

1-28-2015

COLLABORATING TO RUIN? US NATIONAL LABORATORIES AND THE IMPACT OF INTERNATIONAL RESEARCH PARTNERSHIPS

Gerald Hendrickson

Follow this and additional works at: http://digitalrepository.unm.edu/econ_etds

Recommended Citation

Hendrickson, Gerald. "COLLABORATING TO RUIN? US NATIONAL LABORATORIES AND THE IMPACT OF INTERNATIONAL RESEARCH PARTNERSHIPS." (2015). http://digitalrepository.unm.edu/econ_etds/18

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Economics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Gerald Hendrickson

Candidate

Economics

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Alok Bohara , Chairperson

Richard Santos

Jennifer Thacher

Elshan Akhadov

**COLLABORATING TO RUIN? US NATIONAL LABORATORIES AND THE
IMPACT OF INTERNATIONAL RESEARCH PARTNERSHIPS**

BY

GERALD A. HENDRICKSON

B.A., Economics/Political Science, University of New Mexico, 1995

M.A., Economics/Political Science, University of New Mexico, 1998

M.P.A., Public Budgeting and Finance, University of New Mexico,

2000

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctorate of Philosophy

Economics

The University of New Mexico
Albuquerque, New Mexico

December 2014

© 2014, Gerald Hendrickson

DEDICATION

To Pop

and

To Stacey, Celine and Kit

and their varying levels of patience with this project

– near infinite, near zero and absolutely zero.

ACKNOWLEDGMENTS

“Hope deferred makes the heart sick, but a longing fulfilled is a tree of life.”

Proverbs 13:12

As many successful PhD candidates before me have noted, completing a dissertation is more an act of persistent humility than intelligence. Yielding to temptation often steals the momentum discipline generates. To complete this work, one must make guilt and accountability a constant associate, haunting every other activity done. Often, only faith remains - of others in you.

Few dissertations have experienced the number of crises that have swirled around this one. Thus, special thanks goes to Alok Bohara, the chair of my dissertation committee. Dr. Bohara showed uncommon patience in shepherding this project to completion.

The long path this work took to publication was necessary. I learned to “really” program (thanks to Michael Hannah, whose office I now occupy), brief decisionmakers (thanks to Bruce Held) and visualize data effectively (thanks to Nabil Rahal) along the way. The exposure to computational linguistics that helped me solve some of my key problems in performing this research came out of months of working with my wife Stacey on her dissertation.

Thus, this work of paper and ink is not mine alone. Many other people have invested in this success - from my sister, Lisa, who taught me as a child, to Kishore Gawande, who taught me the importance of ethical econometrics. For everyone who contributed, I hope this document is evidence of a positive return on that investment.

**COLLABORATING TO RUIN? US NATIONAL LABORATORIES AND THE
IMPACT OF INTERNATIONAL RESEARCH PARTNERSHIPS**

BY

GERALD A. HENDRICKSON

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctorate of Philosophy

Economics

The University of New Mexico
Albuquerque, New Mexico

December 2014

COLLABORATING TO RUIN? US NATIONAL LABORATORIES AND THE IMPACT OF INTERNATIONAL RESEARCH PARTNERSHIPS

By

Gerald A. Hendrickson

B.A., Economics/Political Science, University of New Mexico, 1995

M.A., Economics/Political Science, University of New Mexico, 1998

M.P.A., Public Budgeting and Finance, University of New Mexico, 2000

ABSTRACT

Following the Cold War, Russian and US research institutions forged new collaborative ties to take advantage of perceived complementarities in conducting scientific research as part of US nonproliferation initiatives. These ties appear to have been successful in the broader nonproliferation context as relatively few Russian nuclear scientists emigrated to perceived rogue states like Iran and North Korea in the years that immediately followed the dissolution of the Soviet Union. Early on, the research benefits of these ties appeared to be significant. Today, as the Russian science and technology cadre is going through a demographic transition and the Russian state is following a corporatist policy in rebuilding its scientific research and development base, the appropriable benefits associated with continuing these policies for US research partners are less obvious.

This assessment is an attempt to gain an empirical understanding of the appropriable benefits from US-Russian research engagement apart from the nonproliferation context. As such, this study examines these collaborations using an alternative network analysis methodology with reference to a knowledge-based model of research and development generation. To assure tractability, the analysis focuses its attention on a subset of institutions that have been broadly ignored in studies of research collaboration – US national laboratories and their Russian counterparts.

The resulting analysis challenges the conventional wisdom of the appropriable virtues of scientific collaboration. For the limited set of relationships examined in this study, this analysis suggests participation in international collaborations between the largest US national laboratories and their Russian counterparts can actually reduce individual researcher's basic research productivity – clearly not a policy goal for a major national research and development establishment. To achieve better appropriability, this finding and its contextual factors are used to demarcate areas of inquiry where Russian-US engagement has an empirical track record of utility and should continue from areas where collaboration has had little success.

Table of Contents

| | |
|--|-----|
| Table of Figures | xi |
| List of Tables | xiv |
| I. The Dilemma of International Collaboration for US National Laboratories | 1 |
| The Rise of Collaboration..... | 1 |
| National Laboratory Collaboration in Nuclear Fusion Research | 3 |
| Risks from Increased International Collaboration | 12 |
| Finding a Balance | 18 |
| A Note on Methods | 21 |
| Overview of this Document | 23 |
| II. Modeling Research Production from a Knowledge Perspective: A Second Best Approach..... | 26 |
| National Laboratories: Practically Important, Theoretically Inconvenient | 26 |
| The Problem Space..... | 37 |
| Modeling Research Productivity | 40 |
| III. National Laboratories, Knowledge and the International Collaboration Interface | 51 |
| The Conventional Wisdom Behind Collaboration..... | 51 |
| Selection of Collaborations of Interest | 56 |
| Knowledge Network Attributes and Transfer Dynamics..... | 58 |
| Institutes, Research Communities and Collaborations | 64 |
| The Kurchatov Institute of Atomic Energy | 65 |
| All-Russia Institute for Experimental Physics (VNIIEF)..... | 78 |
| All-Russia Institute for Theoretical Physics | 85 |
| D.V. Efremov Institute | 91 |

| | |
|--|-----|
| P.N. Lebedev Physical Institute | 96 |
| Educated Expectations | 100 |
| IV. Collaborating to Ruin? Measured Productivity Benefits of US-Russian | |
| National Laboratory Collaborations | 102 |
| Putting Expectations to the Test | 102 |
| Constructing Institutional Publication Record Sets | 104 |
| Organization and Data Processing | 108 |
| Operationalization of Variables | 109 |
| Empirical Analysis | 120 |
| Results | 126 |
| Findings | 135 |
| Conclusions | 138 |
| V. Concluding Remarks | 140 |
| The Changing Mission Mix | 140 |
| Networks as the Analytic Substrate | 145 |
| Knowledge and its Public Good Character | 147 |
| Collaboration Choices | 150 |
| Improving Appropriability | 151 |
| Appendix 1: Pre-processing codes | 154 |
| Reformatting Scripts | 155 |
| Visualization Enabling Scripts | 156 |
| Fruchterman-Reingold layout algorithm (Fruchterman and Reingold, 1991) | 159 |
| Automated Panel Dataset Construction | 159 |
| Appendix 2: Stata do.file | 169 |
| References | 184 |

Table of Figures

| | |
|--|----|
| <i>Figure 1. Joint Major Partner University Publications with Russian Institutions (1977-2012)</i> | 2 |
| <i>Figure 2. Publication Output by Selected US National Laboratory and Major Partner University (Journal Publications in 2007)</i> | 27 |
| <i>Figure 3. Patent Output by Selected US National Laboratories and Industrial Research Organizations (Patent Applications Issued, 2007)</i> | 28 |
| <i>Figure 4. Average Publication Citations, US National Laboratories and Major Partner Universities (Citations of 2004 Publications)</i> | 29 |
| <i>Figure 5. Average Patent Citations, US National Laboratories and Major Industrial Partners (Citations of Patents Granted in 2004)</i> | 30 |
| <i>Figure 6. R&D 100 Awards, US National Laboratories and Universities (1990-2007)</i> | 32 |
| <i>Figure 7. Kurchatov Institute of Atomic Energy Research Community, 2012</i> | 66 |
| <i>Figure 8. KIAE Open Research Activities, 2008-2012</i> | 67 |
| <i>Figure 9. Los Alamos-Kurchatov Collaboration Pairs (1977-2008)</i> | 68 |
| <i>Figure 10. Los Alamos-Kurchatov Joint Scientific Publication Coauthorship Networks (1977 -2008)</i> | 69 |
| <i>Figure 11. Los Alamos-Kurchatov Joint Collaboration Degree Distribution</i> | 70 |
| <i>Figure 12. Lawrence Livermore - Kurchatov Collaboration Pairs (1977-2008)</i> ... | 72 |
| <i>Figure 13. Lawrence Livermore-Kurchatov Joint Scientific Publication Coauthorship Network (1977-2008)</i> | 73 |

| | |
|--|-----------|
| <i>Figure 14. Lawrence Livermore -Kurchatov Joint Collaboration Network Degree Distribution (1977-2008).....</i> | <i>74</i> |
| <i>Figure 15. Sandia-Kurchatov Collaboration Pairs (1977-2008)</i> | <i>75</i> |
| <i>Figure 16. Sandia-Kurchatov Joint Scientific Publication Coauthorship Network (1997-2008)</i> | <i>76</i> |
| <i>Figure 17. VNIIEF Network Neighborhood, 2012.....</i> | <i>78</i> |
| <i>Figure 18. VNIIEF Latent Semantic Topology, 2012.....</i> | <i>79</i> |
| <i>Figure 19. Los Alamos-VNIIEF Joint Scientific Coauthorship Network (1977-2008).....</i> | <i>81</i> |
| <i>Figure 20. Los Alamos-VNIIEF Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008).....</i> | <i>82</i> |
| <i>Figure 21. Lawrence Livermore -VNIIEF Joint Scientific Publication Coauthorship Network (1977-2008).....</i> | <i>83</i> |
| <i>Figure 22. VNIITF Research Community, 2012.....</i> | <i>86</i> |
| <i>Figure 23. Latent Semantic Topology of VNIITF Research, 2012</i> | <i>87</i> |
| <i>Figure 24. Los Alamos-VNIITF Joint Scientific Publication Coauthorship Network (1977-2008)</i> | <i>88</i> |
| <i>Figure 25. Lawrence Livermore-VNIITF Joint Scientific Publication Coauthorship Network (1977-2008).....</i> | <i>90</i> |
| <i>Figure 26. DVEI Research Community, 2012</i> | <i>92</i> |
| <i>Figure 27. Latent Semantic Topology of DVEI Research, 2008-2012</i> | <i>93</i> |
| <i>Figure 28. Sandia-Efremov Institute Joint Scientific Publication Coauthorship Network (1977-2008)</i> | <i>94</i> |
| <i>Figure 29. Sandia-Efremov Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008).....</i> | <i>95</i> |

| | |
|--|-----------|
| <i>Figure 30. PLPI Research Community, 2012</i> | <i>96</i> |
| <i>Figure 31. Latent Semantic Topology of Research at PLPI, 2008-2012</i> | <i>97</i> |
| <i>Figure 32. Sandia-Lebedev Joint Scientific Publication Coauthorship Network (1977-2008)</i> | <i>98</i> |
| <i>Figure 33. Sandia-Lebedev Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008)</i> | <i>99</i> |

List of Tables

| | |
|--|------------|
| <i>Table 1. Professional Society Leadership, by Society and Institutional Type, 2008</i> | <i>40</i> |
| <i>Table 2. Data Table</i> | <i>117</i> |
| <i>Table 3. Pooled GEE Models – US, SNL, LLNL, LANL</i> | <i>127</i> |
| <i>Table 4. Sensitivity Analysis for the US Model – NB (FE), GEE, “IV” GEE</i> | <i>129</i> |
| <i>Table 5. “IV” GEE Models – Pooled US, SNL, LLNL, LANL</i> | <i>130</i> |
| <i>Table 6. Pooled GEE Models – Russia, KIAE, VNIIEF and VNIITF</i> | <i>131</i> |
| <i>Table 7. Nuclear Fusion – US, LLNL, SNL and LANL GEE Models</i> | <i>133</i> |
| <i>Table 8. Rest of Research Fields – US, LLNL, SNL, LANL</i> | <i>134</i> |

I. The Dilemma of International Collaboration for US National Laboratories

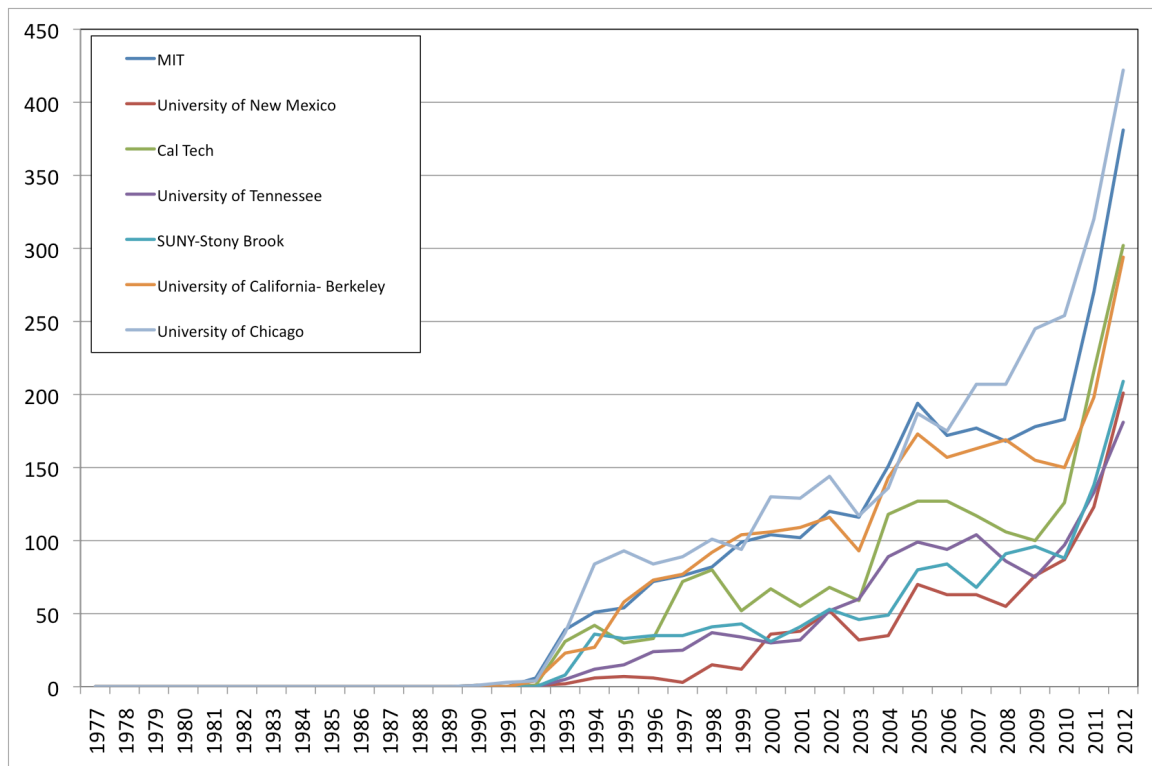
The Rise of Collaboration

Cross-border collaboration is now routine within the global scientific community (UNESCO, 2010; National Science Board, 2008 and The Royal Society, 2011). This trend extends to China and India as well as in the traditional scientific and engineering powers in North America and Western Europe (The Royal Society, 2011). Within the United States, international collaborations as indicated by co-authorships increased for all research and development types during the period 1995-2005 (National Science Board, 2008). There is only one exception to this trend among the G8 countries – Russia. From 2002 to 2008, the number of papers international co-authors produced with Russian collaborators dropped from 8884 to 8788 (UNESCO, 2010).

In the United States, this expansion of international collaboration was led by the academic sector. Colleges and universities increased their rate of

international co-authorship relative to all other US research and development institution types from 1995 to 2005 (National Science Board, 2008). This increase in co-authorship extended to US university collaborations with Russian scientists as well. Joint publications between Russian scientists and faculty members at major US national laboratory partner universities rose significantly following the dissolution of the Soviet Union (shown in Figure 1 below).

Figure 1. Joint Major Partner University Publications with Russian Institutions (1977-2012)



Unlike the United States, government-funded research institutes in Russia led both the initiation and movement to engage in international research collaboration efforts following the dissolution of the Soviet Union. Neither industry nor the academic sector played a major role in Russian research and development efforts in 1990 (UNESCO, 1993). Surprisingly, this distribution of

research and development performance has not changed since that time (UNESCO, 2010).

National Laboratory Collaboration in Nuclear Fusion Research

National laboratories in the US have only made a practice of heavy collaboration with other nations' national research facilities in the past thirty years. Prior to that point, scientific collaboration between national laboratories resulting in publications in peer-reviewed journals was limited even between national laboratories of closely allied countries (e.g., DOE laboratories in the United States and similar British facilities). The limited technical collaboration that existed between national laboratories in the United States and similar laboratories in what was then the Soviet Union or in China was subject to intense political scrutiny. This later set of linkages has only emerged over the last twenty years - since the end of the Cold War.

The broad rise in collaborations between the US and Russian research communities after the Cold War shown in Figure 1 can be traced to two complementary trends. Following the dissolution of the Soviet Union, the US scientific community "pushed" to collaborate with Russian research institutes motivated by the idea that research institutes behind the "Iron Curtain" possessed unique capital assets and a different public knowledge base than possessed by Western scientists and engineers. As Figure 1 shows, this push to collaborate with Russian scientists is seemingly insensitive to the short-run international political context, the number of joint publications with Russian co-authors only took short dips during periods of international conflict (i.e. the Russian incursion into South Ossetia in 2008).

The US national security establishment also funded collaborative basic research within the Russian closed cities that played host to the Russian nuclear weapons laboratories and “pulled” scientists at US universities, US national laboratories and Russian research institutes together to perform that work. At the time, US policymakers hoped by funneling basic research funding to Russian research institutes that Russian nuclear weapons scientists could be encouraged to remain in Russia and not migrate to nation-states that aspired to nuclear weapons state status.

The US scientific community pushing to collaborate with Russian scientists based its interest on firm historical ground. In the immediate years following World War II, nuclear fusion research in the United States, Great Britain and the Soviet Union took place under the constraint of strict secrecy (Artsimovich 1958, Bishop 1958). Despite this secrecy, notable similarities existed between the programs of inquiry could be observed between the nation-states when this constraint was lifted in September 1958 for the Second International Conference on the Peaceful Use of Atomic Energy in Geneva. (Bromberg 1983, Shafranov 2001) Despite these similarities, each nation-state focused the attention of its scientific community in a different manner and made significant complementary contributions to the pursuit of controlled fusion following declassification. Ultimately, a Soviet design concept – the tokamak – became the central concept of interest in nuclear fusion, but only after the concept acquired better diagnostics from Great Britain and various design enhancements (e.g. plasma divertors, neutral beam heating) from the myriad of research efforts funded in the United States (Clery 2013, Dean 2012).

The observed complementarity of the two research communities in the past was clearly behind the push for US national laboratories to re-engage in collaborative work with their Russian colleagues. This past cooperation influenced the pace at which US national laboratories sought re-engagement. This new collaboration began with technical meetings in the late 1980s made possible by Gorbachev's policy of *glasnost*.

Los Alamos National Laboratory re-engaged in significant levels of international collaboration with Russian research institutes first of the three DOE/NSA nuclear weapons design laboratories. This push to collaborate with their Russian colleagues is consistent with the storied history of the institution. The contributions of Niels Bohr (Denmark) and James Chadwick (UK) to the Manhattan Project era effort to build the first nuclear fission based bomb are well known. Since its foundation during the Manhattan Project, Los Alamos has maintained a deep commitment to international collaboration as a key element of conducting world-class scientific research in nuclear physics.

This commitment can be observed in the way that Los Alamos developed its nuclear fusion research programs. James L. Tuck, a British scientist, actually led Los Alamos' nuclear fusion research efforts during the period when the existence of these activities (under what was known as Project Sherwood) was "born classified" (Bishop 1958). Humorously enough, the US did not even acknowledge the existence of the program itself to the UK until 1956. By that time, Tuck had been operating his "Perhapsatron" – a "pinch" concept fusion device – for three years (Bromberg 1983).

Despite its commitment to international engagement in this field, however, Los Alamos never received the same level of investment for nuclear

fusion research that other national laboratories received. Two issues led to this lack of commitment. First, owing to the influence of Tuck, Los Alamos was most well known in the plasma physics community for its work on magnetic “Z-pinch” fusion devices. This Los Alamos’ effort was mirrored and amplified at the Atomic Energy Research Establishment (also known as Harwell) in the UK. Both Los Alamos and Harwell’s devices – the Perhapsatron and the ZETA – exhibited significant plasma instabilities, which limited their ability to reach fusion level temperatures. Z-pinch concept devices largely fell out of favor after operators of the ZETA machine at Harwell made public claims of ZETA achieving fusion in 1958 that had to be retracted soon after.

