Social Networks of Researchers and Educators on nanoHUB.org

Gerhard Klimeck
Network for Computational Nanotechnology, Purdue University, gekco@purdue.edu

George B. Adams III
Network for Computational Nanotechnology, Purdue University

Krishna P.C. Madhavan
Network for Computational Nanotechnology, Purdue University

Nathan Denny
Network for Computational Nanotechnology, Purdue University

Michael G. Zentner
Network for Computational Nanotechnology, Purdue University

See next page for additional authors

Follow this and additional works at: http://docs.lib.purdue.edu/nanopub
Part of the Nanoscience and Nanotechnology Commons

Klimeck, Gerhard; Adams, George B. III; Madhavan, Krishna P.C.; Denny, Nathan; Zentner, Michael G.; Shivarajapura, Swaroop; Zentner, Lynn K.; and Beaudoin, Diane L., "Social Networks of Researchers and Educators on nanoHUB.org" (2011). Birck and NCN Publications. Paper 902.
http://docs.lib.purdue.edu/nanopub/902

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Authors
Gerhard Klimeck, George B. Adams III, Krishna P.C. Madhavan, Nathan Denny, Michael G. Zentner, Swaroop Shivarajapura, Lynn K. Zentner, and Diane L. Beaudoin

This article is available at Purdue e-Pubs: http://docs.lib.purdue.edu/nanopub/902
Social Networks of Researchers and Educators on nanoHUB.org

Gerhard Klimbeck, George B. Adams III, Krishna P. C. Madhavan, Nathan Denny, Michael G. Zentner,
Swaroop Shivarajapura, Lynn K. Zentner, and Diane L. Beaudoin

Network for Computational Nanotechnology, Birck Nanotechnology Center, Purdue University, West Lafayette, IN, USA
{gekco, gba, cm, ndenny, mzentner, swaroop, lzentner, beaudoin}@purdue.edu

Abstract— The science gateway nanoHUB.org is the world’s largest nanotechnology user facility, serving 167,196 users in 2010 with over 2,300 resources including 189 simulation programs. Surveys of nanoHUB users and automated usage analysis find widespread simulation use in formal classroom education, thereby connecting recent research more rapidly and closely to education. Analysis of 719 citations in the scientific literature by over 1,300 authors to nanoHUB.org resources documents use of simulation programs by new research collaborations, by researchers outside of the community originating the program, and by experimentalists. The publication and author networks reveal research collaborations and capacity building through knowledge transfer. Analysis of secondary citations documents the quality of the conducted research with an h-index of 30 after just 10 years of operation. Our analysis proves with quantitative metrics that impactful research can be conducted by an ever growing research community. We argue that HUBzero technology and the user-focused design and operation of nanoHUB.org are success criteria that can be transferred to other science gateways.

Keywords—nanotechnology; simulation; citation network; nanoHUB; science gateway; cloud computing; grid computing.

I. INTRODUCTION

Nanoscience and nanotechnology are nascent areas that have not yet established standard commercial software packages, community software packages, or even in some areas the foundational theory that would need to be implemented in software. The community web site nanoHUB.org was created to fill that void by hosting advanced, research based simulation tools as well as tutorials, seminars, and courses. In the past 10 years the annual user numbers have grown from less than 500 to 167,196 users in the year 2010. nanoHUB.org is the world’s largest nanotechnology user facility.

In this paper we address several basic research questions: (1) Can centralized cyberinfrastructures (CIs) such as nanoHUB.org enable fundamental research? (2) Is it feasible to assess the quality of such research? (3) Can such CI engage independent research groups? (4) What are the fundamental mechanisms for propagating innovations seen within nanoHUB collaboration networks? (5) Can the use of cyber-environments such as nanoHUB lead to increased capacity building within a scientific domain such as nanotechnology? Answers to these questions may not be as relevant in mature areas of computational science such as solid mechanics, fluid mechanics, or quantum chemistry, because these areas already have access to commercial codes and established community codes. However, in emerging areas no such standard codes exist and the community largely relies on the construction of novel codes. Through answering the posed research questions we seek to understand if there are lessons learned that are transferrable to other Science Gateways in other frontier areas of science,

