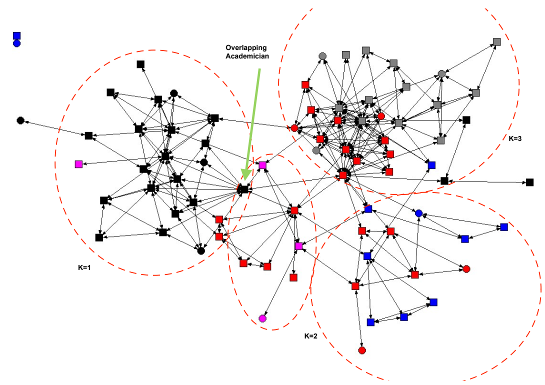

Fig. 9: A social network reflecting the collaboration structure behind our dataset (outliers are omitted for the sake of brevity); dotted clusters identify sub-communities, a single overlap is found.

with X and produce outputs of higher influence than they would by themselves.

Finding 2 confirms our conjecture (1). Also, with respect to finding 2, we noticed that sub-communities across the department under investigation were forming around supervisors and their former doctoral students or research associates even beyond the student/associate's own presence in the department under study.

Finding 3 From a social-network perspective, oldest supervisors have the highest centrality values in sub-communities populated by: (a) former and current students; (b) continued tenured and non-tenured collaborators.

Furthermore, we noticed that gender does not seem to play important role in terms of intended collaboration nor research impact. When counting the number of outputs, the people from the same department were willing and actually did collaborate equally with people of varied gender, also in case of supervision. In fact, the clustering coefficient of both social networks in 9 and 10 for current and intended collaboration were not influenced by gender - removing gender information and comparing with randomly generated null-models, both with-gender social-networks and without-gender social-networks yielded the same and exact clusters.

1) Answering our Research Questions: Cross-Topic Collaboration and Cognitive Distance: As part of answering the first research question, namely:

"does same-topic collaboration reflect higher quantity of research output w.r.t., cross-topic?"

We found that, not only the number of outputs is higher when bigger expertise overlap, but also, this number is exponential. As a corollary to this finding, we noticed that similarities regarding to affiliation to the same department and

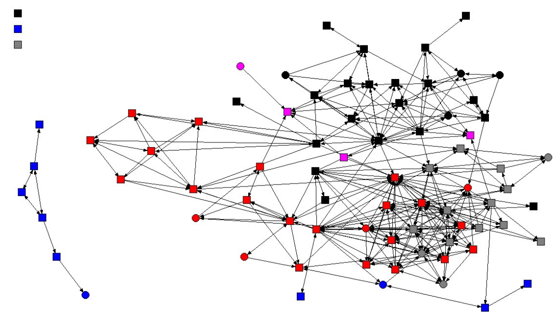

Fig. 10: A social network reflecting the future intentions for collaboration based on our survey.

supervision relationship among scientists affect significantly the number of co-authored outputs. On the other hand the collaboration is negatively related if people have the same position and becomes zero for doctoral students - a possible interpretation is that, on one hand non-tenured doctors tend to compete, creating an even more self-contained and sterile research silo, and on the other hand, doctoral students remain practically isolated in their supervised-research silo. Also in this case we reported no significance with respect to the gender and tenure difference.

Finally, with respect to the quality of impact in the scope of RQ2, the average impact drops considerably, in case of high expertise similarity. Moreover, the values of controlling variables, do not affect hypothesis acceptance. Actually, people with the same position produce less influential output. However, it was clear to us from both results that people who worked together before 2005 showed higher collaboration during the period after 2005. This suggests a stagnancy of research collaboration across software scientists of all ranks and status in the observed organization.

A. Side-Study - Cognitive Distance and Harmonious Quality Research Evaluation

As previously stated, part of our survey was dedicated to understanding the evaluation mechanisms used by practitioners from each sub-area of computing we coded in the department under study. We elicited an average cluster of 5 metrics used in each sub-area of computing. Not surprisingly, a single, recurrent measurement used and considered in all clusters, that is, academic impact factor indicator H-Index. However, when inquired as to weather a higher impact factor from an unknown publication (that is, a publication belonging to a sector other than their own and referring to that of other practitioners with considerable expertise difference) was enough to warrant collaboration and successful publication, we found that:

Finding 4. over 70% of our respondents reported initial reluctancy towards publication outside their usual venues while over 45% of them reported attempts at "re-routing" the collaboration to a familiar venue[8].

The above finding 4 essentially means that collaboration is often tied to venues which are common to a certain area of expertise - researchers therefore are bound to collaborate within the same expertise and "comfort zone". We could not get access to any followup data that would give us the possibility to look further into the matter but we were able to conclude that cognitive distance plays a role in determining where and how collaboration is directed in terms of quality and quantity of output and that it plays a major role in creating a narrow and siloed mindset in doctoral students and their collaboration as well as publication habits.