Harwell received a much more substantial backlash than Los Alamos from this public failure. Wisely, Los Alamos had insisted upon a small-scale research program using small and medium size experimental facilities devoted to developing a theoretical understanding of the physics of plasmas prior to developing large experimental facilities resembling possible nuclear fusion power plants. As a result, the “Perhapsatron” was inexpensive and actually built using discretionary laboratory funds. Despite this commitment to agility, Los Alamos did not have another concept in waiting in case the Z-pinch concept failed and funding flowed to stellarator (Princeton), magnetic mirror (Lawrence Livermore) and ion beam (Oak Ridge) concepts instead. Los Alamos’ experimental fusion efforts never really recovered from this early setback.

Lawrence Livermore National Laboratory (LLNL), Los Alamos’ key US competitor in the nuclear physics community, did not have this same set of reservations about the construction of large scale facilities. As a result, LLNL became the center of gravity for US nuclear fusion research as nuclear fusion

research transitioned from bench top scale experiments to power plant size experimental facilities. Unfortunately, LLNL did not possess a similar depth of commitment to international collaboration to Los Alamos.

Unlike Los Alamos, LLNL did not have the same level of early positive feedback from its interactions at international conferences in nuclear fusion research - where most early international collaboration in the field took place. LLNL's urge to find international research partners in the area experienced a significant setback when researchers from the Kurchatov Institute caustically ridiculed plasma temperature and confinement results from LLNL's "magnetic mirror" machines presented at the first International Atomic Energy Agency (IAEA) Fusion Energy Conference in Salzburg, Austria in 1962. The LLNL researchers adopted the Kurchatov Institute suggestions for improving their mirror machines, but not without significant bitterness about the manner in which they were humiliated on the international stage. This bitterness inclined LLNL fusion researchers to see Russian researchers more as competitors than collaborators (Herman 1990).

Any attempt at broader collaboration between LLNL and Russian researchers took another step backward when LLNL changed its nuclear fusion research program in response to a US Department of Energy (DOE) Office of Energy Research decision to focus its resources on funding the International Thermonuclear Experimental Reactor (ITER) instead of continuing to develop competing approaches to the tokamak concept (Clery, 2013). In what was perceived as a significant scandal at the time, DOE closed down the just completed Mirror Fusion Test Facility (MFTF-B) at LLNL in 1986 the day after it was completed to avoid paying for the operation and maintenance of the test

facility. In reaction, LLNL began pursuing inertial confinement fusion research – funded by the Office of Defense Programs within DOE – with greater vigor.

LLNL had been working on this non-conventional approach for quite some time. LLNL nuclear weapons designer John Nuckolls, the leading pioneer of the inertial confinement approach to nuclear fusion, began his work in the subject in the mid-1950s. As such, LLNL had been performing work in this area for thirty years at the point this emphasis shifted.

Thus, the LLNL fusion research program shift from focusing on mirror machines to laser driven inertial confinement fusion schemes was not as a radical of a change in philosophy as it might appear from outside the US nuclear weapons laboratories. However, it did reduce LLNL's international collaboration in the field of nuclear fusion topics even further. Given its perceived potential to illuminate nuclear weapons research and development, inertial confinement fusion research in the United States was subject to classification under the Atomic Energy Act of 1956. As a result, a greater number of LLNL nuclear fusion researchers became subject to security measures that constrained in their interactions with all former colleagues outside the US nuclear weapons laboratories, including their former Russian colleagues.

Progress in the field was slow. However, these problems persisted after the perceived end of the Cold War, when many of the classification restrictions associated with inertial confinement fusion research were lifted and collaboration became more commonplace. Many of the challenges associated with conventional approaches to nuclear fusion research were echoed in the inertial confinement fusion domain. As elsewhere, LLNL inertial confinement fusion researchers observed hydrodynamic instabilities in their attempts to

implode tiny pellets of hydrogen with lasers that were similar to the instabilities documented by earlier researchers.

The constraints on collaboration are often perceived as limiting the ability of the LLNL nuclear fusion research program to generate inertial confinement fusion breakthroughs because of the increased isolation of these researchers from former colleagues in the conventional nuclear fusion research community. This perception, however, ignores the issue that one of the critical enabling technologies for this approach to nuclear fusion – lasers – lacked the technology maturity associated with many of the tokamak alternative schemes explored in the 1950s and 1960s. The burst of large-scale facilities in the U.S. housing innovative lasers occurred about fifteen years after a similar rapid expansion of facilities exploring the tokamak approach to controlled nuclear fusion took place.

Over the last forty years, LLNL built a series of increasingly more powerful laser facilities in pursuit of inertial confinement fusion research (e.g. Cyclops, Shiva, Nova). The largest and most well known of these facilities is known as the National Ignition Facility (NIF), which went into full-service in May 2009. Built to support the development of nuclear weapons science in a world without destructive nuclear weapons tests, NIF houses a neodymium glass laser, which is the most powerful laser currently known to be in existence.

In the development of NIF, LLNL embraced – some might say finally – international collaboration. It was built with assistance from the other US nuclear weapons laboratories (e.g. pulsed power modules from Sandia), France (beam combiners), Israel (pulse shaping) and technical consultation from Russian fusion researchers. Despite this significant level of collaborative effort, NIF has yet to achieve full ignition of the hydrogen pellet.

Despite falling short of this technical milestone to this point, LLNL is reaping some of the benefits associated with building a “big science” facility of this scale and notoriety. ITER, the international thermonuclear experimental reactor in Cadarache, France, is collaborating with LLNL regarding the design and construction of its facility. LLNL is now experiencing an unexpected resurgence in its pursuit of more conventional nuclear fusion research programs as result. As the generation of scientists that performed LLNL’s mirror fusion and beam research largely departed the site following the closure of MFTF in 1986, LLNL has now recapitalized its conventional nuclear fusion capabilities by attracting significant researchers in the field, including significant figures in the Russian magnetic confinement fusion arena such as former Kurchatov Institute scientist Sergei Krasheninnikov.

Unlike the other two laboratories, Sandia devoted its attention to sensitive inertial confinement fusion concepts in the late 1950s and early 1960s (Van Arsdall, 2007). Those Sandia researchers who participated in this work were subject to the same security measures experienced by LLNL researchers in the 1970s and 1980s. Thus, they had relatively few opportunities to interact with scientists outside of the United States and the United Kingdom. Sandia missed the early expansion of collaborative ties between U.S. and Russian researchers that marked the early period following the declassification of magnetic confinement fusion research in the late 1950s.

Sandia did not begin playing a significant role in the development of conventional, magnetic confinement fusion until much later. Similar to Oak Ridge, Sandia focused its early attention on magnetic confinement fusion through the use of heavy ion beams, an approach largely outside the mainstream

tokamak reactor, mirror machine and laser induced fusion approaches (Yonas, 1978). It took forty years for ties between Sandia and Russian researchers in this field to emerge.

Following the upgrade of one of its particle beam fusion accelerator facilities (PBFA-II) to the Z-Machine in 1996, Sandia moved from the periphery of controlled nuclear fusion research to its center. The Z-Machine revived the promise of the Z-pinch nuclear fusion concept earlier explored at Harwell and Los Alamos by constructing a facility in which large volumes of electrical current could be discharged into wire meshes oriented around target materials in extremely short periods of time (nanoseconds) to spherically implode them using magnetic compression. Its follow-on facility, ZR, routinely records plasma temperatures greater than that produced by any known tokamak and is a brilliant X-ray source (Van Arsdall, 2007).

In an unusual step for Sandia, international collaboration was a key part of the upgrade of the Z-Machine to the ZR. ZR was designed in collaboration with the Kurchatov Institute of Atomic Energy in Moscow. Following collaboration with Russian engineers at the High Current Electronics institute (HCEI) in Tomsk, Sandia's ZR facility increased the repetition rate of "shots" that it can conduct with the development of linear transformer drivers (LTDs). This increase in repetition rate opens the possibility that a multi-chamber inertial fusion engine concept (Z-IFE) based on the ZR design could be a viable candidate for a future nuclear fusion power station.

Like LLNL, Sandia is now experiencing the cross-domain benefits of developing and maintaining a significant "big science" asset. Due to its experience in containing the plasmas generated by "shots" on Z, Sandia is now

designing the plasma first walls for the ITER tokamak in France. Sandia is involved in novel damage mitigation activities within ITER as well. In 2014, the DOE Office of Fusion Energy funded Sandia and the University of California at Davis to develop techniques to form compact toroidal plasmas within ITER when plasma dislocations appear to be imminent.

In review, two general observations can be taken away from patterns of interaction that can be observed across these two sets of U.S. research institutions. Research institution types that follow an open science model in the U.S., such as universities, seem to be quicker to involve themselves in international collaborations than their more “national benefit” oriented national laboratory counterparts. Part of this delay in interaction on the part of U.S. national laboratories can be attributed for a tendency for some novel technologies to be “born classified,” such as in the case of controlled nuclear fusion, which delays the period that national laboratory researchers can engage with external colleagues.

Risks from Increased International Collaboration

The emergent accessibility, positive scientific reputation and diplomatic push to collaborate with these Russian institutes all influenced the timing and pace of new collaboration activity between Russian and US national laboratories. However, it is not clear many proponents of scientific collaboration between these two sets of national assets actively considered the potential risks associated with these joint research arrangements prior to moving into them. This lack of prior consideration, however, does not suggest that risks do not exist.

First, collaborative ties expose researchers to their collaborating partner's operating context. This exposure reveals significant differences in administrative requirements, compensation, research equipment quality as well as environmental and safety regulations. If the disparity between the two collaboration partners is sufficient, these differences can drive researchers to seek employment at other locations. Researcher out-migration erodes a research institution's ability to carry out current and future research activities in research areas impacted by this out-migration and limits the institution's ability to absorb new knowledge by removing the tacit knowledge these individuals possess from the institution's knowledge base and research processes.

In addition, it is clear that collaborative ties can be used to transfer both critical individual and institutional tacit knowledge across nation-state boundaries. Through these ties, tacit knowledge may be transmitted from senior researchers in one nation to junior researchers in another nation as researchers from different institutes work side-by-side. If it occurs, this knowledge transmission should diminish the barriers of entry for those same junior researchers to understanding existing explicit knowledge, such as that contained in technical journals, and push the resulting researcher groups that now possess similar tacit knowledge bases along similar research trajectories, leading to similar future research findings and technology developments. For the researcher that reveals new scientific knowledge, then, this tacit knowledge transfer may reduce individual prestige as unique insights from scientist's tacit knowledge are broadly shared and disseminated without metering. For the applied researcher, this transmission of tacit knowledge can be construed as an intellectual property loss with material losses.

In the context of Foray's basic researcher model (Foray 2004), a reduction of researcher prestige should alter a given researcher's career trajectory. In the academic context, researcher prestige can be directly linked to the probability that a researcher gains tenure or is able to acquire a teaching position at one of the leading institutions in their field. A loss of prestige should have a negative influence on the probability that either of these events occur. Indirectly, loss of prestige may also be linked to whether a researcher "abandons the bench top" and pursues an applied position in industry or pursues a managerial role. Applied researchers at universities may lose an opportunity to begin a business on the basis of a trade secret.

In a national laboratory setting, it is unclear how applicable this model is. Foray's model is based on the idea that academic researchers get paid to teach rather than receive pay contingent upon successful research results. In the national laboratory model, researchers are paid to perform research and deliver technical results to government customers. More successful researchers and technology developers attain greater pay and autonomy than less successful researchers and technology developers. Thus, a researcher that loses prestige in the basic research domain through loss of unique tacit knowledge can perform well in developing national benefits oriented technology and compensate for the loss of status.

If enough basic researchers make this transition, however, the national laboratory will lose its ability to absorb new knowledge as these researchers cease to perform open research in their chosen fields. At the institutional level, the individual risks associated with collaboration are magnified by aggregation. The dispersion of prestige from scientific discoveries can lead to greater

uncertainty in the hiring of scientists who can contribute to institutional mission areas. Even the advantage existing competencies yield these institutions may erode as a result of discouraged researchers leaving their fields for managerial and advisory positions where they make smaller contributions to technical knowledge. Thus, strategic efforts to develop absorptive capacity and long run technical competencies will suffer. In short, these issues can impact whether a national benefit oriented research institution thrives or declines.

There are legal risks that complement these observed risks to medium and long run institutional research productivity capacity. If inadequately supervised, international collaborations can devolve into researchers at national laboratories committing program fraud by conducting “off-the-books” work for research institutes in a partner country, often in return for financial inducement. If these tacit knowledge transfers involve knowledge protected by export control regulations, any individuals making these transfers can be found to be in violation of export control laws and can face significant fines and substantial prison sentences. Patterns of program fraud and export control violations can lead to stiff fines, export debarment, and change of operating contractors.

The issue of legal compliance in this space is not a trivial matter. It is easy for such international collaborations to run afoul of complex regulatory environments focused on the protection of intellectual property deemed critical to the development and maintenance of strategic national technology competencies for domestic economic growth and strategic international political goals. In the United States, such international collaborations have to comply with a myriad of multiple agencies implementing different export control schemes based on legislation passed in the mid-20th century. Owing to its relatively

modern origin, the Russian export control regime lacks the inherent complexity of the US regime, but contains enough ambiguity such that compliance is also not trivial (Beck, Cupitt, Gahlaut and Jones, 2003).

University researchers in the United States have an advantage in developing international technical relationships when compared to their national laboratory colleagues in the same technical fields because compliance with export control regulations is simpler. This advantage emerges from the presumption that university research is “fundamental research” with respect to export control regulations. According to current US Export Administration Regulations (15 CFR §734.8), “fundamental research” is defined as “basic and applied research in science and engineering, where the resulting information is ordinarily published and shared broadly within the scientific community that is not typically restricted for proprietary reasons or specific national security reasons.”

A national laboratory researcher in the United States faces a more complex environment. Such researchers must attempt to comply with export control regulations implemented by a myriad of federal agencies. Researchers must be in compliance with the Arms Export Control Act (AECA) and the International Traffic in Arms Regulations (ITAR) administered by the Department of State, the Export Administration Regulations (EAR) overseen by the Department of Commerce as well as the export provisions of the Atomic Energy Act (AEA) managed by the Department of Energy and the Nuclear Regulatory Commission (NRC). National laboratory researchers may still describe their research as “fundamental,” but whether it actually is viewed legally as fundamental or not for export control purposes depends on how the

research is conducted, with whom, and if it complies with the laboratory's (or Department's) own standards for public release.

Compliance with export control regulations would not be a significant issue if law enforcement agencies only pursued export control cases on only a periodic basis. In such a situation, both universities and national laboratories could pursue a strategy of self-insuring against an occasional unfavorable judgment. In the United States, however, law enforcement agencies are sharply increasing their efforts to identify instances of economic espionage, trade secret theft and other intellectual property crimes. The National Intellectual Property Enforcement Coordination Center reported the Federal Bureau of Investigation (FBI) increased the number of intellectual property investigations cases in 2011 by 56 percent (NIPECC, 2012). During this same period, Homeland Security Investigations (HSI) arrested 530 people for intellectual property theft issues resulting in 304 convictions for export control related criminal violations (Woods, 2012).

Individual penalties associated with observed violations have increased over the past five years as well. In 2012, the US Department of Justice observed that successfully prosecuted intellectual property theft cases are generating sentences with greater length. Sentences of three or more years are now more common (NIPECC, 2012).

Most of the intellectual property cases pursued by the US Department of Justice involve thefts of intellectual property from firms in the United States. However, national laboratory and university researchers collaborating with foreign partners have not escaped the attention of US law enforcement authorities. In 2012, former SNL physicist Jianyu Huang was arrested following

grand jury indictment on charges of lying to a federal official, federal program fraud, and theft of government property. Huang was accused of using unique research equipment at the Center for Integrated Nanotechnology (CINT) to covertly assist research institutes in China performing research in support of Chinese national security aims. In 2008, University of Tennessee engineering professor John Reece Roth was convicted of using graduate students from China and Iran on U.S. Air Force research that was off-limits to foreigners, and taking a laptop with restricted files to China. Roth began serving a four-year sentence in January 2012.

Finding a Balance

Clearly, there are sound economic reasons for engaging in extensive international research collaboration for scientific research or this rise in interaction would not be observed. It is apparent that different nation-states have comparative advantages in performing certain research functions. The heavy U.S. emphasis on “big science” facilities like the National Ignition Facility at Lawrence Livermore has yielded a U.S. advantage in conducting “frontier-expanding research” (Newman, 2000). Similarly, the premium placed on analytic skills within the Russian scientific community is thought by many U.S. researchers to generate a scientific cadre with a unique capability to maximize the observable implications from a given small scale physics experiment – making large scale facilities less of a necessity.

An international network for basic research that applies a cadre of researchers trained in a variety of educational traditions against a scientific research question using a diversity of unique research equipment is an

organizational form that is ideal for the maximization of scientific output, but only in some narrow research environments. The historical observation that groundbreaking basic research and application development has typically taken place in geographic clusters (e.g. northern England for textiles, southern France for fragrances, Silicon Valley for information technology) indicates that these conditions must be fairly restrictive (Lazonick, 2005). Notionally, this organizational form only appears to be optimal when the production in the topical area is a pure global public good, the cost of a given niche basic research facility is high (and feasibly, bearable only by collections of governments of nation-states) and a high degree of complementarity in output exists across the span of national cadres of researchers who can communicate in a low cost manner.

This set of conditions only echoes reality in a few scientific research areas such as high-energy particle physics. Many nation-states have well-established research programs in the field area, an individual's facility with English is a prerequisite for advanced study in the field and large, expensive facilities like the Large Hadron Collider (LHC)¹ are necessary to make scientific advances. However, even passing over the contentious question of whether any scientific output can be considered a global public good, these conditions are not met in a universal way across the broad span of scientific topics.

Research in emergent technology areas, such as advanced semiconductors and meta-materials, is not subject to multi-national investment level barriers to

¹ CERN's LHC was built at the cost of \$9 billion over the course of 10 years (CERN, 2009). Just operating the previous leading supercollider – the Relativistic Heavy Ion Collider (RHIC) – at Brookhaven National Laboratory costs over \$100 million per year. (DOE, 2009)

entry for research, development or production. Graphene – the atomic monolayer form of carbon used to make the smallest transistors on record – is still prepared by pulling adhesive tape off of high quality graphite, a process called “mechanical exfoliation” (Geim and Kim, 2008). A transmission electron microscope (TEM) -- which can cost as little as \$20,000 FY12 USD for used equipment -- is required to examine graphene. As production of graphene does not scale up well at the moment, most production consists of small batches harvested by hand.

Thus, a question emerges. Is the rise in international collaboration by U.S. national laboratories and universities across the broad span of technical fields truly warranted? For nuclear fusion research, for which many of the conditions asserted above apply, international collaboration appears to be valuable and useful. However, given the information above, it seems unlikely that international collaboration in nanotechnology is going to result in any dramatic improvements in scientific output over a nation-state forming its own small number of geographic clusters of nanotechnology focused research parks and eschewing international engagement – at least initially.

Unfortunately, the question of how and when national research institutions should collaborate cannot be answered on the basis of scientific output alone. As indicated earlier in this chapter, international research collaborations are viewed as ties that can serve an unofficial diplomatic function in international affairs by governmental agencies. Forming these ties requires that U.S. national laboratories are brought out from buffered isolation and into interaction with laboratories like the D.V. Efremov Institute, the P.N. Lebedev Institute, the Kurchatov Institute, the All-Russia Institute for Experimental

Physics and the All-Russian Institute for Theoretical Physics – all research facilities supporting a foreign power’s national security apparatus. From a knowledge perspective, any unproductive collaboration between these two sets of partners could be a transfer mechanism for tacit knowledge from one country to another – potentially eroding the core competencies of the laboratory (or country) transferring the knowledge.

The central concern of this document is to assess whether US national laboratory research productivity (as well as related core competencies) have been negatively impacted by collaborations with Russian research institutes and identify important sources of heterogeneity that influence the productivity impact of these collaborations in order to improve US national laboratories ability to benefit from these external interactions. This study has already discussed two important sources of productivity impact heterogeneity - institution and research field. The following chapters will describe and explore how differences in tacit knowledge, network neighborhood and connectivity characteristics are other important factors that influence the impact of individual co-authorship relationships, and at an aggregate level, the impact of a given institutional collaboration.