nanoHUB.org supports three broad categories of use – simulation, interaction, and downloading – with many users engaging with the site in multiple ways. Of the 2010 users, 9,805 logged on to run 372,404 interactive, graphical simulations of nanoscale phenomena, materials, and devices from within their web browser. Those users who are not logging on but who interact with the site for at least 15 minutes in a single session did so from 51,732 distinct internet addresses. Each of these interactive users spent on average about 4.5 hours on nanoHUB. Users from 144,314 addresses downloaded content. While new content is published continually, by the end of 2010 nanoHUB hosted the work of 741 authors in the form of some 2,300 content items, including 189 simulation programs, 46 undergraduate- and graduate-level courses, and 2,100 seminars and teaching materials.
	nanoHUB.org is a dual use science gateway that simultaneously serves education and research. Instructors are using nanoHUB.org in higher education, doing so in a known 487 classes at 158 institutions to date. An analysis of usage patterns in formal classroom training is the subject of a separate publication. nanoHUB.org users are conducting research as evidenced by their 719 published research papers that cite nanoHUB.org resources.

This paper focuses on the processes that transition research codes, written by small research groups, into education and into research of other groups. We explore the effects, broader use of scientific computing and research programs, and new forms of resource sharing.

Traditionally, social network analysis has been used in “studies of kinship structure, social mobility, science citations, contacts among members of deviant groups, corporate power, international trade exploitation, class structure, and many other areas” [1]. Valente [2] has proposed the use of social network threshold models for understanding the success or failure of collective action and the diffusion of innovations. The power of a CI is not only in its ability to facilitate research and education at an individual level but, more importantly, to allow groups of researchers to collaborate and share knowledge in new ways. Social network analysis has also been used as a methodology to show how a network of innovators forms within a problem
space [3]. Most studies using social network analysis as a methodology use centrality measures as the primary means for understanding the topology of the underlying networks. Borgatti [4] points out that “centrality measures, or at least popular interpretations of these measures, make implicit assumptions about the manner in which traffic flows through a network. For example, some measures count only geodesic paths, apparently assuming that whatever flows through the network only moves along the shortest possible paths.” Several studies [5,6,7,8] have discussed the nature of scientific collaboration networks. We use social network analysis as the theoretical basis for this paper.

In the context of large-scale cyber-environments such as nanoHUB.org, there are more fundamental problems that need to be addressed before a full social network analysis of the underlying collaboration networks can be undertaken. In our context, questions that take higher priority to address the fundamental need for our work include: (1) Is the CI facilitating fundamental research? (2) What is the quality of the facilitated research? (3) What types of innovations are seen within the network? This is precisely what this paper addresses.

II. BACKGROUND

nanoHUB.org [9,10] is operated by the NSF-funded Network for Computational Nanotechnology (NCN), which was founded in 2002 to support the U.S. National Nanotechnology Initiative (NNI). The rationale for NCN was that modeling and simulation are significantly underutilized in research and education because of obstacles to their deployment and difficulties in their use. To alleviate these issues, NCN created two key pieces of software. The first was the Rappture [11] toolkit for rapidly adding interactive graphical user interface to an unmodified simulation program, enabling exceptionally low cost web deployment. The second was HUBzero® [12], a platform to manage a wide variety of computational resources and connect them to a standard web server for delivery to any Flash- and Java-enabled browser. Both Rappture and HUBzero were created by the NCN and have been released as open source software. To date, Rappture has been used to make 219 programs web-ready, and in addition to nanoHUB, HUBzero is powering 18 HUBs focused on a variety of scientific fields. In combination, Rappture and HUBzero let nanoHUB.org provide a seamless user experience that connects advanced research simulation programs, powerful graphical user interfaces, and appropriate computational resources to a vast set of users. In addition, the centralized delivery of services allows nanoHUB.org to gather detailed usage data for the resources it hosts.