Conversely, however, the data we reported and statistically modelled in Figures 8 and 7 clearly indicates that wherefore those collaborations across multiple disciplines are successful, the impact of out-coming results are indeed exponentially higher.

1) RQ3: Cognitive Distance and Organisational Structure: When analysing the collaboration social-networks reflected by our dataset (see Figure 9) we observed that: (a) previous collaborations reflect a tightly knit organisational network where 4 clusters reflect strong seniority$-$ $>$juniority relations as well as high expertise overlaps; (b) about 13% of the network members have collaborated in the 5 years previous to our study; (c) cliques of about 33% size of the entire network form around 3+ expertise-overlaps, with a single exceptional member (the *decaan* of the organisation in question) overlaps between two clusters. From this scenario, one fact appears clear:

> **Finding 5.** a clear-cut number of strong, self-contained and potentially stagnant silos are forming around highest expertise-overlapping collaborators - also, the trend of generating *more* as opposed to *better* research is more and more predominant the more junior researchers: doctoral students who team up with previous or supervising personnel.

Finally, when analysing future collaboration intentions the situation *gets worse* rather than improving. Figure 9 shows the formation of pretty much identical clusters in the right-hand side of the network (a graph isomorphism analysis with R reveals 83% similarity for this area of the network) with the further addition of an increasing number of separate sub-graphs or other elements.

B. Limitations and Threats to Validity

Like any study of comparable magnitude and scale, this study is affected by several limitations, assumptions, and threats to validity; this section discusses the major ones we spotted in our study design or execution. First, two major limitations are obvious: (a) the study was dedicated to a single computing research and education department within a single research organization and hence could be hardly generalisable; (b) the study used a combination of statistical and social-networks analysis techniques. Although common in empirical research we recognised a few limitations as we applied them in this domain.

Internal and Sampling Validity. Internal validity refers to the internal consistency and structural integrity of the empirical research design, more specifically it refers to how many confounding factors may have been overlooked. For example, the process of peer-review revealed to us that the results we attain may be compromised by the lack of clear sampling over the age, tenure-length, and publication strategies of the involved participants. In the case of age, there may be a bias towards too young non-tenured staff being conditioned to "following" the lead of more senior personnel. Similarly, some lengthy-tenured professionals may be looking outside of their confort zone in view of multi-disciplinary research much more than their young-tenured peers. We strengthened this threat with an accurate sampling strategy whereby we controlled as many variables as possible to: (a) ensure a meaningful variety of our sample; (b) ensure that basic controllable variables for the objects of our study.

Conclusion Validity. Conclusion validity represents the degree to which conclusions about the relationship among variables are reasonable. In the scope of the discussions of our results we made sure to minimise all possible interpretations to a minimum, designing the study with reference to known hypotheses and facts from previous literature. Also, our conclusions were drawn from sound social-networks analysis measurements[9] statistical modelling and analysis of our dataset.

VII. Conclusions

In this article we summarise and report on a 2-year longitudinal study of a large European research institution, very active in computing research. We were able to understand the key limitations behind increasing collaboration across software science from insights we gathered in this study. More in particular, in studying the observed limitations, we proposed high-level corrective strategies which were subsequently adopted in the very same institution we studied. At this point, almost 2 years after our study has ended, we are re-enacting the same study in an additional set of 2 large computing departments in the hope of fully replicating the insights elaborated in this article for their further confirmation.

Up to this point in both of these studies, we found that, although collaboration across software science is indeed high, it looks as though said collaboration is stagnant and mainly rotating around previous work relations reflecting very low cognitive distance. Conversely, we conclude that a more communitarian vision is needed to change this trend into a more productive environment fostering the generation of innovation and academic impact.

Although our study was carefully planned, crafted and executed using triangulation of sources as well as analysis, its limitations are clear, for example: (a) the study was dedicated to a single computing research and education - this hinders generalisability; (b) the study used analysis techniques common in empirical research but with several few limitations as applied in our domain of study.

[9]Logs for our analysis are available online: http://tinyurl.com/mfobvrd

REFERENCES

[1] P. Naur, "Computing versus human thinking," *Communications of the ACM*, vol. 50, no. 1, pp. 85–94, 2007.

[2] D. L. Parnas, "Stop the numbers game," *Commun. ACM*, vol. 50, no. 11, pp. 19–21, 2007.

[3] S. Schmid, "Organizational learning in innovation networks: exploring the role of cognitive distance and absorptive capacity ; an agent-based model," Ph.D. dissertation, Uni Friedrichshafen, 2014.