A Note on Methods

The research in this volume uses a novel set of analytic methods to arrive at its conclusions. Exploratory social network analysis is used to explicate the position of researchers at the interface of interactions between Russian and US national laboratories and examine their connectivity mixing patterns. The methodology used to identify research fields for collaborating scientists and

develop publication output metrics for topical research programs relies upon visualized latent semantic analysis for content analysis. Briefly described in the fourth chapter, this technique for quantitative analysis of unstructured text is used in computational linguistics and applications such as IBM's "Watson" artificial intelligence application, but is rarely observed in economics. More conventional empirical techniques are also employed (e.g. negative binomial regression modeling of count data), and are consistent with the latest techniques applied in innovation performance assessment (Baba et al. 2009).

This final choice of analysis techniques in this project differs from that in the initial research plan. The original plan was to use more conventional time series techniques to detect regime changes in publication frequency and quality as a result of collaborations. Unfortunately, observed micronumerosity and a lack of observed author independence limited the utility of approaching individual collaboration assessment using time series means. Instead, this study followed "an auxiliary network analysis" model (Wassermann and Faust, 1994) choosing the links between individual authors as the fundamental basis for analysis rather than just considering author attributes in a vacuum as many other such studies.

As hinted at earlier in this section, the knowledge-focused approach of this study is inspired by the work of Dominique Foray in knowledge economics (Foray, 2004), but its empirical instantiation in this examination is novel. Operationalizing some of Foray's concepts (e.g. tacit knowledge) required adoption of the aforementioned unstructured text analysis techniques to detect topics in publication datasets. To this author's knowledge, the kind of classification exercise engaged in within this study to empower its completion is

also novel within research and development economics, if not within small sample scientometric studies (Small, 2011). Indeed, the promise of this application of latent semantic analysis may help “democratize” the economic study of knowledge as well as emerging science and technology by allowing policy analysts with narrow or shallow scientific knowledge to identify emergent topics and critical enabling technologies without having to resort to solicitation of subject matter experts who could benefit from the outcome of a given cost-benefit analysis or research portfolio decision.

Each of the software tools used in this set of studies is publicly available. The exploratory social network analysis makes use of Gephi (available in alpha from <http://gephi.org>) – a noted benchmark in knowledge visualization. This project relies upon Sandia developed LDRDView for visualized latent semantic analysis (available for free on request from <http://www.osti.gov>). All of the conventional empirical analysis was conducted using Stata 13 (available from <http://www.stata.com>). Any customized programming developed in the course of this inquiry is included for completeness in one of the attached appendices.

Overview of this Document

This document consists of this introductory section, three chapters describing research in the area and a concluding section summarizing the research results and what those results may mean for the participation of US national laboratories in international collaborations with other nation’s national laboratories. The intent of this introductory section is to set the stage for the volume’s treatment of the problem of assessing the value of national laboratory-national laboratory linkages on the international stage. The research sections that

follow begin with a description of the theoretical, knowledge-oriented model used to generate the primary research questions at the heart of this research attempt. These questions are then examined through the lens of the structure of international national laboratory-national laboratory linkages (Chapter III).

The fourth chapter is an examination of the hypothesis developed in Chapter II and III. The main content of the chapter is an empirical examination of collaborations that makes use of simple joint publication counts at the individual and the institutional level to explore if there are any indications that localized research productivity changed as a result of international collaborative efforts. Due to its use of simple publication counts, this research follows in a tradition of work related to publication production similar to that conducted on patent production prior to 1995 (see Pakes, 1985; Cockburn and Griliches, 1988 for examples).

It should be noted that the exploratory focus on structured transfer of knowledge via diffusion over known social linkages represents a break from the typical knowledge diffusion literature (see Caballero and Jaffe, 1992, for an example) that fails to adequately deal with the idea of tacit knowledge. The approach used is similar to techniques employed by Sorenson and Singh in studying nanotechnology (Sorenson and Singh, 2007) that adopt more refined social network analysis attributes to study the transfer of economically relevant knowledge.

The final chapter uses these research results to paint a path forward for continuing research collaborations between US national laboratories and their fellow international laboratories that better preserve benefit symmetry. A predominant theme in this treatment is how to structure (or restructure)

collaborations to take advantage of complementary knowledge when appropriate for furthering technical goals, not just seeking collaboration for the sake of maintaining an external political relationship. This theme includes injunctions to more carefully structure collaborations to avoid unnecessary knowledge transfer as well as preserve external-internal knowledge exchange by developing mechanisms for the intergenerational transfer of tacit knowledge.

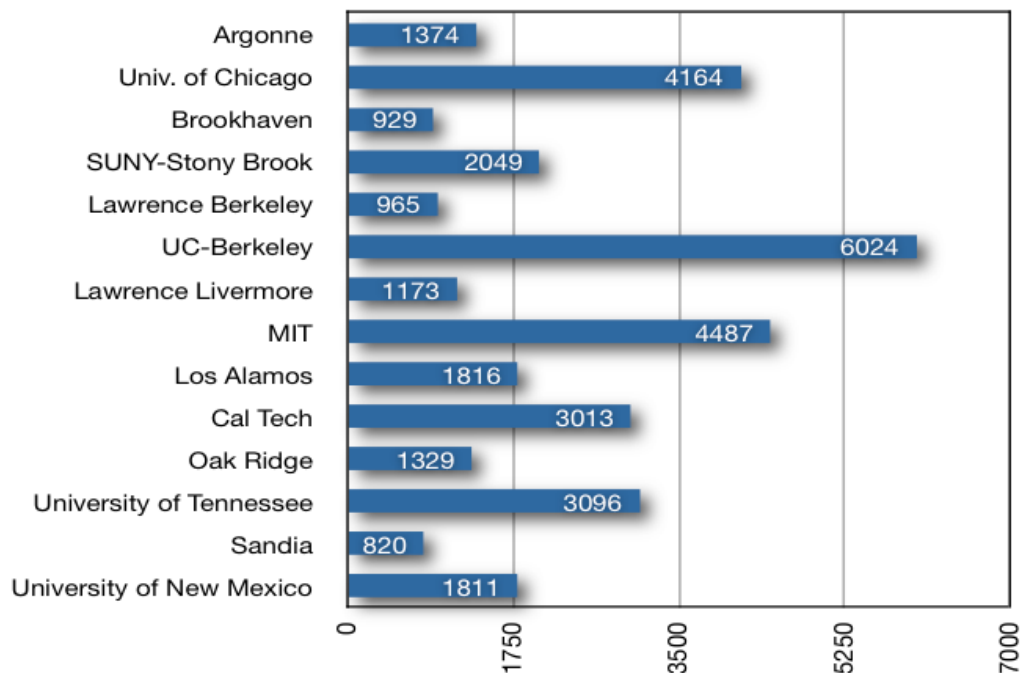
II. Modeling Research Production from a Knowledge Perspective: A Second Best Approach

National Laboratories: Practically Important, Theoretically Inconvenient

Unfortunately, the consideration of many of the questions raised in the introduction almost takes place in a theoretical vacuum. Despite an impressive list of path-breaking scientific and technical accomplishments, national laboratories are often left out of academic discussions concerning key institution types in the U.S. research and development system. Discussions about U.S. national laboratories as part of a national innovation system from non-academic actors, such as the General Accounting Office (GAO) or the Henry L. Stimson Center, typically limit their consideration of these facilities to their instrumental value to government agencies, such as the Department of Defense (DoD) or the Department of Homeland Security (DHS). As a result, there is clearly a gap in knowledge about how the national laboratories generally contribute to the advancement of research and development in the United States.

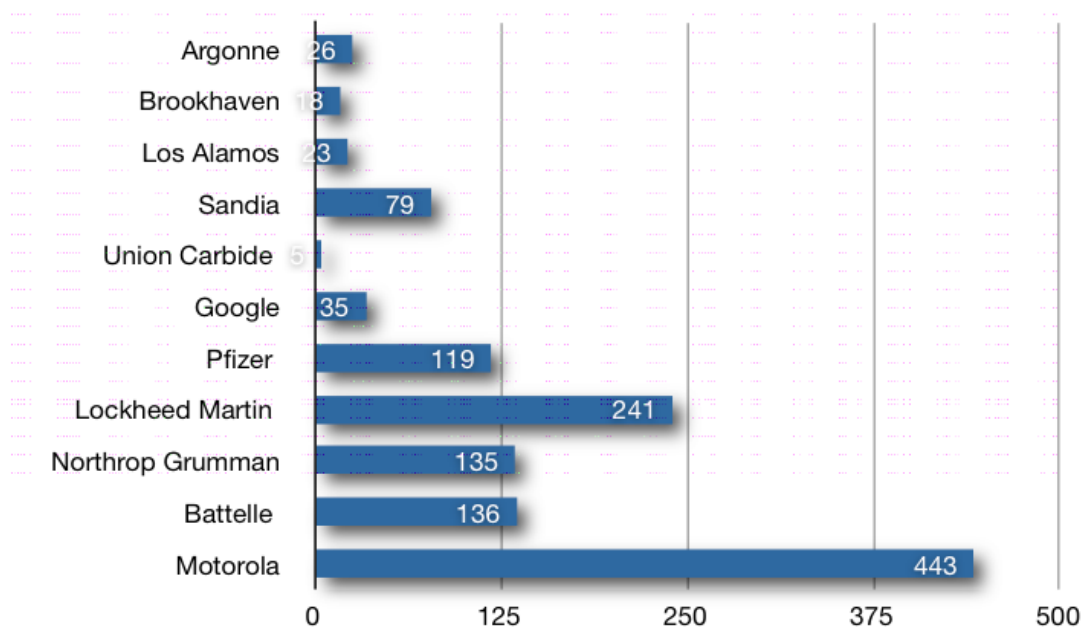
Given the absence of a deep body of literature on the topic, it is necessary to set the stage for understanding why national laboratories are important to the national innovation system. To demonstrate this importance, the typical output of these research and development institutions in generating basic research and inventions should be compared to other, better-documented forms of research and development institutions – universities and large research and development driven firms. As can be observed from Figure 2, the more basic science research focused national laboratories, such as Los Alamos and Argonne, produce as many publications as a regional research university (e.g. University of New Mexico). Laboratories with more of an engineering orientation produce about half of the publications of that same university type. The key difference is that within the national laboratories almost all of the publications are in scientific and technical areas.

Figure 2. Publication Output by Selected US National Laboratory and Major Partner University (Journal Publications in 2007)



A similar observation can be made with respect to the generation of inventions. As shown in Figure 3 on the next page, engineering focused national laboratories, such as Sandia, receive more issued patents than either relatively young commercial firms known for their innovation (exemplified by Google) or firms in mature industries (here represented by Union Carbide in the chemical industry). It should be noted that national laboratories with less of a focus on developing inventions still were issued more patents in 2007 than the exemplar organization from the mature industry.

Figure 3. Patent Output by Selected US National Laboratories and Industrial Research Organizations (Patent Applications Issued, 2007)

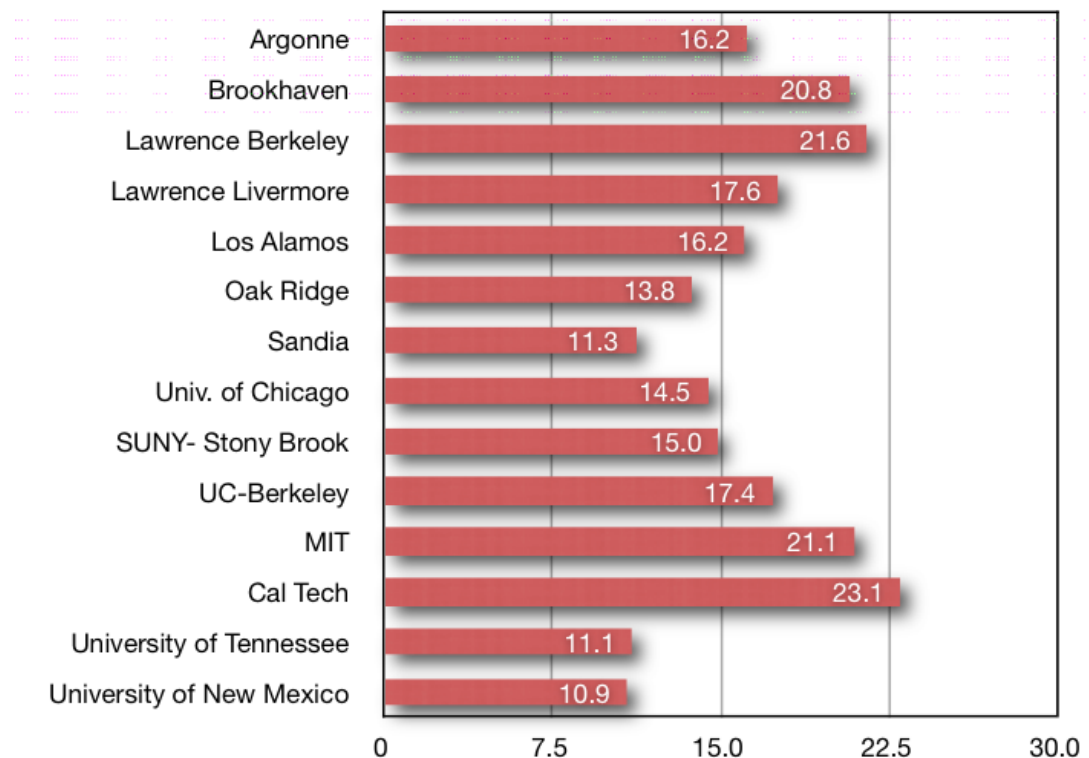


There is one other notable trend that can be gleaned from this data. The firms that operate national laboratories - Lockheed Martin and Battelle in this sample - generate more than twice the number of annual patents than national research facilities themselves. Unlike their operating corporations, however, the

patents generated by the national laboratory facilities tend to have stronger ties to cutting-edge basic research (Jaffe, Fogarty and Banks, 1998).

If national laboratories are simultaneously generating enough basic research to compare to a regional research university and also putting together enough patents to rival innovative firms in industry, the quality of the generated publications and patents could exhibit poorer research quality - a manifestation of the typical quantity/quality tradeoff. However, this tradeoff does not appear to have materialized. As can be seen in Figure 4, the average publication citation rate for journal articles - a common indicator for research quality - produced within national laboratories is roughly equivalent to that observed from their partner universities.

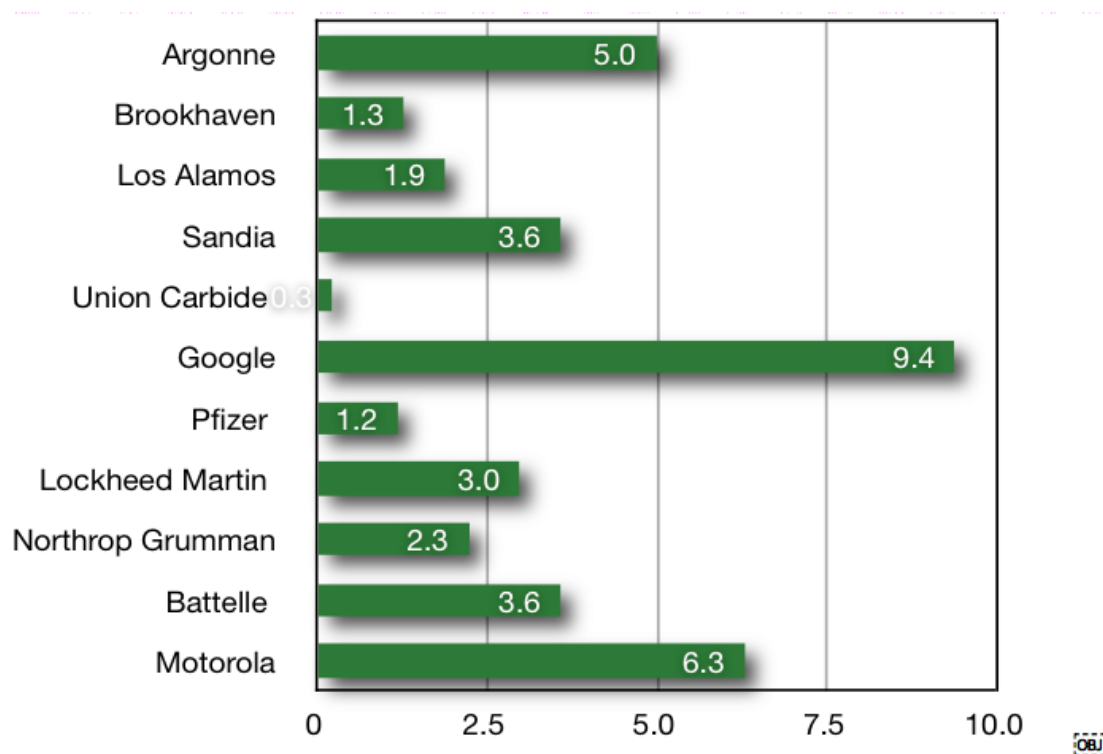
Figure 4. Average Publication Citations, US National Laboratories and Major Partner Universities (Citations of 2004 Publications)



National laboratory articles receive more citations on average than publications generated by large state universities in this sample. Publications from Brookhaven or the Lawrence Berkeley National Laboratory receive about the same number of average citations as articles from MIT and California Institute of Technology. There appears to be little difference in overall quality between the two institution types.

This trend is similar to what can be observed by comparing a selected sample of national laboratories and some of their major industrial partners in a cross-section examining average patent citations across each institution (Figure 5). On average, national laboratory patents are cited less frequently (2.91 citations/patent) than patents developed by their major industrial partners (3.71 citations/patent) in this exemplar sample. However, the number of patent

Figure 5. Average Patent Citations, US National Laboratories and Major Industrial Partners (Citations of Patents Granted in 2004)



citations per patent is roughly equivalent for the national laboratories and firms who perform research and development work primarily for government agencies (e.g. Lockheed Martin, Northrop Grumman, Battelle). Notably, two national laboratories (Argonne and Sandia) outperformed each of the defense contractors.²

The pattern shown in this previous figure is suggestive in its support of conventional wisdom. Entities in each of these research and development organization types produce high-quality innovations that are used as foundational knowledge for future innovation. Commercial firms appear best suited to generate innovation in support of direct technology development in rapidly evolving industries as the performance of Google and Motorola in the above chart suggests. Likewise, national laboratories that have a more fundamental research bias in their research portfolio (e.g. Los Alamos) demonstrate difficulty in making the transition to a greater focus on applied research. These national laboratories appear to produce the least-cited innovations. National laboratories with strong engineering traditions do not seem to have these same transitional issues.

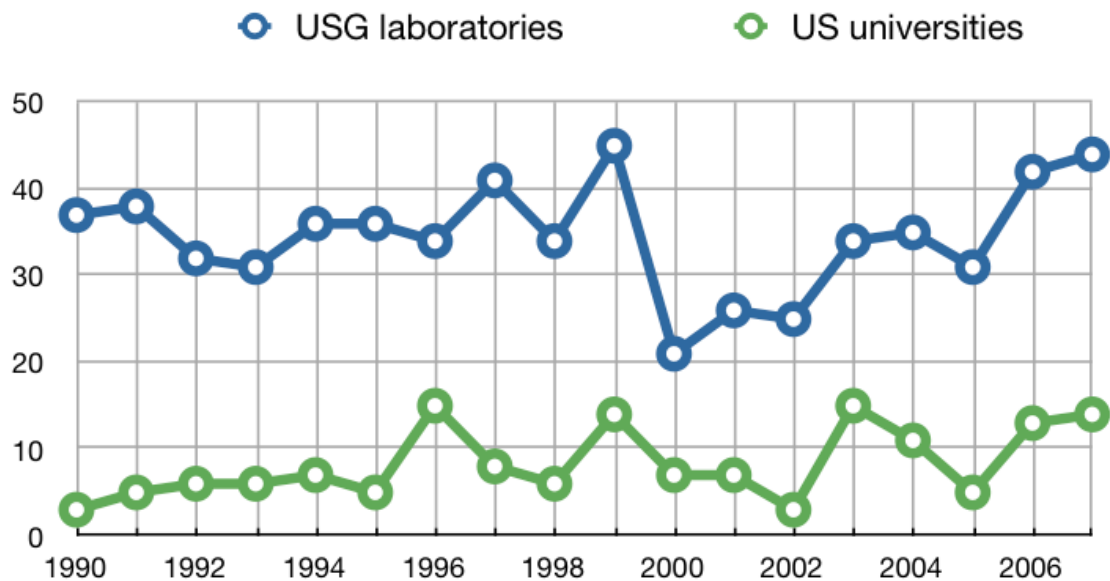
Despite this nuance, the lack of notice given to national laboratories in the academic literature does appear to be inconsistent with both the laboratories' level of research output as well as the high quality of publications and patents that these facilities produce. This inconsistency becomes even more pronounced

² It should be noted that national laboratories do not actively seek firms that are infringing upon their intellectual property in the same manner that corporations do. Instead, national laboratories often rely upon external law enforcement and security agencies to search for infringing activities. Patent infringing firms takes advantage of the absence of infringement search activity by these national laboratories by not citing national laboratory patents.

when the level of success these institutions exhibit in developing significant innovative technologies is considered. As an example, 44 R&D 100 awards³ - often referred to as the “Oscars” of innovation - went to US governmental laboratories in the same year depicted in the earlier graphs concerning patent and publication productivity (2007). In comparison, US universities only garnered 14 such awards this same year.