III. CITATION NETWORKS

The openness of the nanoHUB.org web site and informal, voluntary engagement with the users create a significant challenge to impact assessment. Citation of online resources in the scientific literature is still often informal. Despite introducing digital object identifiers (DOIs) for each simulation tool on nanoHUB and requesting that authors use these DOIs in the reference section of their papers, most authors continue to cite nanoHUB tools informally in the text (e.g. “We used the online simulation tool Schred on nanoHUB to ...”) or in a footnote (e.g. “[Simulation tool Schred on nanoHUB.org”). Finding such references requires a full text search on the whole paper which the widely-used literature search engines Web of Science or IEEE Xplore do not provide. Google Scholar [13] performs the required full text search on the complete publication and enables NCN personnel to find papers that cite nanoHUB.org.

For this paper a database was created from results of searching Google Scholar. The search string used combines “nanoHUB” with the some of the most popular tools: [“nanoHUB” OR “HUBzero” OR ”Schred” OR ”FETToy” OR “nanoMOS” OR “BioMOCA”]. This string is an evolving compromise designed to execute a search that uncover as many papers as possible while reducing the number of false positives. False positives are expensive because the search results (papers) are read to determine whether the paper a) does not cite nanoHUB or its content (false positive), b) cites nanoHUB or its content to provide context or background for the work of the paper, or c) cites nanoHUB or its content as contributing to or supporting the intellectual contribution made by the paper.

As of January 2011 our search and vetting has identified 719 papers that cite nanoHUB.org since the year 2000 (Figure 1). For brevity we will call the papers that cite nanoHUB.org resources “nanoHUB papers” and imagine nanoHUB.org as an active contributor to the research.

Each nanoHUB paper is further analyzed to classify it by topic, choosing from nanotechnology, cyber technology, education, and a mix of education and nanotechnology. Each paper is analyzed for the specific resource that it references, such as a simulation tool, a lecture, or computational resources. Nanotechnology papers are further analyzed to determine if they contain physical experimental data or involve an experimentalist. Computational researchers may use previously published experimental data to validate or test their novel simulation engine or test a modeling hypothesis. Such work relates immediately to experimental work, but does not influence it directly. The presence of an experimentalist in a published paper implies the design or execution of an experiment that may be assisted by nanoHUB resources. Such
resource usage constitutes a deeper engagement with and service to experimentalists.

Figure 1 shows the growth of the collection of nanoHUB papers and of the community of authors of these papers through the end of 2009. A strong growth in both the number of papers and the authors is clearly visible. We do not include the 2010 and 2011 papers in this graph, because at the time of the writing of this document in February 2011 not all 2010 papers have been archived in Google Scholar (or Web of Science or IEEE Xplore). Our experience is that these services lag publication dates by 4 to 6 months.

Figure 1 reveals that both the number of papers and the number of authors not affiliated with the NCN are larger than the NCN-affiliated numbers. In the definition of NCN affiliation we have chosen to be very conservative and deem a paper as NCN affiliated if there is an NCN affiliated co-author. Interestingly the ratios of total authors to total papers remains steady at 1.9 over 11 years indicated by the identical slopes of the blue curves in Figure 1.

Papers citing nanoHUB.org linked by common author

Figure 2. Network of 719 scientific publications that cite nanoHUB.org and its content as contributing to the paper’s intellectual content. Dots represent papers; lines connect papers with common authors. Papers are categorized by inside/outside NCN (+ and *) and field of science: nanoresearch (red), cyberinfrastructure (green), education (blue), and mixed education/nano (purple). Paper networks independent of NCN are clearly visible.

Figure 2 shows the 591 papers in Figure 1 plus additional papers from 2010 and 2011 (total 719) color-coded by research topic. A pair of papers is linked by a line when there is at least one author in common. Papers with one or more NCN-affiliated authors are shown with + signs; those with no NCN-affiliated authors by dots. NCN papers are strongly networked (high connection density shown in light gray). However, research networks independent of NCN are developing as shown by the linkages between papers represented by dots.