[4] A. I. M. J. Rahman, R. Guns, R. Rousseau, and T. C. E. Engels, "Expertise overlap between an expert panel and research groups in global journal maps." in *ISSI*, A. A. Salah, Y. Tonta, A. A. A. Salah, C. R. Sugimoto, and U. Al, Eds. Bogaziçi University Printhouse, 2015. [Online]. Available: http://dblp.uni-trier.de/db/conf/issi/issi2015.html#RahmanGRE15

[5] D. A. Tamburri, P. Lago, and H. van Vliet, "Organizational social structures for software engineering." *ACM Comput. Surv.*, vol. 46, no. 1, p. 3, 2013. [Online]. Available: http://dblp.uni-trier.de/db/journals/csur/csur46.html#TamburriLV13

[6] O. for Economic Co-operation and Development., *Measuring innovation : a new perspective*. Organisation for Economic Co-operation and Development Paris, 2010. [Online]. Available: http://titania.sourceoecd.org/vl=592502/cl=16/nw=1/rpsv/cgi-bin/fulltextew.pl?prpsv=/ij/oecdthemes/99980002/v2010n7/s1/p1l.idx

[7] M. Q. Yang, A. Niemierko, J. Y. Yang, and Y. Jin, "Promoting inter/multidisciplinary education and research in bioinformatics, systems biology and intelligent computing." *I. J. Computational Biology and Drug Design*, vol. 2, no. 3, pp. 207–220, 2009. [Online]. Available: http://dblp.uni-trier.de/db/journals/ijcbdd/ijcbdd2.html#YangNYJ09

[8] J. Ni and J. J. Dongarra, Eds., *Interdisciplinary and Multidisciplinary Research in Computer Science, IEEE CS Proceeding of the First International Multi-Symposium of Computer and Computational Sciences (IMSCCS|06), June 20-24, 2006, Zhejiang University, Hangzhou, China, Vol. 1*. IEEE Computer Society, 2006. [Online]. Available: http://dblp.uni-trier.de/db/conf/imsccs/imsccs2006-1.html

[9] J. Younglove-Webb, B. Gray, C. W. Abdalla, and A. P. Thurow, "The dynamics of multidisciplinary research teams in academia," *The Review of Higher Education*, vol. 22, no. 4, pp. 425–440, 1999. [Online]. Available: http://muse.jhu.edu/journals/review_of_higher_education/v022/22.4younglove-webb.html

[10] F. S. Roberts, "The challenges of multidisciplinary education in computer science." *J. Comput. Sci. Technol.*, vol. 26, no. 4, pp. 636–642, 2011. [Online]. Available: http://dblp.uni-trier.de/db/journals/jcst/jcst26.html#Roberts11

[11] H. Dryburgh, "Underpresentation of girls and women in computer science: Classification of 1990s research," *J. Educational Computing Research*, vol. 23, no. 2, pp. 181–202, 2000.

[12] F. T. Piller, A. Vossen, and C. Ihl, "From social media to social product development: The impact of social media on co-creation of innovation (december 21, 2011)," *Die Unternehmung*, vol. 65, no. 1, 2011. [Online]. Available: http://ssrn.com/abstract=1975523

[13] J. Corbin and A. Strauss, "Grounded theory research: Procedures, canons, and evaluative criteria," *Qualitative Sociology*, vol. 13, no. 1, pp. 3–21, 1990.

[14] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *Journal of Information Science*, vol. 28, no. 6, pp. 441–453, 2002.

[15] S. Wasserman and K. Faust, *Social Network Analysis : Methods and Applications*. Cambridge University Press, 1994.

[16] M. C. Binz-Scharf, "Collaboration, communication, and control: The effects of ict-enabled innovation projects on informal organizational structures." in *HICSS*. IEEE Computer Society, 2005. [Online]. Available: http://dblp.uni-trier.de/db/conf/hicss/hicss2005-8.html#Binz-Scharf05

[17] H. Levesque and R. Brachman, "Expressiveness and tractability in knowledge representation and reasoning," *Computational Intelligence Journal*, vol. 3, pp. 78–93, 1987.

[18] P. Jacsó, "Differences in the rank position of journals by eigenfactor metrics and the five-year impact factor in the journal citation reports and the eigenfactor project web site." *Online Information Review*, vol. 34, no. 3, pp. 496–508, 2010. [Online]. Available: http://dblp.uni-trier.de/db/journals/oir/oir34.html#Jacso10b

[19] T. Verbick, "Women, technology, and gender bias," *Journal of Computing Sciences in Colleges*, vol. 17, no. 3, pp. 240–250, 2002.

[20] P. Li, S. H. Dau, G. J. Puleo, and O. Milenkovic, "Motif clustering and overlapping clustering for social network analysis." *CoRR*, vol. abs/1612.00895, 2016. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1612.html#LiDPM16

[21] J. Fox, *Applied regression analysis and generalized linear models*. Sage Publications, 2008, ch. Bootstrapping regression models, pp. 587–606.

[22] L. C. Freeman, "A set of measures of centrality based on betweeness," *Sociometry*, vol. 40, pp. 35–40, 1977.