As can be observed from Figure 6, US national laboratories consistently outpace their US university counterparts in acquiring innovation awards by a minimum of a three-to-one margin. As is discernible from the figure, the gap between the two institution types appears to have closed somewhat since 2000. Overall, however, this relationship has remained fairly steady for the bulk of the near fifty years the R&D 100 awards have been offered.

Figure 6. R&D 100 Awards, US National Laboratories and Universities (1990-2007)



³ The R&D 100 awards are determined annually by the editors of R&D Magazine to honor the 100 most technologically significant products developed in a given year.

Careful examination of the corpus of R&D 100 winners shows this difference in recognition exhibits both an extensive and an intensive margin. Many of the larger US Department of Energy laboratories (e.g. Los Alamos, Lawrence Livermore, Sandia) appear multiple times on the list annually. Very few academic institutions appear on this list at all. MIT is the only US university that routinely appears on the list of award winners.

Out of the Mainstream?

Given the significance of their productivity level, research quality and innovation significance, the contemporary innovation literature may lack depth on the topic of national laboratories because these research and development environments are outside the “invisible colleges” that characterize most mainstream scientific activity (Crane, 1972). This isolation could arise easily - national laboratory researchers work in environments where there are more constraints on professional interactions than what university scientists experience. As such, these constraints should limit the number of laboratory scientists and engineers who can rise through the ranks of professional societies to serve as officers. A lack of participation at these highest levels could reflect the possibility that national laboratory researchers are not perceived as a key part of the scientific cadre crafting each discipline’s research programs.

Examination of professional society leadership statistics indicates there may be some truth in this assertion. As can be observed in Table 1, national laboratory figures tend to be high profile members of certain professional societies while largely ignored in most societies. This lack of representation in leadership positions in professional societies calls into question the true

Table 1. Professional Society Leadership, by Society and Institutional Type, 2008

| Professional Association | Universities | National Laboratories | Industry |
|--|--------------|-----------------------|----------|
| American Association for the Advancement of Science (AAAS) | 12 | 1 | 0 |
| National Academy of Sciences (NAS) | 16 | 1 | 0 |
| American Chemical Society (ACS) | 5 | 0 | 9 |
| American Physical Society (APS) | 3 | 1 | 0 |
| American Nuclear Society (ANS) | 1 | 2 | 1 |
| American Society of Mechanical Engineers (ASME) | 3 | 2 | 6 |
| Institute of Electrical and Electronics Engineers (IEEE) | 7 | 5 | 16 |

integration of national laboratory researchers into the “global component” that connects researchers in a given technical field (Newman 2001).

Some biasing forces are clearly at work here. The national laboratories portfolio of research includes work that requires facilities too expensive for either a university or typical commercial concern. As a result, researchers from national laboratories tend to dominate technical arenas where cutting edge research requires use of such facilities - such as research into nuclear fusion. This dominance is expressed in the professional society corridors as well - the head of IEEE’s Nuclear & Plasma Science Society is typically an individual from one of the national laboratories.

However, this large capital requirement in some fields is a clear barrier to widespread replication of results. As such, this inability on the part of most university researchers to replicate research results constrains the level of awareness of national laboratory research. This factor could reduce the visibility of national laboratory researchers in more general scientific professional societies.

An Ill-Fit for the Typology

The social aspect observed above likely contributes to a lack of visibility of national laboratory researchers within the broader American technical community. However, the primary reason the academic innovation literature is largely silent on the issue of how to consider the work of national laboratories is because neither of the conventional models used to explain research institution behavior can be directly applied to the study of national laboratories. Much of this innovation literature focuses on two ideal types -- open science or the appropriable technology innovation model (Foray 2004) -- and the institutions that embody those principles the best -- universities and industrial laboratories.

In the United States, national laboratories occupy the role of basic research hub, technology transfer intermediary and technology developer simultaneously. Some of these facilities even engage in full-scale manufacturing activities. Neither of these conventional models is appropriate for dealing with an entity whose innovation activity occurs in both areas simultaneously without significant adjustment. At best, it could be argued that these two models could be blended together to explain national laboratories' work.

Another complicating factor is that each of these governmental laboratories maintains a national security research portfolio as well. Clearly, the

output characteristics of this part of the national laboratories' research and development portfolio will be unobservable to academic researchers outside the national laboratory system. However, the presence of this national benefit oriented portfolio impacts the ability of the innovation researcher to fully observe national laboratory innovation activities.

This hidden portfolio creates a hidden information problem. The national laboratory focused innovation researcher will not be able to make use of many of the traditional research and development trending measures in conducting research activities. Information such as research and development funding per site per annum has been viewed in the past by the government and the national laboratories as sensitive unclassified information protected from public disclosure.⁴

The absence of this data creates a situation where it is difficult for a researcher to get a sense for the changing nature of the incentive structure for scientists and engineers to produce journal publications within these organizations. Salaries of individual researchers are not disclosed at the national laboratories because the researchers at these institutions are actually employees of private holding corporations, instead of the government itself. This situation is contrary to the public university case, in which even normally private information such as researcher salaries are often publicly disclosed.

⁴ It should be noted that there is increased interest in generating greater transparency associated with government-funded research in the United States. In 2013, the Office of Science and Technology Policy (OSTP) within the Executive Office of the President issued an instruction that "direct results of federally funded research be made available" to public access. This movement to transparency is unlikely to resolve the problem discussed above, however, which is based on in established federal codes rather than in administrative instructions.

To complicate matters further, the importance of the national security portfolio item varies across the national laboratories. The impact of this variance across the cross-section of the DOE national laboratories is such that imputing input and output from proxies is difficult without some internal intuition about how these entities actually function. If the DoD national laboratories are added to this mixture, the picture becomes even muddier owing to the extreme variance in researchers from these facilities ability to publish in open journals across different presidential administrations. As such, it is no wonder that many academic researchers (see Foray 2004 for an example) are unsure how to treat this institutional type and quickly relegate national laboratories to the periphery of theoretical development as a non-generalizable special case.

The Problem Space

At this point, it is useful to gain a perspective on the problem as stated so far. Assessing whether or not the US national laboratories' pursuit of collaborations with other nations' national laboratories exposes these same facilities to competency erosion requires a significant amount of novel work owing to the lack of depth on the subject in the innovation literature. However, tackling this question means eschewing many of the common ways of examining research and development institutions, because of the clear hidden information problem concerning both the inputs that go into the laboratories and what the laboratories produce as part of their national security portfolio. In such circumstances, the national laboratory researcher is left with a quandary concerning how to study international national laboratory collaboration impact

in a replicable manner without having to resort to the cultivation of anonymous internal sources or gain access to protected unclassified information.

A potential answer to this dilemma is to change the analysis focus. It is true that understanding national laboratory innovation in academic research suffers from a hidden information problem. This issue is a serious obstacle to an information oriented perspective and understanding the characteristics of this missing piece of the puzzle are important. However, understanding information is not integral to answering the research question driving this work -- understanding knowledge is.

Fortunately for this work, the knowledge base that produces that hidden information is not hidden itself. To get a perspective on how this last statement can be true, it is important to understand how government laboratories protect information in association with executive order restrictions and legal guidelines and then reflect on how national laboratories are organized. The upshot is that these laboratories still produce observable results linked to their knowledge base, even if a significant portion of the information they produce remains protected.

Much of national laboratories' national security related work in the United States is classified as "Official Use Only." This class of sensitive, but technically unclassified information is protected because its disclosure has the "potential to harm commercial, governmental or private interests" and falls into one of an evolving set of exemption categories: commercial/proprietary, privileged information, personal privacy, law enforcement information, etc. Public release of this information requires formal adjudication under the Freedom of Information Act (FOIA).

In the United States, the national laboratories associated with the Department of Energy also process formally classified information. Such information falls into three categories of information: national security information (NSI), restricted data (RD) and formerly restricted data (FRD). Each category has three levels of sensitivity: confidential, secret, and top secret.

Each of these categories of classified information finds its origin in US law. The classification of national security information comes from a series of presidential executive orders (see EO 13526 as the latest addition in this series) beginning with a foundational executive order on the topic issued during the World War II era Roosevelt administration. In the case of restricted data and formerly restricted data, the Atomic Energy Act of 1956 dictates these information classification types.

However, unlike many researchers at universities and industrial laboratories, researchers at national laboratories live in an environment where there is encouragement and ample opportunity to work on both open science projects and research outside the public domain at the same time. The same researcher is likely to be working on a basic research question and an applied technology project at the same time. A flexible organizational structure -- the matrixed organization -- is often deemed desirable to coordinate this multiple tasking of single assets and facilitate the flow of innovative knowledge from one area of the laboratory to another.

The broad adoption of this organizational form in US national laboratories means that a national laboratory researcher with a given research expertise will produce multiple kinds of research products while maintaining a singular technical position. In many cases, the average researcher will not be confined to a

functional silo where they do nothing but perform basic research or technology development or national security work. The constant is the same people - with the same tacit knowledge -- generate all of the products.

Thus, it is conceivable that an innovation interested researcher could make use of these openly available indicators as proxies for other better known research and development indicators. For instance, it is well known that simple patent counts are highly correlated with research and development effort oriented inputs like spending (e.g. Griliches, 1984). Thus, simple publication counts at the institutional level could shed some light on changes in the level of effort that a given institution is applying to the development of research and development activities over time. Changes in publication count generation at a lower level of analysis (e.g. the individual or research organization) as a result of an international collaboration could actually be an indirect measure of knowledge flow across coauthor pairs (Goyal, 2007).

Modeling Research Productivity

To examine the nature of the relationship between collaboration and research productivity, it is necessary to more formally explore the factors that drive research publication generation. Such a consideration in this context begins with the observation that national laboratories have to choose a level of institutional effort (E) to apply to the pursuit of what Dasgupta and David (1994) refer to as open science research. Within national laboratories, funding for this institutional effort comes from internal taxation on technology development contracts with entities other than the laboratories' sponsoring agency. It should be noted this funding mechanism deviates from the model proposed by the

Stanford “new” economics of science school in that researchers are actually paid to generate research, not teach students or manage programs (Mirowski and Sent, 2002).

Following Niskanen (1975), this treatment assumes that research and development managers at national laboratories maximize budgets. Note that this treatment deviates from the typical Niskanen model in that the bureaucrat is maximizing discretionary budgets for the future in the present

$$(1) \quad B_1(X(X_{t-1}, P_{t-1})) + (1 / (1 + r))(B_2(X_{t+1}(X_t, P_t)))$$

where B , X , P and r represent the revenue generated by technology development projects for other agencies, the number of these technology development projects, the number of publication signals and the institutional discount factor, respectively. In this form, X_t is fixed as both the number of past development projects and publications are known. Since X_t is fixed, the research and development manager’s choice variable is P - the number of publications generated in the current time period.

Decision-makers in this environment are clearly constrained by an intertemporal resource allocation issue. The key question they must answer is “How much basic research should the institution under their control pursue relative to quicker pay-off work for other federal entities?” To get to this answer, the intertemporal budget equation (1) needs to be combined with a current period spending constraint as in (2), where a and b represent the average development project cost in resources and the average cost of generating a publication, respectively.

$$(2) \quad C = aX + bP$$

Maximizing the resulting expression with respect to P , taking first order conditions and simplifying yields (3). In essence, the maximization rule suggests the optimal level of publication will be that where the average cost of publication multiplied by one plus the institutional discount rate is just equal to the marginal

(3)
$$(\partial B_2 / \partial X_{t+1})(\partial X_{t+1} / \partial P) = b(1 + r)$$

budgetary increase from additional projects prompted by the last publication. The level of institutional effort in producing basic research (bP) that achieves this optimization is referred to in this model as F^* .

The level of spending on basic research F^* implies a certain level of institutional effort (E^*). This effort (E) can be disaggregated into the institution's dedication of researchers (L), research capital (K) and enabling labor (H) to the research enterprise (4). E^* is the allocation of enabling labor and research capital that flows from that funding allocation given current wages and capital prices.

(4)
$$E = F(L, H, K)$$

Reflecting reality, labor and enabling labor are mobile in the short run. The stock of research capital is "sticky" in the short run.

In the national laboratory research context, enabling labor can be thought of as the individuals that maintain the capability of the research capital and allow it to be made ready for conducting experiments. In the case of the high energy physics research facilities like the Z facility at Sandia, enabling labor can include all those individuals who fabricate and lay out compression objects, calibrate diagnostic sensors, implement safety procedures, monitor environmental toxins, etc. It is important to note these people are not the scientists who design, perform, and interpret the results of experiments - all of

whom are viewed in the context of their individual levels of embodied knowledge.

There are two primary components to the research capital stock - equipment and facilities. Equipment, such as computers and relatively low cost laboratory apparatus, are viewed as variable production factors in the short run. In this treatment, facilities describe unique research equipment and its surrounding infrastructure. For example, the current best-in-class transmission electron microscope (TEM) only works properly when it is housed in a specially constructed building with a thicker foundation to dampen seismic vibration and wall and ceiling construction that soaks up acoustic vibration from ambient noise.

All three of these components are assumed to be twice differentiable. The first derivatives of institutional effort with respect to labor, enabling labor and research capital are all positive. Due to congestion effects, the second derivative with respect to labor and enabling labor have negative signs. The sign of the second derivative with respect to research capital is assumed to be negative in the short run as well. If part of a production facility is made available for research efforts, the second derivative of effort with respect to research capital will exhibit a negative sign, owing to competing uses. This same factor applies to new research-only facilities, which if significant enough, are often backlogged and exhibit congestion effects.

Now that the meaning of E^* has been established, it is necessary to explore the underlying production process behind “open science” publications within

these national laboratories. Research generated publications⁵ (designated as q in the upcoming equations) emerge from a production process (R) that combines embodied knowledge $L(A)$ with institutional effort (E^*), yielding:

$$(5) \qquad q = R(L(A), E^*)$$

The first derivatives of the research publication function with respect to labor, the knowledge embodied in that labor and open science oriented research inputs (institutional effort) are all positive. The second derivative of the research production function with respect to institutional effort is less than zero, reflecting the declining marginal physical productivity of labor, capital and enabling labor. This formulation is similar to that contained in Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), but applied at a lower level of analysis.

The sign of the second derivative with respect to knowledge is ambiguous. Whether the sign of the second derivative is positive or not depends upon the nature of the research that a given researcher is engaged in. If the researcher is engaged in what Newman (Newman, 2001) refers to as classification oriented research, the marginal productivity of knowledge is declining. If, however, the researcher is participating in an effort that can be classified as a pioneering research effort, the second derivative of the research production process with respect to knowledge will be positive. Pioneering research efforts have the ability to shift the entire research heuristic within a field, potentially leading to an explosion of new work.

⁵ Technically, this production function generates a certain amount of research publications at a fixed quality level. Achieving greater levels of quality requires greater levels of effort and embodied knowledge to achieve the same quantity.

In the majority of situations, the second derivative of research with respect to knowledge acquisition will be negative because the majority of research is classification oriented (Newman, 2001). Diminishing marginal productivity of knowledge arises from classification-oriented research because such research uses established methodology against a known problem set. The application of existing methods typically only result in incremental improvements in methods, which in turn only generate incremental improvement in knowledge over that problem set. This diminishing generation of knowledge is reflected in a declining marginal productivity of knowledge.

Following Foray (Foray, 2004), the knowledge component of this equation can be further deconstructed into the two separate components observed in (6): the stocks of explicit (A_E) and tacit knowledge (A_T). Explicit knowledge is that

$$(6) \quad A = A(A_E, A_T)$$

knowledge which is documented in some codified form. This codification allows multiple people to have access to and benefit from the knowledge, making it a non-rivalrous good. Arrow (Arrow, 1965) classically refers to this kind of knowledge as “information” and categorizes it as a public good owing to its low cost of copying, which makes it virtually non-excludable.

This characterization is misleading in this domain. There is a complicating factor associated with the public good nature of public knowledge with respect to new scientific information. Even if the new scientific information is widely distributed, access is only secured by gaining access to the journals carrying the new information by visiting institutions with physical copies of the journals, maintaining an institutional linkage with an electronic subscription or paying for such a subscription out-of-pocket. Thus, the public good nature of new scientific

knowledge is circumscribed by excludability stemming from location, connectivity and income. Even given physical or virtual access to new scientific information, only a small cross-section of the population actually can understand the subject matter, if it is valid and what implications the findings have for future work, if any. All these restrictions imply that the ability of individuals to absorb new scientific information - a private good - may be more relevant than the overall stock of public knowledge when assessing the contribution of public knowledge to researcher productivity.

Tacit knowledge may be best described by the words of Polanyi (Polanyi, 1966) who coined the phrase. Polanyi described tacit knowledge as knowledge that could not be readily codified, famously indicating “we know more than we can tell.” Tacit knowledge, also referred to as “intuition,” is viewed as the stock of knowledge that results from carrying out a task repetitively and, as such, is commonly connected to the idea of learning-by-doing.

Cardiff University sociologist H.M. Collins has further refined Polanyi’s concept of tacit knowledge into three distinct categories: *weak* tacit knowledge, *somatic affordance* tacit knowledge, and collective tacit knowledge. *Weak* tacit knowledge describes knowledge that could be made explicable with sufficient effort, but for any number of reasons that effort has not yet been expended. *Somatic affordance* tacit knowledge refers to knowledge that humans possess due to the composition of their bodies. *Collective* tacit knowledge is knowledge that emerges from the participation of humans in society (Collins, 2010).

Unlike explicit knowledge, which is viewed as a public good and is transmitted easily, a researcher’s tacit knowledge is a private good and will only be communicated with difficulty. As indicated previously, *weak* tacit knowledge

can be encoded and transmitted if enough effort is expended. In comparison, transmission of *somatic affordance* and *collective* tacit knowledge requires repeated contact over a significant period of time, creating greater excludability.

Traditionally, *somatic affordance* tacit knowledge is gained through “learning-by-doing” and transmitted via coaching while learning. Similarly, *collective* tacit knowledge is acquired through accumulated interactions with other members of a group and is transmitted primarily via apprenticeship and working closely together.

All of these observations can be summarized in (7), which integrates (4) and (6) in (5). Qualitatively, this expression can be interpreted as indicating that

$$(7) \quad q = R(L(A(A_E A_T)E(L^*, H^*, K(D^*, N'))$$

the production of publishable basic research is the result of a research process combining tacit and public knowledge embodied in labor with budget optimized levels of researchers (L^*), equipment (D^*) and enabling personnel (H^*) with a given set of research facilities (N'). What is missing in this specification, outside of a functional form choice, is a constrained mechanism for acquiring knowledge. To accomplish that task, it is necessary to give some attention to two processes: how knowledge is actually gained and how knowledge acquisition is balanced against other work activities.

Further decomposition of the knowledge variables is required to reveal the acquisition process. Given the tacit knowledge acquisition mechanism presented earlier, an individual’s tacit knowledge emerges from past research experience (r) and what diffuses from their first degree research network, known as their first degree neighborhood (g) as shown in (8). Parameters modify

$$(8) \quad A_T = A_T(r, g)$$

each of these variables (γ, ρ) - restricted to values between 0 and 1 - which signify that both past experience and existing network connections will not be perfectly applicable to any new research environment. To simplify consideration, public knowledge (A_E) will be assumed to be constant for all researchers. This assumption is based on the idea that in every technical field there is a common body of knowledge that all researchers who publish work in refereed journals possess.

Substituting (8) in (7) yields the researcher's objective function. The key constraint on the production process is the researcher's time T , which can be observed in (9). The researcher either performs research or makes/maintains contacts in this simplified conception. Having a greater number of research partners shrinks average project duration (F) by allowing for research task

$$(9) \quad T = F / (1 + g) + mg$$

specialization. That advantage, however, must be balanced against the time cost of maintaining each of these relationships (mg).

Using this objective function and constraint in a conventional Lagrangian constrained optimization yields the following optimization condition (10). In this

$$(10) \quad F(1 + g) \frac{dA_t}{dg} = (F + m(1 + g)^2) \frac{dA_r}{dr}$$

specification, the marginal rate of tacit knowledge transfer to the whole network from adding a new research associate within the project's duration will just equal to the marginal rate of tacit knowledge gain from performing the project and maintaining the existing network at the optimum. This condition is consistent with the principle of preferential attachment among research colleagues as individuals with the largest networks and research experience are likely to be the

first connections sought because they add the most tacit knowledge to a research activity.