An NCN goal is service to experimental researchers. Extending the use of simulation tools from computational experts to experimentalists is a critical item in translation of nanoscience to nanotechnology. Such translation as well as a change of culture be observed in the nanoHUB citation map. Experimental data are contained in 25% of the 605 nano research publications (Figure 3). Experimental researchers drive 13% of these papers.

One of the main open questions for centralized CI is can it enable research and education within a problem space – in our case, nanoscale science and engineering. The 719 papers published by over 1,300 authors provide significant evidence that research can be conducted and enabled within the context of an open community resource. The fact that over 79% of these authors are not affiliated with NCN indicates that a broadening of participation beyond the community founders is possible. The relationship of the nano research papers with experimental data and the presence of experimentalists on the list of co-authors demonstrate that nascent nano research codes can transition from computational research into practical experimental usage. The next question is whether this novel approach to research is productive.

Papers citing nanoHUB.org for nano research

Figure 3. Network of 605 scientific publications that reference nanoHUB for nano research. 25% of these publications contain experimental data, and 13% of these publications can be identified as experimental science.

One measure of the productivity and impact of a research effort is its $h$-index [16]. A set of $S$ research papers has index $h$ if $h$ papers of $S$ have been cited by other publications at least $h$ times. Through Google Scholar we have identified over 3,200 secondary citations to the 719 nanoHUB papers with a distribution of citations such that the $h$-index of the set is 30. The personal $h$-index for a typical academic researcher increases by one with each year. Thus, a professional with 10 years of experience may be expected to have an $h$-index of 10. The first papers citing nanoHUB appeared in 2000. In 11 years the nanoHUB $h$-index has increased more rapidly than a typical researcher and shows the nature of the impact of the community formed around it. We are not aware of a similar evaluation of a science gateway and are not sure if this is a particularly high or low $h$-index. However, we do claim that this $h$-index indicates that the nanoHUB papers are appreciated by subsequent authors using that work in their citations.

Figures 2 and 3 together show that over and above enabling fundamental research in nanoscale science, nanoHUB also has an impact on associated fields such as CI and education. This analysis shows that nanoHUB has enabled the growth of a community around the core services it offers. The growth of such communities is critical for the propagation of research and educational innovations. Figures 2 and 3 not only show a strong core, but also the growth of weak links [14]. Therefore, the topology of the resulting network is highly suitable for enabling the rapid diffusion of scientific innovations [15].
IV. TOOL NETWORKS

Networks may also be defined by the resources researchers use. Nanoelectronics is one example of a research community that uses simulation to advance the field but that has not had access to commercial simulation software. The relatively small number of researchers in the field, the nascent nature of the understanding of electron transport within small devices, and the hotly-debated choices of computational and modeling approaches all prevented the formation of commercially viable simulation software. Consequently, a culture of research groups developing their own software arose, and with it, a value system in which the hallmark of important research was its foundation on a purpose-built simulation tool authored by the researchers themselves. The perception was that good research cannot be conducted with someone else’s tool.

nanoHUB.org has changed the way some nanoelectronic device researchers work and has changed their culture. Today, simulation tools are being shared among independent research groups. Below we give examples of tools i) developed in the NCN that have impact outside of NCN and ii) contributed from outside NCN that have impact inside and outside NCN.