This theoretical researcher productivity model suggests several behavioral patterns that should exist in any general research production context where researchers perform research under managerial supervision. Higher levels of institutional effort should lead to greater numbers of observed publications. Individual researcher productivity should increase with greater past publication experience and connectivity will be more productive than more isolated authors with shallower publication records. That same productivity will decline if the institutional discount rate increases, such as when a research organization community is reorienting its capabilities in order to meet a critical production deadline, or if the work-for-other government agencies' return on generating publications declines, such as when a research organization faces conditions of project congestion.

For the question under consideration in this study, this model suggests that researcher collaborations within research institutions will be more productive than researcher collaborations between two research institutions. This productivity impact is due to the issue that the cost of maintaining internal collaborators is lower than maintaining external collaboration partners. As such, the network neighborhoods of researchers at both US national laboratories and Russian research institutes will be likely to be biased toward having more internal collaborations partners than external collaboration partners. The external collaboration partners that are selected, however, should make significant enough contributions to project completion that the productivity difference that could be gained by adding an additional Russian coauthor or a

coauthor from the US should not be significantly different from zero. This expectation should stand for “atomistic” collaborations where individual scientists at Russian research institutes and US national laboratories elect to collaborate in response to their own research questions.

National laboratory researchers also participate in joint scientific efforts where their collaboration partners are chosen for them. This governmental intervention generates one of three states in participating researchers. In the most likely case, researchers participating in the collaboration will have greater numbers of external partners than they would possess in an atomistic collaboration situation. As such, adding an additional collaborator of that type should result in productivity loss. In a less likely scenario, researchers participating in the collaboration could have fewer numbers of external partners than they would possess in an atomistic collaboration situation. As such, adding an additional collaborator of that type would result in productivity gains. Or, finally, and least likely, the governmental intervention could produce the same researcher network neighborhood mix as under atomistic collaboration, yielding no productivity impact at all.

Given the perceived likelihood of each scenario, it is likely these “bilateral” collaborations should be less productive than collaborations that come about through the practice of normal science. The aggregation of these inefficiencies across researchers should magnify the productivity impact at the institutional level. Thus, institutional level collaborations dominated by these “bilateral” collaborations between sets of researchers at US national laboratories and Russian research institutes should also be less productive than collaborations these researchers form with colleagues at their own institutions.

III. National Laboratories, Knowledge and the International Collaboration Interface

The Conventional Wisdom Behind Collaboration

At the fourth Decade of the Mind conference, James Olds, the current director of the Krasnow Institute of Advanced Studies at George Mason University, made a categorical statement concerning the linkage between international scientific collaboration and research quality in his field of neuroscience. In opening up the technical portion of this conference, Olds declared “science is only high quality when it is internationalized.” Olds’ opinion was not merely a comment made in passing - he underlined this same sentiment at the end of his remarks that “truly excellent science is always international.”

What is interesting about Olds’ statement is how it ignores the progress of research even within his own field. The research presented at the conference he opened with this set of statement alone is testimony to the inspired work of local

research teams composed of individuals from a single country. For example, one of the key developments in cognitive science, the concept of associative memory was developed by Jay McClelland and David Rumelhart at Stanford -- both American academics. The connectionist revolution in cognitive science that resulted from this work is transformative, but not the result of international collaboration. In more conventional physiological neuroscience, James Albus extended David Marr's work on the cerebellum to form the Marr-Albus model of the cerebellum, a still relevant construct forty years after its initial publication. The joint model could be considered to be a serial international collaboration, as Marr was British and Albus was American, but the two scientists never coauthored a paper together.

These examples do not suggest that collaborations do not lead to good work. While McClelland's work with Rumelhart started the connectionist revolution, it was McClelland's work with Thomas Rogers - a Canadian psychologist -- that popularized this approach to memory. Some of the current work with the greatest potential impact -- Christof Koch's collaboration with Francis Crick on a theory of consciousness -- meshes a German biophysicist with a British biochemist has already led to path breaking developments in the understanding of the frontal motor region of the brain and promises more. Work at the Howard Hughes Memorial Institute's Janelia Farm is premised on the necessity of international collaboration and is attempting to transform the study of neuroscience by mapping how genetics affects brain structure via comprehensive work in fruit flies using what are known as "forward genetics" techniques - themselves a major breakthrough over lesion and fMRI studies.

This opinion is not merely held by individuals working in the field of neuroscience. It is understandable how the idea that international collaboration and high quality research go hand-in-hand came to be. Any study of the history of science in the 20th century shows that teams of scientists whose members come from different nations -- and academic traditions -- can accomplish dramatic technical feats and make significant discoveries. The Manhattan Project is just one of many such international projects that could be cited in defense of this perspective. The Human Genome Project is a more recent example in the same vein.

This conventional wisdom does not match well with the observation that innovative research and development activities tend to occur in geographic clusters. The best current example of such clustering is Silicon Valley in California. Silicon Valley's success has been attributed to a unique confluence of the region's history, relatively easy access to venture capital markets and large number of knowledge workers. The region was noted for its innovation prior to attracting large numbers of migrants with technical backgrounds (Saxenian, 2007). Similar stories could be told about British industrial districts that dominated the textile industry in the nineteenth century (Lazonick, 2005).

There are multiple reasons cited for why these geographical innovation clusters occur. Saxenian observes that knowledge workers in Silicon Valley tend to work in start-up firms to develop skills that might be useful to them as they pursue their own start-up firm in the future (Saxenian, 2007). Lazonick hits a similar note by connecting the dominance of British industrial districts to the ease with which a worker could gain an apprenticeship with multiple concerns (Lazonick, 2005). As indicated in the past chapter, this side-by-side work is one

of the key ways that tacit knowledge can be transferred from one person to another. Geographic proximity makes this tacit knowledge transfer easier by making repeated contact easier, improving the ability of the novice to imitate the master and allowing for the novice and master to experience the same environment.

Given this set of observations, it is more likely the idea of international collaboration producing excellent research is simply an artifact of joint research projects where the collaborating partners have different, but complementary sets of embodied knowledge that help them produce highly cited research. As such, the nationality of a given researcher may be just a proxy for the kind of human capital investment typically made within a given country's educational establishment. In this kind of perspective, similarity between researchers in their background and technical competence will tend to generate research projects that are incremental that do not push the boundaries of technical knowledge. However, if researchers are different in the tradition of their training, they will be more likely to conduct research that is more likely to be at the frontier of knowledge when joined together.

To see how this complementarity might be generated through the process of developing a publication record, it is important to acknowledge that different nations have educational establishments that make a distinct set of choices which impact what a typical researcher from that country will have expertise in and what level of domain expertise they will have. On average, Russian and Chinese researchers tend to have a higher level of mathematical competence than their American colleagues. Both the Russian and Chinese educational establishments place high premiums on the development of mathematical mastery, and that

mastery is often required to advance to unique topical studies. By and large, American researchers tend to have an advantage in domain expertise, due to the primacy of mastery of the technical field over the mastery of mathematics in the American system.

Similar national difference in attributes can be observed in experimental fields. American scientists tend to have greater experience than their counterparts in developing facilities for conducting unique, but replicable large-scale experiments. This expertise at replication manifests itself as leadership in particle and plasma physics research fields. Due to comparably relaxed environmental and safety standards relative to Western standards, Russian scientists tend to have an advantage in conducting hazardous field experiments. This relaxed attitude toward work with hazardous materials allows Russian scientists the unique opportunity to conduct complex real world experiments and, as such, develop tacit knowledge about the physical processes.

If different regulatory environments can encourage different types of scientific competencies, it should also be true that distinctly different scientist life cycles should have consequences for the kinds of countrywide competencies that a given country is likely to demonstrate. The lack of a formal retirement structure for scientists and engineers in Russia means that research careers often span twice the years of an American or Western European researcher working in the same field (Yegorov, 2009). If the lack of environmental regulation and enforcement generates a Russian advantage in technology that advances from explosive tests, the longer careers of Russian scientists should yield a Russian advantage in innovation that emerges from the individual accumulation of tacit knowledge. This age structure gives the Russian system some institutional

memory that the American and European systems do not have - a clear mechanism to revisit early research paths in the light of new enabling technology.

While not addressed in this study, this same thinking can be applied to the other font of tacit knowledge for researchers – their network neighborhoods. If two researchers have distinctly different network neighborhoods with few shared ties, the collaboration between the two potentially will be able to benefit from a significant amount of breadth in the interests and competencies of the researcher communities their collaboration would tie together, but a global search of that community may take a significant amount of time. A similar researcher pair with a large number of shared ties within their joint network neighborhoods will have less breadth of knowledge, but any queries for relevant knowledge should be answered more quickly.

Selection of Collaborations of Interest

Examination of the entire set of DOE/NNSA national laboratories' historical collaborations with all Russian research institutes would be a time-prohibitive task for even a team of researchers. On the other hand, limiting the collaborations of interest to this study to just those between national laboratories with well-known nuclear weapons activities would significantly reduce the degree to which the results of this study can be more broadly generalized. Such an approach would also limit the insight this study could derive from understanding collaborations between these highly national benefit oriented national laboratories and other national laboratories with greater open science orientations.

This study takes a middle path and chooses to focus on “best case” collaborations between Russian and U.S. national laboratories that could illuminate the importance of the input knowledge complementarities in the individual research publication production process as outlined in the previous section. In aggregate, this selection criterion pushed this study to select nine pairs of collaborations between three DOE/NNSA national laboratories’ collaborations with similar Russian institutions.

All three U.S. nuclear weapons laboratories’ relationships with the Kurchatov Institute of Atomic Energy, which is often thought of as Russia’s “Los Alamos” equivalent - are examined. Kurchatov is the epicenter of Russian nuclear fusion work. It has recently emerged as a nanotechnology research and development center following investment by the Russian joint stock company Rusnanotekh.

This study also examines Los Alamos’ and Lawrence Livermore’s collaborations with two laboratories in two of Russia’s formerly closed cities - the All-Russian Scientific Research Institute of Experimental Physics (VNIIEF) and the All-Russian Scientific Research Institute of Technical Physics (VNIITF). In the late 1990s, Los Alamos’ collaborated with VNIIEF to conduct explosive tests of magnetic flux compression generators in Siberia as part of a bilateral joint inertial fusion collaboration. These tests would not have been allowed in the United States due to “unacceptable environmental impact” (Reinovsky, 2005).

While Sandia does have limited ties to these two Russian nuclear weapons institutes, their late emergence in the data set makes statistical analysis relatively unfruitful. Thus, Sandia’s ties to two other Russian research institutes are considered - the P.N. Lebedev Physics Institute in Moscow and the D.V. Efremov

Research Institute in St. Petersburg instead. The P.N. Lebedev Physics Institute is Russia's oldest physics research institute and is noted for its broad experimental and theoretical contributions to condensed matter physics. The D.V. Efremov Research Institute, on the other hand, possesses a far narrower focus on developing particle accelerators and plasma containment devices.

Notably, this set of institutional dyads offers an opportunity to examine institutions with a variety of open source / national benefit balance positions. The Los Alamos-Kurchatov Institute of Atomic Energy dyad is the relationship that should demonstrate the most open science character on both sides. Conversely, the Lawrence Livermore – VNIITF dyad brings together where both institutions are driven by national benefit. Of the three US nuclear weapons laboratories, Sandia tends to maintain the most national benefit focused approach, but its collaborations with Russian laboratories tends to be with institutions that follow – comparatively at least – the collaborative open research model outlined above.

Knowledge Network Attributes and Transfer Dynamics

To study knowledge transfer across these US-Russian national laboratory collaboration linkages of interest, it is important to gain an understanding of the observable static network topology, the research topics examined by each institute, and the dynamics associated with these networks' evolution. In this case, a credible understanding of the static network topology associated with these collaborations can be gained by visually mapping the research communities contained within each institution, the co-authorship network associated with the publication of basic research emerging from joint collaborations between these selected US and Russian laboratories and by

characterizing these same networks in terms of their key nodes, connectedness, and degree distribution. A similar understanding of research competencies can be obtained by mapping the topics of institutional research publications via visualized latent semantic analysis. The key concern of the dynamics section will be an examination of the apparent mechanisms driving the inclusion of new U.S. and Russian laboratory scientists into these networks to see how well these additions conform to the preferential attachment model results observed in other scientific collaboration networks (Barabasi and Albert, 1999).

Each of the visualizations of institutional research communities and joint collaborations between selected US and Russian nuclear weapons laboratories are based on relationships demonstrated via co-authorship of articles in international, peer-reviewed journals. A node on these graphs represents a scientist who participated in these collaborations. Each linkage represents co-authorship of at least one publication joining two scientists. The size and color of each node vary by a nodes connectedness to other nodes (i.e. its degree). Larger and more red nodes (or links in larger mappings) correspond to scientists with more observed co-authors. Where practical, author labels are applied to nodes.

The distribution of the nodes in these graphs is generated through the use of the Fruchterman-Reingold layout algorithm. This algorithm attempts to display nodes in such a manner that the nodes possess links with uniform lengths between them that do not overlap. One of a class of what are known as “force directed algorithms,” the Fruchterman-Reingold layout algorithm virtually replace each node with a conceptual “proton” and replaces each edge with a balanced “nuclear-weak force, nuclear strong force” to form a system where nodes repulse each other when they get too close, but attract each other

when they are connected (Fruchterman and Reingold, 1991). In this case, the nodes are initially clustered in a square layout and the layout allows the “nuclear forces” on the virtual “protons” to move the system over the course of a pre-designated number of iterations of the algorithm. While it was not the objective of the designers, a key advantage of this layout algorithm is the nodes that are most central in a network are almost always in the center of the plot.

Fruchterman and Reingold’s original algorithm is included for completeness in Appendix I.

Constructed using five years of co-authorship links between authors with an affiliation to one of the institutions under examination in this study, the following visualized institutional research communities are actually depictions of the entire network neighborhoods surrounding a given institution. These visual representations show the institutions under consideration in this study vary with respect to the connectivity of the entire set of researchers at a given institution, number of research communities within that institution and how those researchers are clustered. The five-year window used in these visualizations is an attempt to expose salient, but latent relationships with researchers who may be carrying out national benefit activities that may still influence researchers. As such, this multi-year perspective yields a better-informed picture as to how connected researchers (and research communities) within these institutes truly are.

Mapping out national laboratories’ research and development portfolio starts with attempting to understand the kind of publicly available, authoritative and representative data source that could provide a window into the entire portfolio of research being conducted by the national laboratory. Given the

external-to-internal environment information exchange dynamic that appears to be present in many of these institutions, a dataset including the title, authorship, abstract and publication date of all modern, peer-reviewed publications produced at each facility should possess each of these necessary characteristics (such as that available from Thomson ISI Web of Knowledge). Constructing a full publication record (i.e. a document corpus containing the full-text of every publication) is unnecessary, and indeed, unhelpful, because even the most modern semantic analysis algorithms have difficulty carrying out operations other than searching for keywords and their usage context in large numbers of documents containing significant amounts of highly unstructured text.

Once this corpus is developed, this assembled text can be analyzed via any of a number of computational linguistic techniques. Of these techniques, latent semantic analysis is probably the most widely applied and is used in this examination. Latent semantic analysis (LSA) evolved from earlier vector space models, and attempts to draw meaning from large document sets based on the singular value decomposition of a term-document matrix derived from this set.

LSA is an unstructured text analysis technique that attempts to identify themes in a corpus of documents through the examination of the terms used in those documents and how similarly they are used across the set of documents. In its most basic form, LSA involves representing a set of documents (titles and abstracts of scientific publications) as a document by term frequency matrix. This matrix is then subjected to singular value decomposition to generate rank restricted singular values. When multiplied by the term by rank matrix, a term vector is formed that can be used to calculate term similarity to other terms via the application of the Euclidean distance formula or cosine similarity. Similarly,

when the singular values are multiplied by the document by rank matrix, a vector is generated that can be used to measure the similarity of the documents. This approach to unstructured text analysis was first proposed by Landauer (1988), but only gained popularity as electronic processing and storage became less costly due to the large memory requirements necessary for processing even sets of small documents.

To improve the ability to comprehend the resulting similar research clusters, these clusters are presented in graphical form. Many implementations of latent semantic analysis force the analyst to conduct a textual clustering analysis, assign labels in a qualitative fashion, and then seek the confirmatory opinion of a subject matter expert. This method is time-consuming and subject to significant variance based on the domain knowledge of a given analyst (Hendrickson, 2009). Graphical representation, depending on the layout algorithm being used, can give the analyst an additional sense for how similar each of the clusters are to each other. If the interest is in assessing research programs, much can be learned by mapping topical clusters, as is performed later in this chapter. If the focus of an analyst's interest is acquiring an understanding of what is known as "enabling technology," (i.e. the research equipment used to carry out the research and understand key chokepoints) mapping clusters formed from publication abstracts would be appropriate.

There is a vast array of graph layout algorithms that would be amenable to this kind of task. Some early efforts used a graphical implementation of the Boltzman algorithm (see Borner, 2000 for an example). Others use a force-directed algorithm like the Kawai Kamada graph layout algorithm (a representative treatment can be found in Zhu and Chen, 2007). This study uses

an implementation of VxOrd, a force-directed graph layout algorithm that preserves both global and local structure for a range of graph sizes (Boyack and Rahal, 2004).

Unlike the visualizations of institution research communities, the visual co-authorship mappings are not constructed using a five-year window. Instead, these interfaces between institutional research community network neighborhoods are presented in their entirety. This choice is purposive as the longer time horizon yields valuable information into the kind of collaboration that is taking place between the two interfacing research institutions.

There are three distinct types of collaborations that can be observed in these visual co-authorship mappings – atomistic, bilateral institutional and large-scale multi-national collaborations. Atomistic collaboration occurs when two researchers spontaneously form research relationships primarily as a result of the preferential attachment mechanism proposed by Barabasi and others. Bilateral institutional collaborations emerge when two research institutions perceive an advantage from forming a research partnership. Participation in large scale multi-national collaborations is typically motivated by the desire to move fundamental scientific understanding forward when the cost of carrying out the research to drive that understanding forward is beyond the means of most nation-states separately such as in the high-energy particle physics case discussed in the first chapter.

It should be noted that institutional decision-maker cognizance of collaborative activities varies across these collaboration types. Atomistic relationships may or may not be sanctioned by their broader institutions. In comparison, institutional decision-makers are often instrumental in driving

bilateral relationships forward. Like the bilateral relationships above, participation in large-scale international collaborations usually occurs with institutional decision-maker assent. However, unless the research institution is developing new facilities to support such an effort, it is less likely that institutional decision-makers will be as involved or as informed as in the bilateral case.

This distinction between participation in scientific collaborations involving atomistic, small numbers of participating institutions and large-scale international collaborations has implications for a given institution's likelihood to take part in collaboration. For example, it is less costly to maintain a large number of acquaintanceship ties (which are viewed as weak ties) in these large international collaborations than the intermediate strength ties required to form and maintain a bilateral research collaboration between two national laboratory partners or the strong ties required for atomistic collaboration. As such, there should be a significant productivity difference between research organizations that favor participation in massive international collaborations over engagement in strategic bilateral partnership or encouragement of atomistic collaboration.

Institutes, Research Communities and Collaborations

The following section reviews each of the Russian institutions under consideration in this study and their research collaborations with the US national laboratories discussed in Chapter 1. It begins with a brief description of each Russian research institute, their associated research communities, institutional research foci and their collaborations with each relevant US national laboratory.

Each section attempts to address whether the collaboration dynamics associated with each relationship are consistent with expectations.