Papers citing nanoMOS on nanoHUB.org

- one or more NCN-affiliated co-author(s) → no NCN-affiliated co-author

| 94 total |
| 35 NCN-affiliated |
| 59 not NCN-affiliated |

Cited 94 times, the tool NanoMOS [17] is a good example of an impactful NCN-developed tool. Since the year 2000, nanoHUB data show that NanoMOS has had 1,700 users who have logged 22,000 simulation runs. NanoMOS has undergone several stages of development. It is written in Matlab and simulates current flow using the NEGF formalism. Of the 94 citations, 35 are by NCN-affiliated authors and 59 are by authors unaffiliated with NCN (Figure 4). In many of the unaffiliated papers the authors state that they downloaded NanoMOS source code to study the NEGF formalism before applying it to their particular study, or to establish a baseline from which to add new capability. The analysis of the cited papers has shown that the release of the NanoMOS code as open source and the usability on the web has helped some authors to build, benchmark, test, improve, or expand their own tools. As such nanoHUB and NanoMOS helped to establish NEGF as the standard quantum transport theory used in the field of nanotransistor research.

Most of the NCN affiliated-author papers document NanoMOS improvements that resulted in new insights into electron flow through nanotransistors. As the code matures, new code elements are migrated into the public version and published on nanoHUB. nanoMOS now has a history of ten years of continual improvements and publications by the NCN team. The innovation propagation model seen here is inside-out, meaning that the innovation occurred within the core network and spread to the broader community.

The tool Schred [18] boasts the greatest number of citations among all nanoHUB tools. It was contributed to nanoHUB.org from outside NCN and is primarily serving a network of authors outside NCN (see Figure 5). While Schred does not embody the most sophisticated theory, and thus appeals to a few theorists, it is exactly what is needed by experimentalists to calibrate their experiments. As such, the deployment of Schred on nanoHUB is clearly a community service. Another critical element is that Schred is relatively easy to use and requires about 1 minute to execute on a recent Intel/AMD CPU. In 2007 alone, 464 users ran 8,107 Schred simulations (including classroom use) resulting in 16 publications. In all, Schred has 1,850 users who have run 59,000 simulations. We also note here that the heavy community use of codes developed outside the core networks propagates through the weak networks seen in Figures 2 and 3; an example of outside-in propagation of a research innovation.

Papers citing Schred on nanoHUB.org

- one or more NCN-affiliated co-author(s) → no NCN-affiliated co-author

| 104 total |
| 12 NCN-affiliated |
| 92 not NCN-affiliated |

Analysis of nanoHUB papers reveals other instances where external theory-driven researchers used a nanoHUB tool that was contributed from outside of NCN. For example, Kureshi and Hasan, working at a university in India, published their study [19] in which they conducted over 1,600 runs of the simulation tool Carbon Nanotubes Interconnect Analyzer (CNIA [20]). CNIA was written and contributed by Tanachutiwat and Wang of the University of Albany. We do not know of any relationship between these two research teams other than through CNIA. CNIA is also cited in a University of Cincinnati Master’s thesis from 2005. CNIA demonstrates how members of the broad nanoHUB community outside of NCN have shared (published) their program, and it was found useful by unrelated research groups.
V. AUTHOR NETWORKS

There are 1,300 authors of the 719 nanoHUB papers. Here we study the “genealogy” of authors and the temporal expansion of authorship.

Figure 6 depicts each author with a red or green vertical sliver, indicating affiliated or not affiliated NCN, respectively. Each author is Author symbols are plotted in the year of their first nanoHUB publication. Co-authors are connected via joint nanoHUB papers in different years. Connections within a single year are omitted. Color and width of the connecting lines were chosen to improve the ability to distinguish between different years of collaborations.

It is interesting to note that the first nanoHUB papers came from outside the NCN! Further, the authors of the first papers have continued to publish throughout the life of nanoHUB. Indeed there are several subsequent authors that continue to collaborate on publications throughout the years.

VI. DUAL USE IN RESEARCH AND EDUCATION

Schred and nanoMOS are particularly interesting nanoHUB tools because of their documented dual use: 1) they are used for education at the graduate and undergraduate level in classroom demonstrations and for homework or project assignments and 2) they are highly referenced, with 104 and 94 citations in the scientific literature, respectively. Both Schred and NanoMOS originated in research and transitioned into education and from the original researchers into other people’s research. The usage in the classroom has been positively identified by faculty interviews and by usage pattern analysis.