The Kurchatov Institute of Atomic Energy

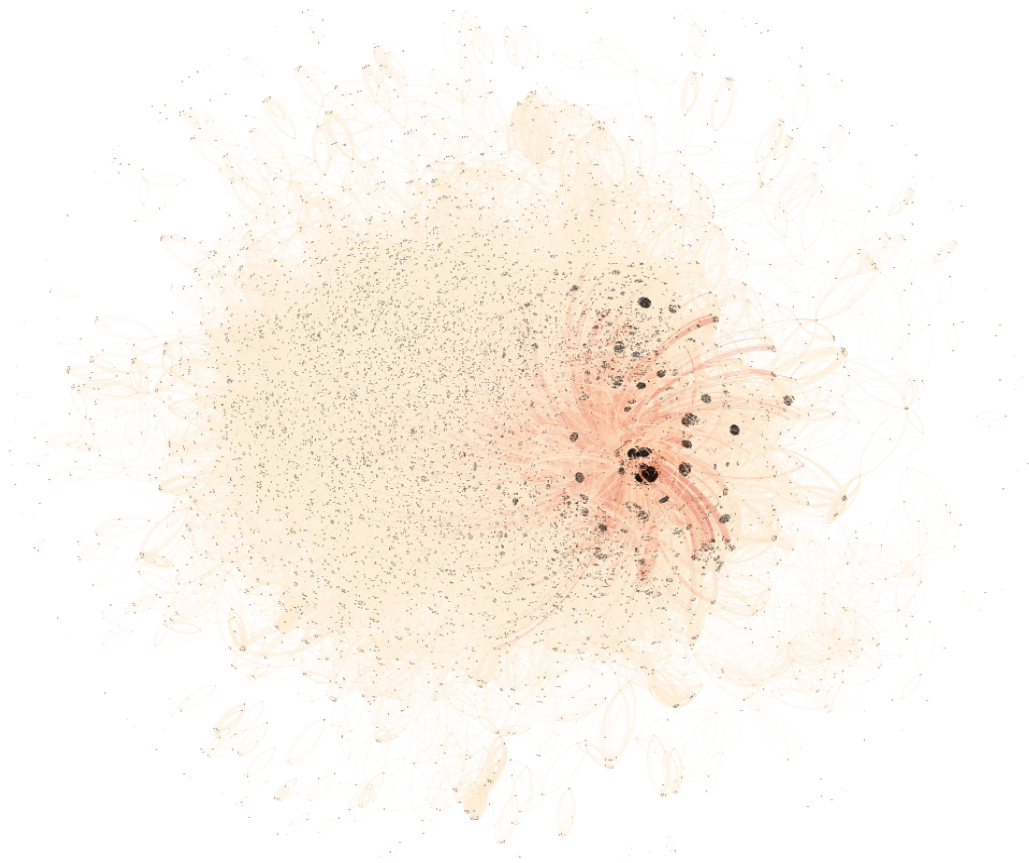
The Laboratory No. 2 of the USSR Academy of Sciences was founded in Moscow in 1943 in order to develop a nuclear weapon for the Soviet Union. Renamed the Kurchatov Institute of Atomic Energy in 1960 to honor Igor Kurchatov, the director of the Soviet Union's nuclear weapons effort, the Kurchatov Institute is known for its contributions to developing Soviet thermonuclear bombs, the first nuclear reactor to contribute electricity to a power grid, as well as atomic reactors for icebreakers, submarines and space vehicles. Since the fall of the Soviet Union, the Kurchatov Institute – like many of its U.S. nuclear weapons laboratory counterparts – diversified its set of competencies and is pursuing research programs in nanotechnology, cognitive sciences and biology in addition to its nuclear science programs.

Static network analysis of the Kurchatov Institute's research community in 2012 shows a network neighborhood composed of 22,354 researchers linked by 5,067,800 co-authorship relationships into 167 network components that can be segregated into 280 distinct communities. In this context, components are subgraphs from which all nodes are reachable via a traverse across a known edge. Communities, on the other hand, are densely connected groups of nodes bounded by more sparsely connected nodes (Blondel, Guillaume and Lambiot, 2008).

Together, these two network statistics imply the open research community within Kurchatov is only linked in a diffuse manner. The graph of the Kurchatov

Institute's institutional research community (Figure 7) is consistent with that observation. As can be seen by a small concentration of red edges in one central region of the graph with orange links permeating the rest of the network structure, most members of the Kurchatov Institute research community are only lightly connected with only one real area of dense connection in the graph. This pattern of diffuse connectivity and absence of ties can be interpreted as being evidence of *structural holes* in the Kurchatov Institute research community which

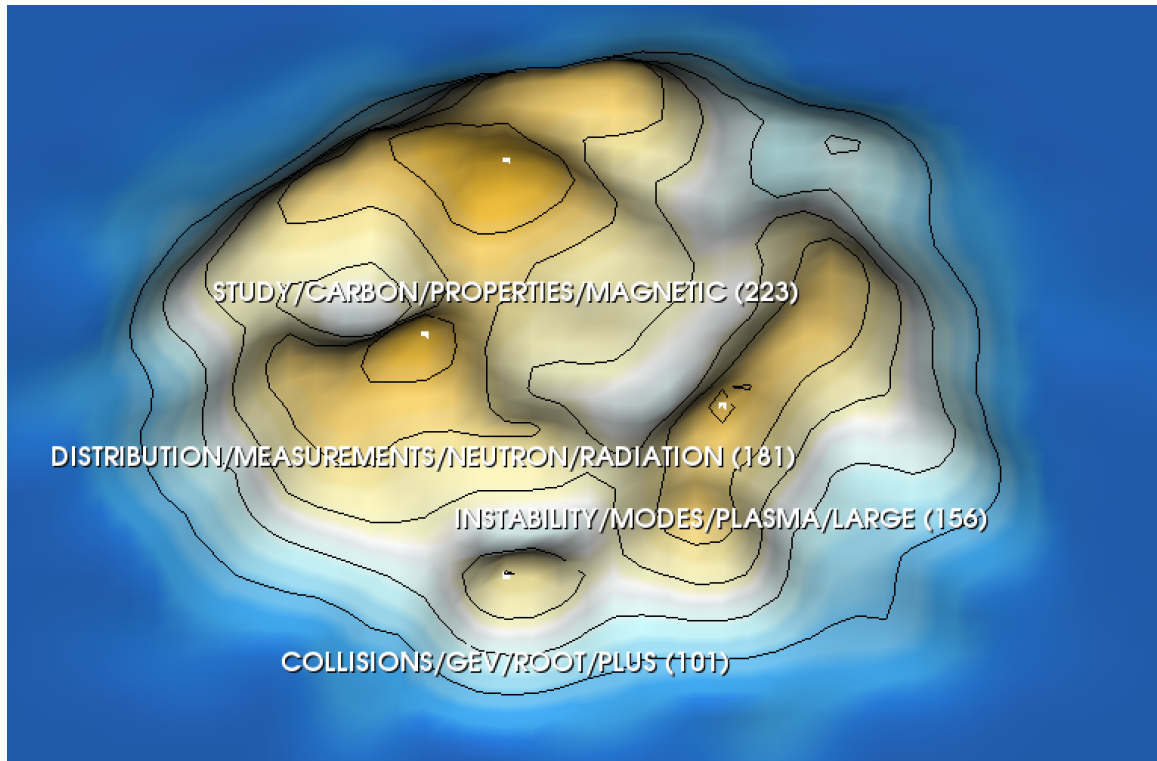
Figure 7. Kurchatov Institute of Atomic Energy Research Community, 2012



likely limits the diffusion of tacit knowledge between community members and creates the opportunity for multiple, overlapping competencies to exist (Jackson, 2008).

The densely connected community in the above graph corresponds to a group of individuals carrying out research in the field of particle physics. Like many other elite research institutions, the Kurchatov Institute participated in the design and development of the ALICE experiment within the Large Hadron Collider (LHC) at CERN to conduct quark gluon plasma research. As latent semantic topographical map shows in Figure 8 below, the study of particle physics (as depicted with by the peak labeled “Collisions/GEV/Root/Plus”) is actually a relatively minor pursuit at the research institute. Instead, the research institute devotes a significant proportion of its research activities to nuclear fusion, such as study of tokamak plasma instabilities (indicated by the

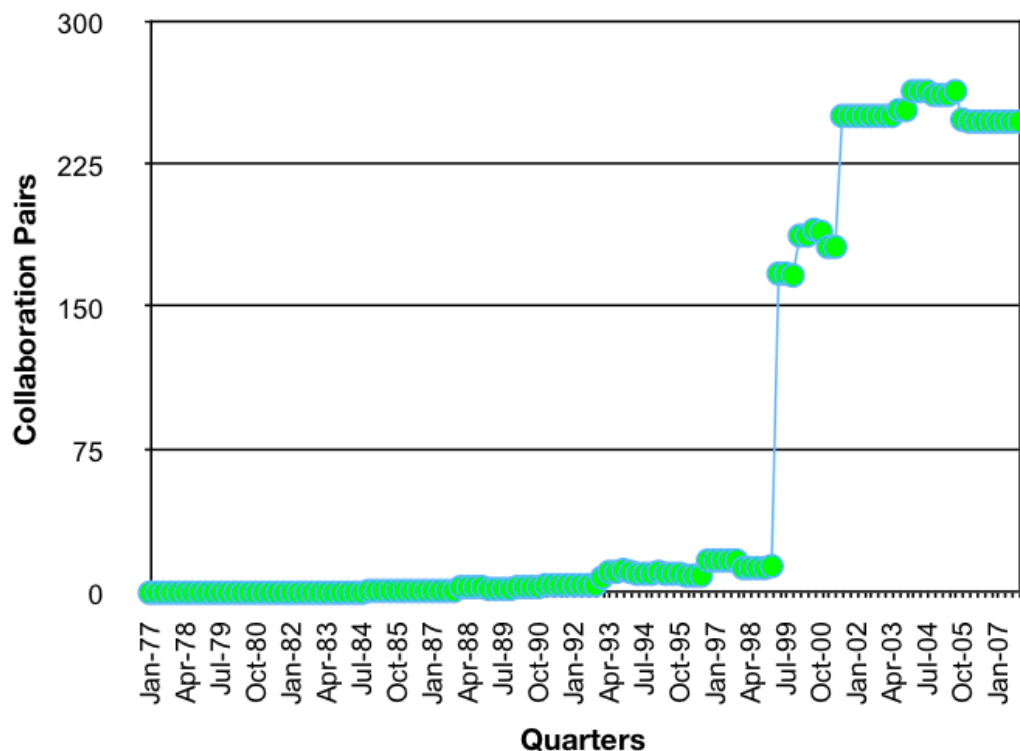
Figure 8. KIAE Open Research Activities, 2008-2012



“Instability/Modes/Plasma/Large” peak) and how the ablation of plasma facing components in the tokamak affects magnetic fields in the plasma (“Study/Carbon/Properties/Magnetic”).

Based on the publication record, Los Alamos and Kurchatov began collaborating in the early 1980s – shown in Figure 9. This collaboration focused on reviving nuclear fusion research coordination that had largely ceased in the late 1960s. These interactions intensified in the later 1980s when Mikhail Gorbachev became General Secretary of the Communist Party of the Soviet Union and instituted his program of *glasnost*, or “openness,” which opened up the Soviet Union to both international cultural and scientific exchanges.

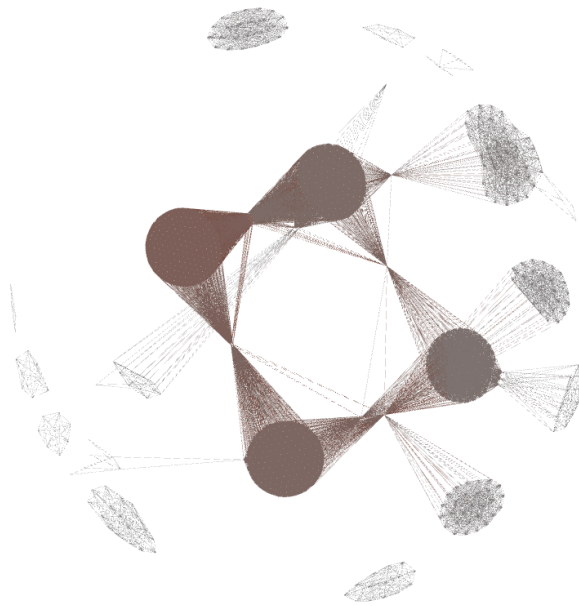
Figure 9. Los Alamos-Kurchatov Collaboration Pairs (1977-2008)



Examination of the network topology associated with these collaborations is instructive. If the last five years of the period of interest are excluded (i.e. 1977

-2008), the graph of the co-authorship networks these collaborations are nested in (as shown in Figure 10) shows eight connected giant components. There are also 17 discrete communities of densely connected groups of nodes bounded by more sparsely connected nodes (Blondel et al, 2008). This structure gets lost if the last five years of research collaborations gets added to the mix, because the k-core structure associated with the LHC quark gluon plasma experiments mask all other structure.

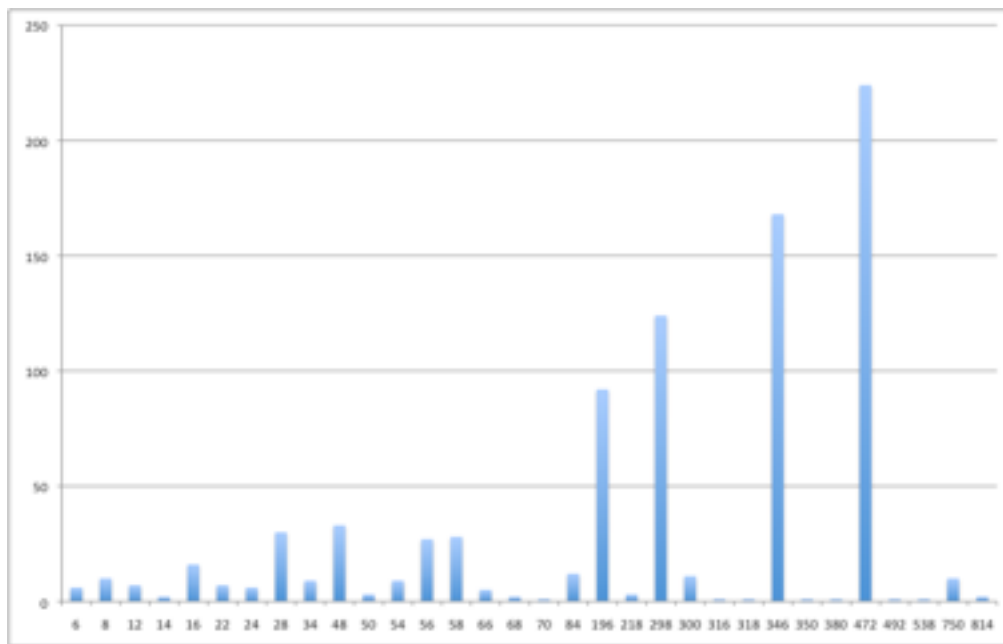
Figure 10. Los Alamos-Kurchatov Joint Scientific Publication Coauthorship Networks (1977 -2008)



Visual inspection of the degree distribution shows that nodes in this network tend to be linked via assortative mixing – nodes with similar connectivities or degrees tend to link to each other (Barrat et al, 2006). Inspection of the degree distribution contained in Figure 11 indicates a power law is not in effect in these collaborative interfaces as the figure shows a large number of very

well connected nodes and relatively few lightly connected nodes. The most well-connected Los Alamos and Kurchatov scientists in the network in this earlier time period are associated with the JT-60, a Japanese Atomic Energy Research Institute (JAERI) operated magnetic confinement fusion reactor: Sergei V. Neudatchin for Kurchatov and Glen Wurden for Los Alamos. Four key external figures -- T.J. Renk of Sandia, David R. Smith of Princeton Plasma Physics Laboratory, Seiji Ide of the JAERI Naka Fusion Research Establishment, and Ki Sang Lee of Gangneung-Wonju National University in South Korea -- play uniquely significant bridging roles in the network and connecting these disparate research activities together.

Figure 11. Los Alamos-Kurchatov Joint Collaboration Degree Distribution



Co-authorship ties between large numbers of co-authors in a global collaboration are necessarily weaker than a corresponding set of ties between small numbers of local co-authors (Newman, 2001). Indeed, it would be difficult

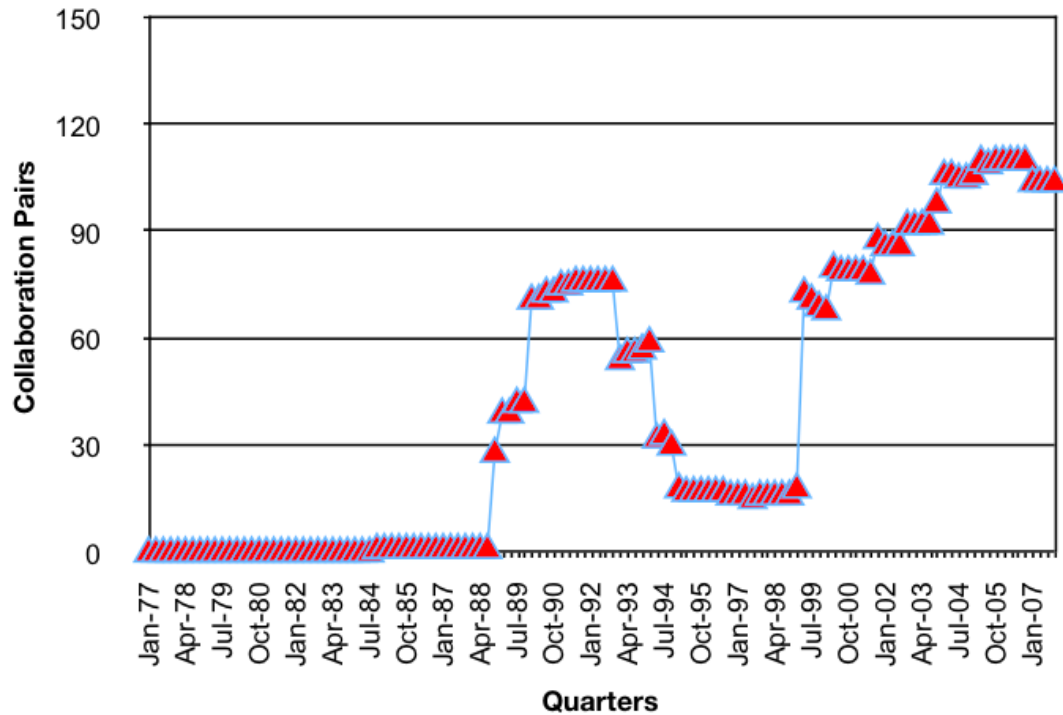
to argue that over 470 coauthors, such as all of those scientists who participated in the JT-60 have socially significant, symmetric ties. However, the number of repeated publications by this group suggests at least some acquaintanceship between most of these individuals even if they are not strong ties as depicted. Clearly, the different teams at different facilities contributing specialized elements to this project possess strong ties given the nature of the JT-60 task and its connection to the multi-national ITER magnetic confinement fusion project.

Dynamically, the subgraphs containing the collaborating Los Alamos and Kurchatov scientists show that these researchers tend to add new linkages through a preferential attachment mechanism in this community. Kurchatov scientists had an average of 760 unique co-author linkages when beginning their collaborations. By comparison, their Los Alamos colleagues only possessed 620 such observed links. While this measure may actually substantially over-count actual numbers of collaborators due to the evolving naming conventions for scientific publications of the time, this difference implies that in the absence of political coordination, scientists from Los Alamos should gravitate to their Kurchatov counterparts in the natural conduct of science.

There appears to be a unique complementarity at work in this relationship. While it is true that the Kurchatov scientists may be better connected than their Los Alamos counterparts, the Los Alamos scientists appear to have far greater publication experience. The average Los Alamos scientist participating in these interactions possessed just over 36 publications at the initiation of these relationships. Scientists from Kurchatov only possessed half of this experience on average (i.e. just under 17 publications).

Lawrence Livermore reinvigorated its collaborations with Kurchatov around the possibility of developing nuclear fusion as an energy source at an International Atomic Energy Agency (IAEA) forum in Yalta in 1986. This topic was also the impetus for early post-Cold War collaborations between the two

Figure 12. Lawrence Livermore - Kurchatov Collaboration Pairs (1977-2008)

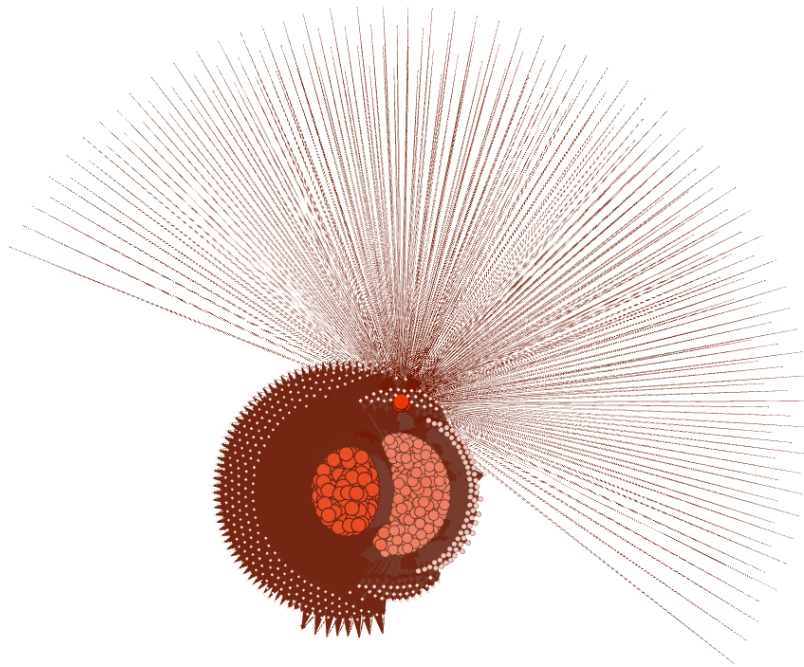


organizations having to do with the D-IIID Tokamak at General Atomics and the T-10 Tokamak at Kurchatov. In the mid-1990s, the topic set for this collaboration diversified to include free electron lasers and radionuclide monitoring, but dipped in the number of collaboration relationships as can be seen in Figure 12. The last decade has been dominated by joint work on the PHENIX detector for the Relativistic Heavy Ion Collider (RHIC) housed at Brookhaven National Laboratory in New York.