It is significant to note that applications that would be considered “educational” by some are used in research work by others, and the results are published in high quality journals. The tool CNTbands was originally envisioned as an educational classroom tool. It is however also cited 14 times in the literature, with 10 citations from non-NCN-affiliated users. We believe that such dual use illustrates not only the transition of technology from research into education, but another dimension of how the social network of nanoHUB.org grows by educating some of its future research participants.

VII. NANOHUB USER COMMUNITY FAST FACTS

- User location
  - 35% from the United States
  - 33% from Asia
  - 23% from Europe
  - 91% of users affiliated with an academic institution

- Simulation tool use: 60% of simulation tool users are US based, accounting for 70% of all simulation runs

- Educational Reach in the US:
  - 100% of top 50 Engineering schools reached (U.S. News and World Report)
  - 99% of Carnegie RU/VH schools reached
  - 95% of Carnegie RU/H schools reached
  - 25% of all minority serving institutions granting degrees in STEM fields reached

VIII. NANOHUB.ORG AS A WEB CITIZEN

Many web sites have been the subject of social network study, including for example Facebook, LinkedIn, and Wikipedia. Open questions remain as to how such popular social sites will interact with each other and with science.
gateways. nanoHUB.org has begun an exploration of such interactions through participation in iTunes U and Wikipedia. While it is still early to report on the impact these services have on the social network fostered by nanoHUB.org, our initial results indicate that it is likely to have some effect: i) as of December, 2010 nanoHUB tracks were downloaded an average of 1200 times per week from iTunes U. Beyond Campus and 8000 subscription feeds were established and ii) 16 animated GIF images were contributed to add value to 31 relevant Wikipedia articles that are viewed about 25,000 times per day. These animated images grew the number of visits to nanoHUB from Wikipedia from a baseline of 200 to over 4,000 per month since December 2010. Experience with iTunes U and Wikipedia demonstrates that there is a valued need for nanoHUB content in these large arenas.

IX. CONCLUSIONS

We have shown that nanoHUB.org has fostered a social network. nanoHUB.org data show that the distinction between research and the educational use of nanoHUB.org content is blurring. The role of nanoHUB.org in research has shown connections between scholarly authors both inside and outside of NCM. This network has grown over time, and has shown translation of research related codes to practical experimental use. Research clusters completely unaffiliated with NCM have formed. We have shown in addition to the networks exhibited in the published literature, that the use of nanoHUB.org simulation programs also implies connections between otherwise unrelated researchers. Further, patterns in tool use have proven the migration of research codes into graduate and undergraduate education. We have found a substantial number of citations in the scientific literature (42, comprising 7% of the total) that discuss nanoHUB use in education, proving that nanoHUB.org is impacting research on nanoscience in education and that the site has played a vital role in developing researchers.

By bringing tools to one common place for the nanotechnology community, nanoHUB.org is helping to migrate tool use from the originating nanotechnology sub-domain community to others. This collaborative effort enables work that would not have occurred otherwise. nanoHUB.org is broadening its reach and impacting new communities.

Given time-efficient access to cutting-edge resources, persons at all knowledge levels, from student to expert, are able to produce relevant research and travel further in their intellectual journeys. Some will even find that what began as education can seamlessly take them into the territory of research. In this sense, the nanoHUB.org environment has great potential for simultaneously advancing research while preparing the workforce and continuing its education.

ACKNOWLEDGMENT

Mark S. Lundstrom founded nanoHUB.org in 1998. In 2005, Michael McLennan created the Rappture Toolkit and Rick Kennell wrote the scalable middleware of HUBzero that, respectively, enable and power interactive nanoHUB simulations. Search, vetting, and categorization of the 719 papers citing nanoHUB and its content was performed by Michael J. Anderson, Denis Areshkin, Philatlia Rufaro Bolton, A. Arun Goud, Yu He, SungGuen Kim, Saumitra R. Mehrotra, Kai Miao, Margaret Shepard Morris, and Margarita Shalaev.

REFERENCES

[13] scholar.google.com