This latter collaboration dominates the joint scientific publication co-authorship network these institutional ties are embedded in (as depicted in

Figure 13 below). Unusually, despite the length of time these collaborations have been going on, this co-authorship network displays an unusual level of continuity. The resultant graph shows only one connected giant component and one discrete community. This topology is similar to what can be observed in the LHC ALICE experiment.

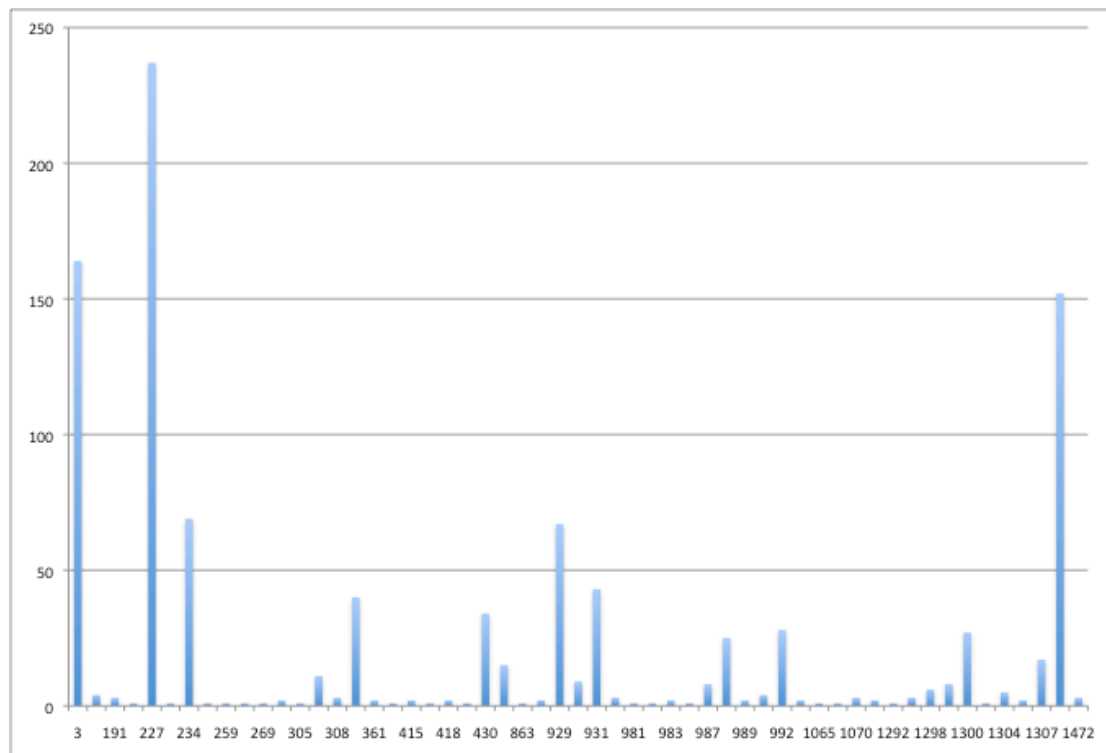
Figure 13. Lawrence Livermore-Kurchatov Joint Scientific Publication Coauthorship Network (1977-2008)



There is additional structure worth noting about this graph in the way that nodes are clearly partitioned between center core, center-periphery, periphery and “tendrils” node partitioning. This partitioning comes through in the degree distribution contained in Figure 14. Instead of following a power-law, there are peaks in this distribution at 3, 227, 929 and 1307 linkages. This multi-peak distribution is a signal there is a transitioning, but central group of scientists at the heart of these collaborations who participate in most of these relationships.

Clearly, the most well connected Lawrence Livermore and Kurchatov scientists in this network are associated with the PHENIX detector work for the RHIC magnetic confinement fusion experiment: S.L. Fokin, A.V. Kazantsev, V.I. Manko, A.S. Nyanin, A.A. Vinogradov, I.E. Yushmanov for Kurchatov and A. Enokizono, E.P. Hartouni, M. Heffner, S.C. Johnson, J. Klay and J. Newby for Lawrence Livermore. Unlike the Los Alamos-Kurchatov co-authorship network, there are no unique bridging roles in the network connecting these disparate research activities together. The density of ties is such that this activity is difficult to interrupt.

Figure 14. Lawrence Livermore -Kurchatov Joint Collaboration Network Degree Distribution (1977-2008)

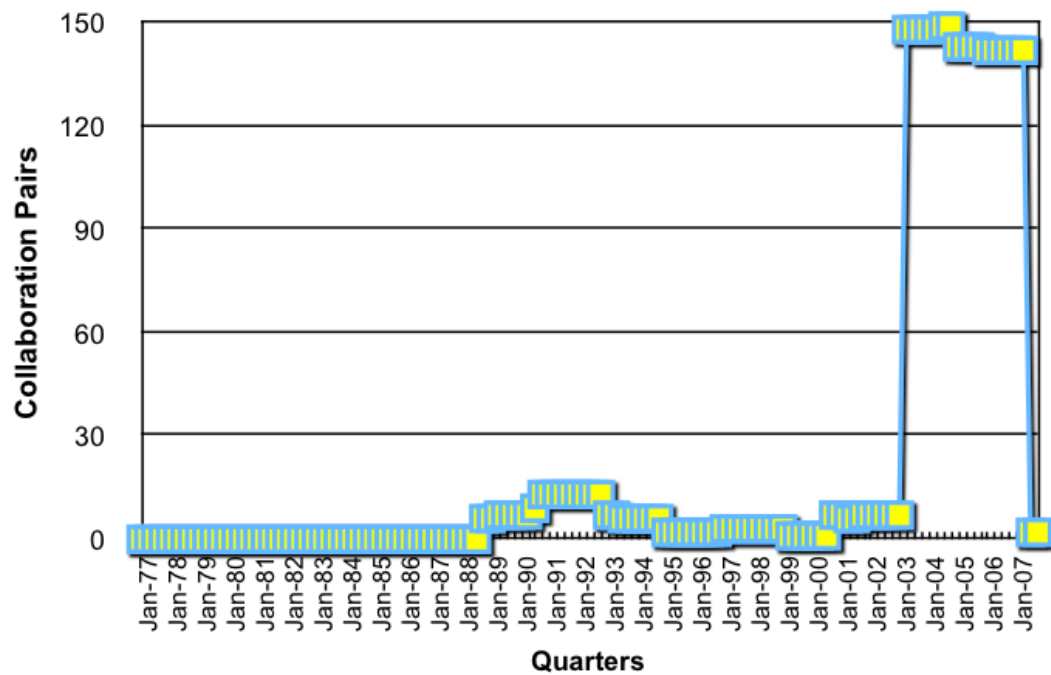


Both Lawrence Livermore and Kurchatov scientists were less well linked to their respective co-authorship communities than their Los Alamos-Kurchatov colleagues when their collaborations were initiated. The average Kurchatov and

Lawrence Livermore scientist share similar connectivity (i.e. 371 linkages to 378, respectively) and publication experience (18 publications to 20). The distribution of these two attributes across these sets of collaborators are similar in publication experience with the best-connected Kurchatov scientists considerably better connected than their most well connected Livermore colleagues. Collaboration-by-collaboration assessment suggests assortative mixing is occurring, with new scientists attaching to scientists of similar connectivity. There are relatively few instances of preferential attachment, which is contrary to most observations of scientific collaboration networks (Newman, 2001).

Like Lawrence Livermore and Los Alamos, Sandia entered into limited collaboration with Kurchatov with a focus on magnetic confinement fusion in the early post-Cold War period. This technical collaboration continued at a low level

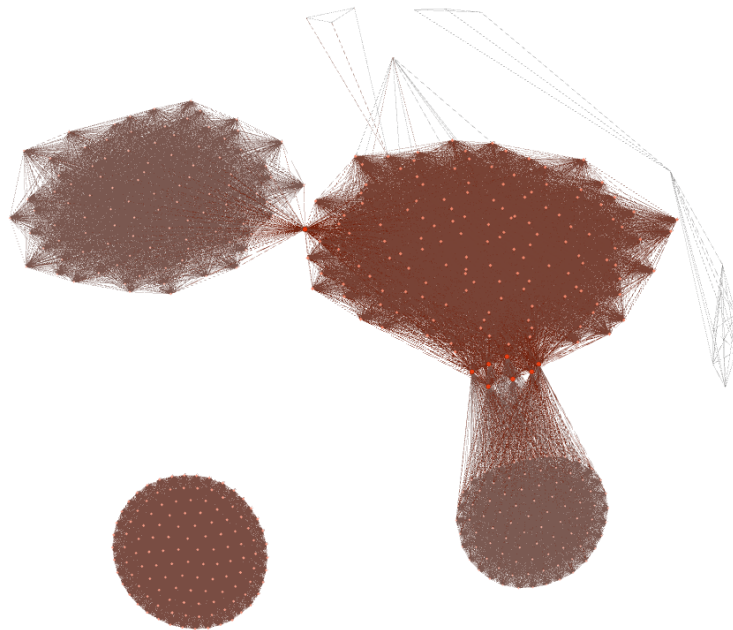
Figure 15. Sandia-Kurchatov Collaboration Pairs (1977-2008)



for several years as Sandia carried out its work on the D-IIID tokamak at General Atomics in San Diego and the National Spherical Torus Experiment at Princeton Plasma Physics Laboratory. As shown in Figure 15, this collaboration began increasing in 2004 as Sandia began its planning for refurbishing its Z pulsed power facility and was exploring the concept of an inertial confinement fusion engine concept for a scale nuclear fusion plant.

This collaboration linked to the Z facility refurbishment dominates the joint scientific publication co-authorship network between Kurchatov and Sandia (as shown in Figure 16). This subgraph within the network has aspects that are similar to that seen in the Los Alamos–Kurchatov relationship. The coauthors participating in this activity are far more experienced and connected than the

Figure 16. Sandia-Kurchatov Joint Scientific Publication Coauthorship Network (1997-2008)



individuals who had participated in the bulk of the collaborations in the post Cold War period. In total, there are three giant components in the larger

network at distinctly different scales (i.e. self contained structures containing small, medium and large numbers of nodes respectively). In addition to these components, there are seven distinct communities within this network as well.

Like in the Los Alamos case, visual inspection shows that nodes in this network tend to be linked via assortative mixing – nodes with similar connectivities or degrees tend to link to each other. Inspection of the degree distribution shows a large number of very well connected nodes and relatively few shallowly connected nodes (i.e. a power law is not in effect in these collaborative interfaces). The most well connected Sandia and Kurchatov scientists in this network are associated with ongoing magnetic confinement fusion experiments over time: Igor Semenov for Kurchatov and William Wampler for Sandia. Among others -- David R. Smith of Princeton Plasma Physics Laboratory, M.J. Schaffer and L.L. Lao from General Atomics, and Martin Rensink of Lawrence Livermore – all play unique bridging roles in the network and connecting these disparate research activities together.

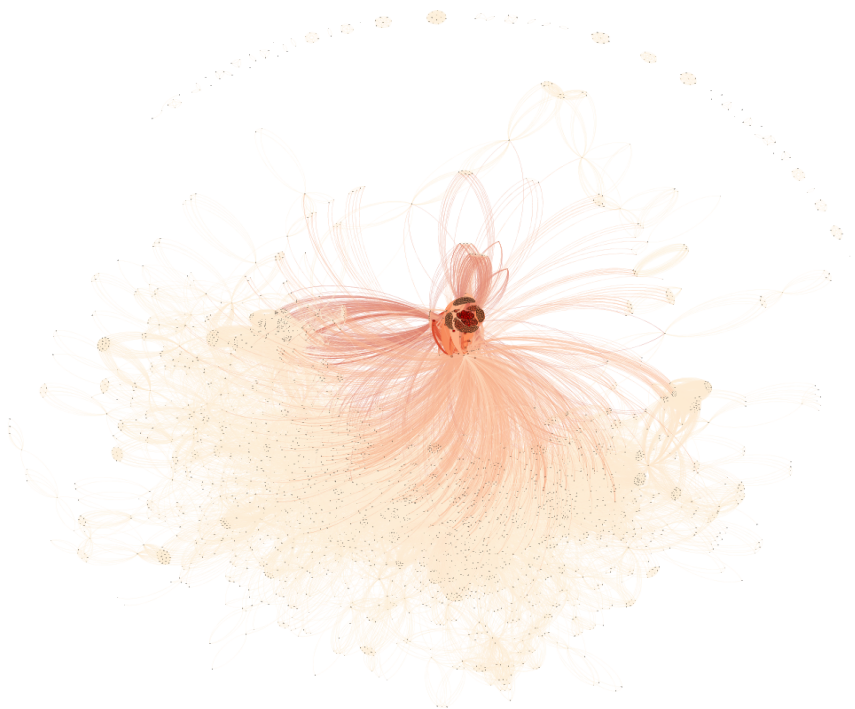
Dynamically, the subgraphs containing the collaborating Sandia and Kurchatov scientists demonstrate a preferential attachment mechanism. There is a steep connectivity and experience gradient drawing Kurchatov Institute scientists into collaborations with Sandia researchers outside of this ZR refurbishment activity. These Kurchatov scientists only had an average of 81 unique co-author linkages when beginning their collaborations. By comparison, their Sandia colleagues possessed 154 such observed links. On average, these same Sandia researchers possessed over three times the publications of their Kurchatov colleagues at the initiation of their collaborations.

All-Russia Institute for Experimental Physics (VNIIEF)

The Soviet Union established Design Department N 11 (KB-11) in April 1946 to carry out nuclear weapons development activities in Sarov. Three years later, the first Soviet atomic bomb RDS-1 was assembled by KB-11 in Sarov in August 1949. From that point forward, VNIIEF – referred to as Arzamas-16 in the United States -- played a key role in the design and manufacture of Soviet thermonuclear weapons and delivery systems, including the development of multiple independent re-entry vehicles (MIRVs) for Russian intercontinental ballistic missiles (ICBMs).

Static network analysis of VNIIEF's research community in 2012 shows a network neighborhood composed of 5202 researchers linked by 3,537,164 co-authorship relationships into 31 network components that can be segregated into

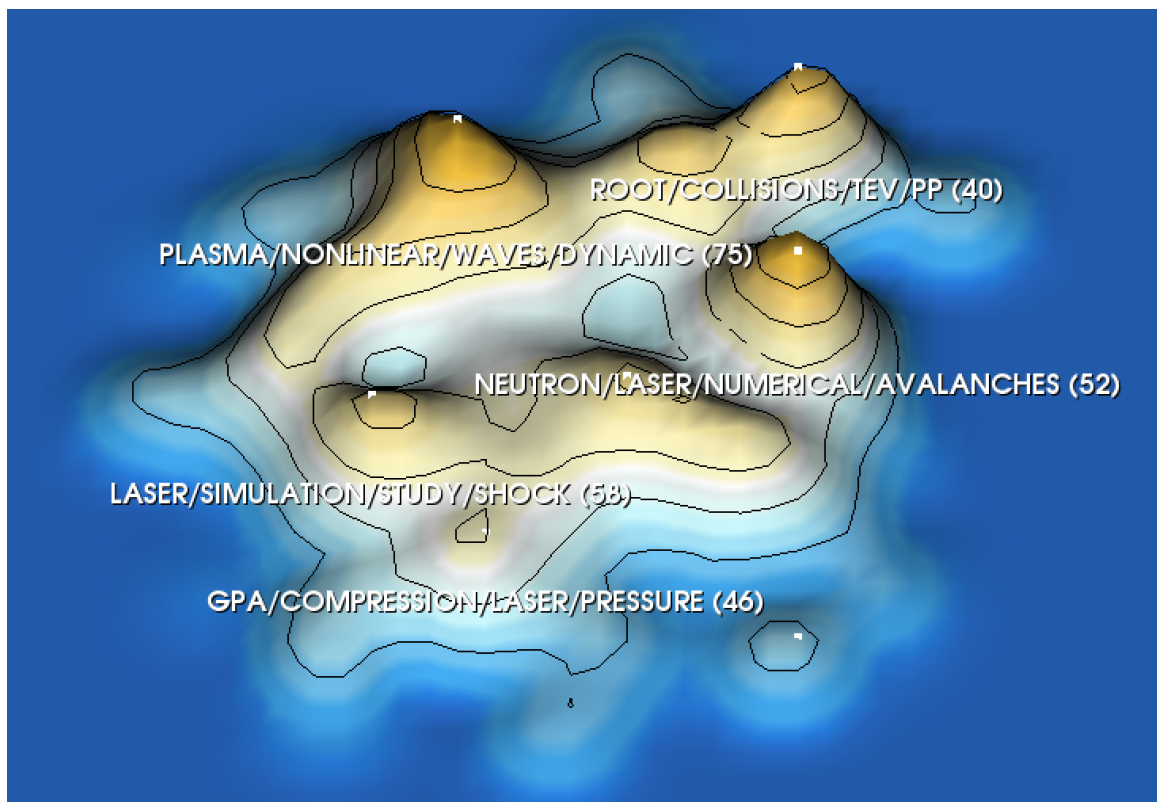
Figure 17. VNIIEF Network Neighborhood, 2012



52 distinct communities. The open research community at VNIIEF is roughly a quarter the size of the Kurchatov research community, but is more densely linked. There is a dominant, highly connected center with relatively few isolated communities (as can be seen in the unconnected subgraphs at the top of Figure 17). Compared with the Kurchatov Institute, VNIIEF appears to have fewer structural holes and as such, is less likely to possess multiple, redundant competencies in isolated areas.

From the latent semantic topology represented in Figure 18 below, VNIIEF performs a significant amount of research linked to direct drive inertial

Figure 18. VNIIEF Latent Semantic Topology, 2012



confinement fusion (i.e. using lasers to compress hydrogen pellets). These research interests can be seen in the “GPA/Compression/Laser” and

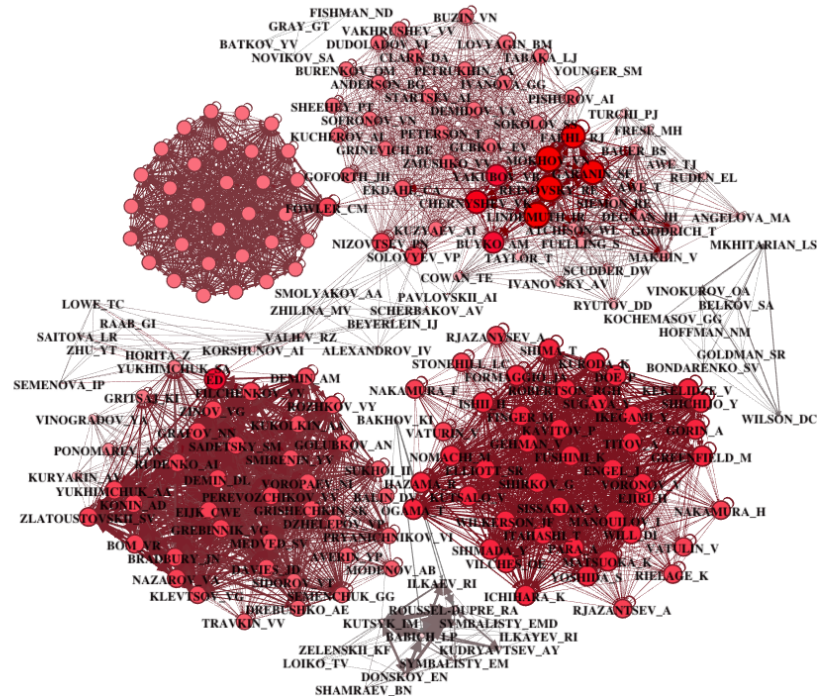
“Neutron/Laser/Avalanches” labeled research peaks. Like the Kurchatov Institute, VNIIEF perform some fundamental particle physics research at the LHC (as indicated by the “Root/Collisions/TEV/PP” peak). The key difference between the two is that but this research area appears to be a less peripheral part of the basic research portfolio at VNIIEF than it was within the Kurchatov Institute.

Based on the publication record, Los Alamos and VNIIEF began collaborating on magnetic flux compression topics in the late 1990s -- considerably later than Los Alamos’ interactions with Kurchatov. High magnetic field research still is a dominant feature of the collaborations between the two facilities. Over the past five years, the topics of joint interest between researchers at the two sites have diversified to include neutrino detection.

This collaboration concerning how materials behave in mega-gauss high magnetic fields dominates the joint scientific publication co-authorship network between VNIIEF and Los Alamos (see Figure 19 below). This network has aspects that are similar to that seen in both the Los Alamos and Sandia relationships with the Kurchatov Institute. There are five giant components in the network and eight distinct communities within this network. Visual inspection suggests that similarly connected scientists cluster together, an assortative mixing feature.

The Los Alamos and VNIIEF scientists with the highest degree are Vladimir N. Mokhov for VNIIEF and Robert Reinovsky for Los Alamos. Both of these individuals remain key players in the high magnetic field research arena. While there are not any external individuals serving as “bridges” in the

Figure 19. Los Alamos-VNIIEF Joint Scientific Coauthorship Network (1977-2008)

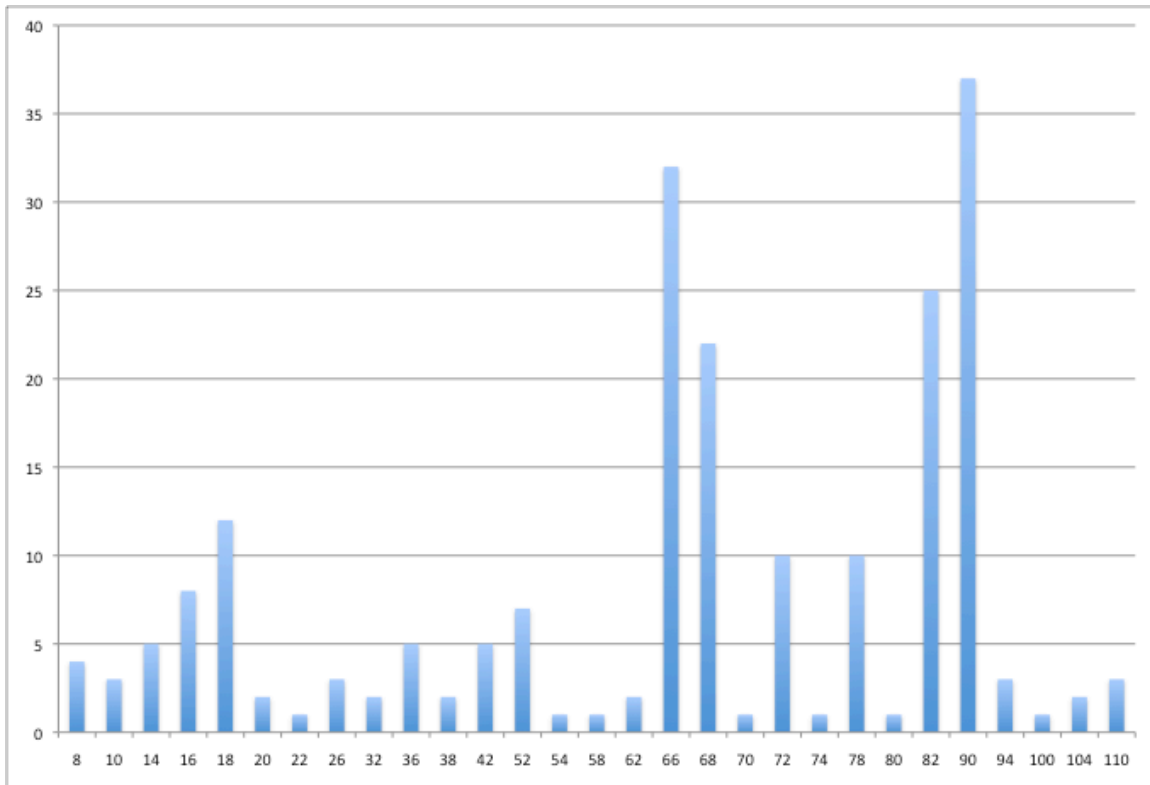


network, it is clear that Clarence Fowler at Los Alamos – a Los Alamos physicist noted as the father of magnetic flux compression generation – served a crucial role in connecting separate high magnetic field research programs over time. Inspection of the degree distribution contained in Figure 20 shows two peaks of well-connected nodes, but relatively few shallowly connected nodes (i.e. a power law is not in effect).

The subgraphs containing the collaborating Los Alamos and VNIIEF scientists appear to evolve via a preferential attachment mechanism. Like in earlier cases where preferential attachment can be observed, in the absence of political obstacles, Russian scientists from VNIIEF should be drawn to their Los Alamos colleagues as collaboration partners. VNIIEF scientists had an average of only 48 unique co-author linkages when initiating their collaborations. By comparison, their Los Alamos colleagues possessed on average of 64 such links.

The best-connected Los Alamos scientists have at least 24 more co-authorship ties than their VNIIEF colleagues, offering them a considerable advantage in

Figure 20. Los Alamos-VNIIEF Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008)



monitoring technical discoveries throughout the broader technical community.

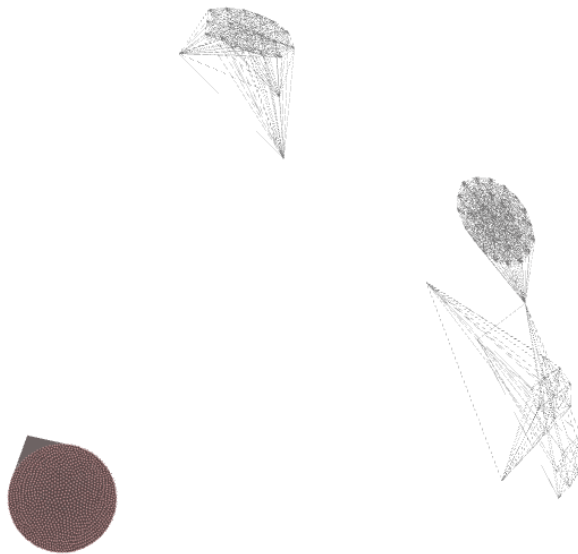
This connectivity advantage is not the only advantage that VNIIEF researchers gain by working with Los Alamos scientists. At initiation of these collaborations, it appears that Los Alamos scientists enjoy a significant tacit knowledge advantage over their VNIIEF peers. On average, Los Alamos researchers have more than double the peer-reviewed publication experience of their VNIIEF colleagues.

Lawrence Livermore began their collaborations with VNIIEF in the early 2000s. The initial focus of this collaboration was radionuclide monitoring using

biological sources. This collaboration then shifted to shock compression studies of mutual interest to the two laboratories.

Lawrence Livermore and VNIIEF scientists worked together on the ALICE experiment for the Large Hadron Collider (LHC) at CERN. This collaboration is shown in the central giant component in the joint scientific publication co-authorship network between VNIIEF and Lawrence Livermore (refer to Figure 21). The topical diversity of this collaborative relationship is reflected in the presence of three giant components in the network with five distinct communities within this network in all.

Figure 21. Lawrence Livermore -VNIIEF Joint Scientific Publication Coauthorship Network (1977-2008)



The degree distribution associated with this joint scientific publication co-authorship network indicates that scientists with similar degree appear to be clustered together, demonstrating an assortative mixing property. The Lawrence

Livermore and VNIIEF scientists with the greatest network prominence are all ALICE participants: V.V. Basmanov, D. Budnikov, V.V. Demanov, S.Filchagin, R. Ilkaev, A. Mamonov, S. Nazarenko, A.Punin, V. Punin, O. Vikhlyantsev, Y.Vinogradov for VNIIEF and both A. Enokizono and J. Klay for Lawrence Livermore. There are not any external individuals serving as significant “bridges” in the network.

The subgraphs containing the collaborating Lawrence Livermore and VNIIEF scientists appear to evolve via a selective attachment mechanism. VNIIEF scientists had an average of 685 unique co-author linkages when initiating their collaborations. By comparison, their Lawrence Livermore colleagues possessed 615 such observed links. If connectedness was all that mattered, Lawrence Livermore scientists should seek out VNIIEF colleagues for research collaborations.

Typically, a significant publication record should accompany this level of connectivity. However, in this case, there is a considerable tacit knowledge gap between the Lawrence Livermore and VNIIEF researchers that can be observed at the initiation of these collaborations. Livermore participants published an average of four times more often than their VNIIEF colleagues (45 publications to 11) at the time they entered into these collaborative interactions. Indeed, the most well published one percent of VNIIEF researchers at that initiation point is still less well published than the average Livermore researcher.

Together, these two factors suggest that a tacit VNIIEF strategy may be in place. VNIIEF researchers may be preferentially entering into large international collaboration efforts (instead of more concentrated bilateral ties with a research partner) to increase its exposure to international scientific discoveries with the

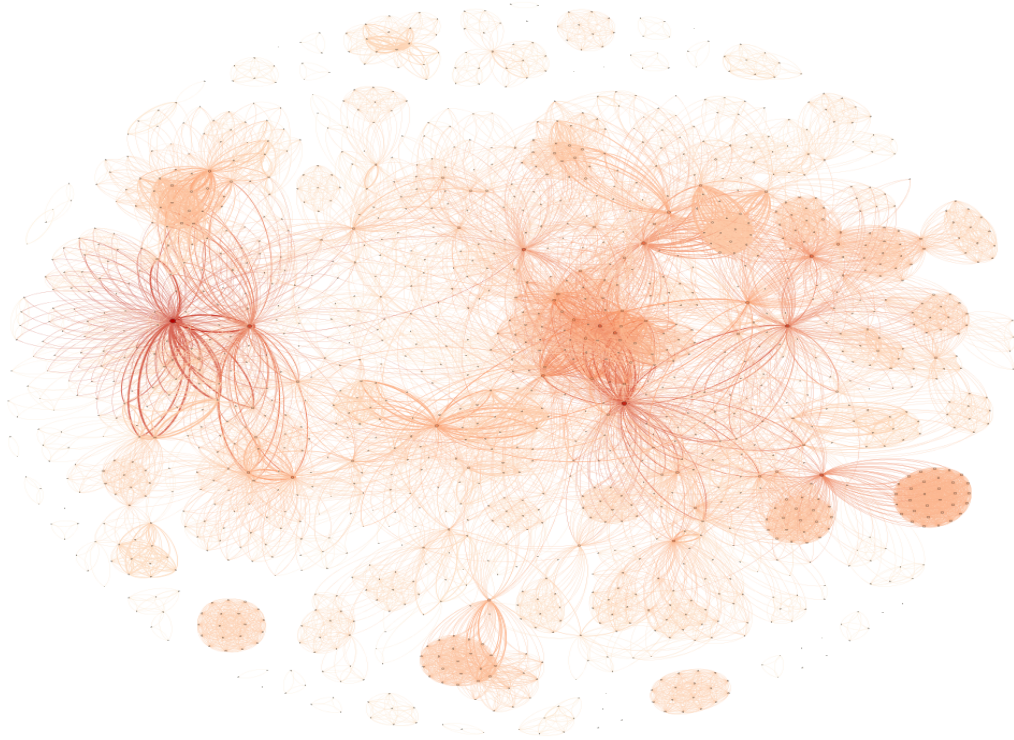
least cost. This choice to not provide more seasoned researchers to these efforts exposes less VNIIEF tacit knowledge to leakage. Unfortunately, this lack of experience suggests the relative contribution of these VNIIEF researchers to these larger efforts may be limited as well.

All-Russia Institute for Theoretical Physics

Initially known as Scientific Research Institute 1011 [NII-1011], the All-Russian Institute of Technical Physics (VNIITF) – referred to as Chelyabinsk-70 in the United States during the Cold War -- was established in Snezhinsk in April 1955 to assist VNIIEF in Soviet nuclear weapons development efforts. VNIITF's primary mission is designing thermonuclear weapons and providing scientific support to the Russian nuclear weapons stockpile throughout their lifecycle. Russian authorities did not publicly acknowledge the existence of VNIITF prior to 1992.

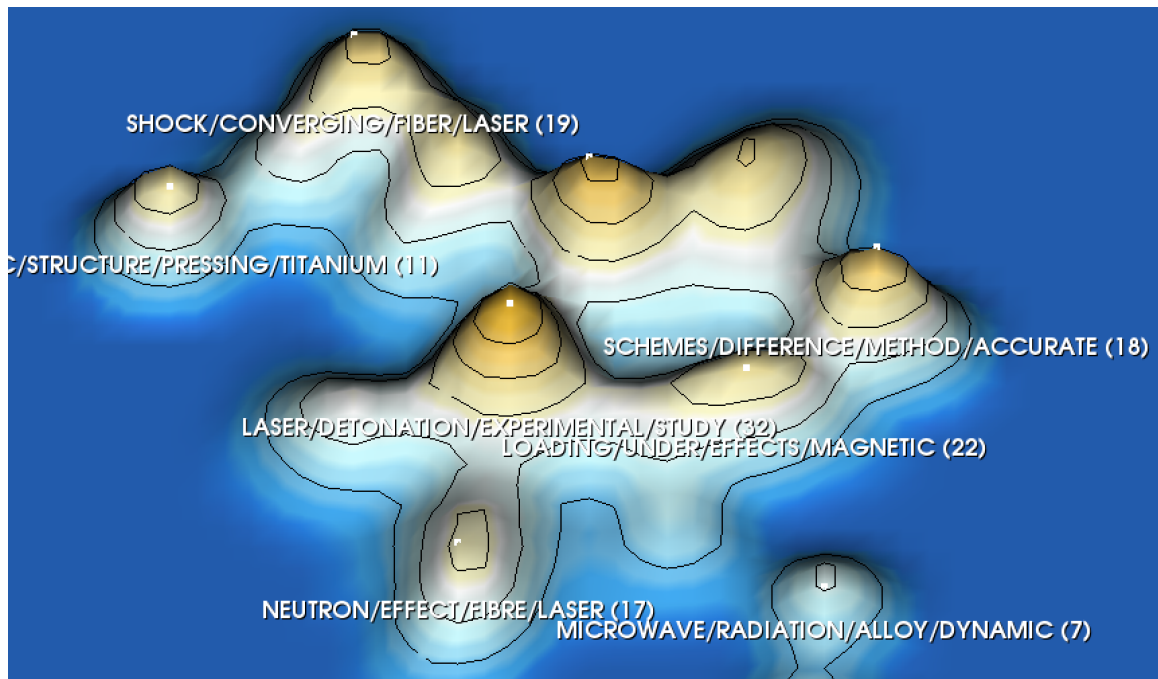
Static network analysis of VNIITF's research community in 2012 shows a network neighborhood composed of 1344 researchers linked by 17,610 co-authorship relationships into 26 network components that can be segregated into 63 distinct communities. The open research community at VNIITF is far smaller than either the Kurchatov or VNIIEF research community, but is densely linked. This community appears almost multi-polar, with no dominant, highly connected center, but with few isolated communities (as can be seen in the unconnected subgraphs on the perimeter of Figure 22).

Figure 22. VNIITF Research Community, 2012



As can be seen in Figure 23, the contemporary research conducted at VNIITF resembles much of the stockpile stewardship research performed at nuclear weapons laboratories in the United States. For example, there is nuclear safety research going on indicated by the laser initiated high explosive work “Laser/Detonation/Experimental/Study”). There are also multiple research peaks demonstrating interest in replacing expensive gas and glass lasers with less expensive fiber lasers to perform experimental work (i.e. the “Shock/Converging/Fiber/Laser” and “Neutron/Effect/Fibre/Laser” peaks).

Figure 23. Latent Semantic Topology of VNIITF Research, 2012



Unlike both the Kurchatov Institute and VNIIEF research portfolios, while there is some nuclear fusion related research (“Loading / Under / Effects / Magnetic”), there appears to be little to no work in the field of particle physics.

As the existence of VNIITF was not made public until 1992, it should come as little surprise that the first collaborations between Los Alamos and VNIITF began in the early 2000s. Plasma physics topics dominated these early collaborations. The focus of these collaborations moved to plutonium metallurgy research soon thereafter.

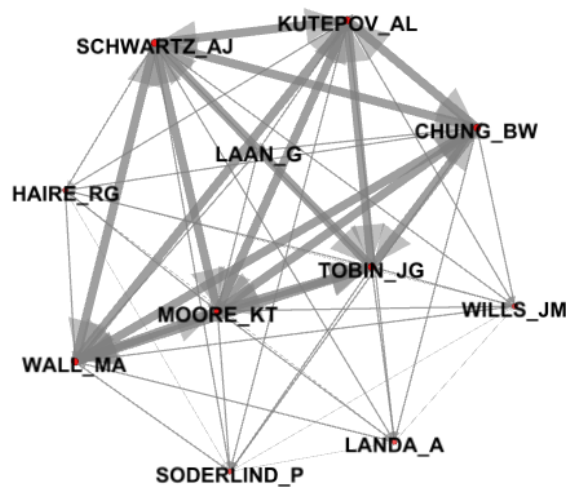
The early collaborative efforts in plasma physics between the two institutions left their mark in the joint scientific publication co-authorship network between VNIITF and Los Alamos. The scientists who participated in this work (see Figure 24 below) still compose the most connected component subgraph of research linkages. Given the low volume of collaborations and the diversity of topics, it is no surprise that there are three giant components in and

such observed links. Like in earlier cases where preferential attachment can be observed, Russian scientists from VNIITF would be expected to seek out their Los Alamos colleagues for research partnerships in the absence of government edicts to do so.

The desire to connect to the broader international technical community is not the sole factor driving VNIITF's efforts to connect with Los Alamos researchers. This tacit knowledge gradient is even steeper in this case than the Lawrence Livermore-VNIIEF interaction discussed previously. At the initiation of their interactions with VNIITF, Los Alamos researchers had authored an average of more than six times the technical publications written by their VNIITF contemporaries. Similar to the Lawrence Livermore-VNIIEF situation, VNIITF appears to be following a low tacit knowledge exposure strategy, only in this case there does not appear to be a preference evidenced for engaging only in highly international scientific collaborations.

Like the previous Los Alamos case, the primary focus of Lawrence Livermore interactions with VNIITF is plutonium metallurgy. This collaboration has the lowest participation and is the shortest of the nine relationships examined in this chapter. It only contains one small component and community (note Figure 25), making it an ideal case to use to observe micro-scale phenomena leading to network evolution. The scientists with the greatest network prominence are J.G. Tobin, P. Soderlind, A. Landa, K.T. Moore, A.J. Schwartz, B.W. Chung and M.A. Wall for Lawrence Livermore and A.L. Kutepov for VNIITF. Inspection of the associated degree distribution shows increasing frequency with greater connectivity.

Figure 25. Lawrence Livermore-VNIITF Joint Scientific Publication Coauthorship Network (1977-2008)



The subgraphs containing the collaborating Lawrence Livermore and VNIITF scientists should evolve via a preferential attachment mechanism. VNIITF scientists had an average of only 23 unique co-author linkages when initiating their collaborations. By comparison, their Lawrence Livermore colleagues possessed 58 such observed links. Like in earlier cases where preferential attachment can be observed, in the absence of political obstacles, Russian scientists from VNIITF should have natural affinity for their Lawrence Livermore colleagues.

As in the Los Alamos-VNIITF relationship, the tacit knowledge gradient between Lawrence Livermore and VNIITF researchers is steep. At the initiation of their interactions with VNIITF, Lawrence Livermore researchers had authored an average of more than three times the technical publications written by their VNIITF contemporaries. Similar to the Lawrence Livermore-VNIITF and Los Alamos-VNIITF cases, VNIITF appears to be following a low tacit knowledge

exposure strategy, only in this case there does not appear to be a preference evidenced for engaging only in highly internationalized scientific collaborations.

D.V. Efremov Institute

The D.V. Efremov Scientific Research Institute of Electrophysical Apparatus (DVEI) was spun off from a former Electrosila manufacturing plant dedicated to the Soviet Union's attempt to build its first cyclotron in the 1930s. Begun as a joint German-Russian stock company in 1898, Electrosila was known for its development of the large-scale power generation equipment associated with hydropower. When the Soviet nuclear weapons project started, the facility, which had designed the vacuum chambers and other components of the cyclotron, was designated as a "Special Design Bureau" to develop electromagnetic transducers for the project.

Static network analysis of DVEI's research community in 2012 shows a network neighborhood composed of 3309 researchers linked by 445,385 co-authorship relationships into 19 network components that can be segregated into 35 distinct communities. The open research community at DVEI is larger than VNIITF, but still far smaller than either the Kurchatov or VNIIEF research community. This community is bi-polar and densely connected (as can be seen in Figure 26).

Today, DVEI is the primary designer of Russian equipment for conducting fundamental research in nuclear physics, high-energy physics, and controlled nuclear fusion. As can be seen in the latent semantic research topology map in Figure 27, DVEI's openly published work includes a large amount of research and development work supporting the development of the ITER tokamak facility

in Cadarache, France. This work includes development work associated with ITER's divertor ("ITER/Divertor/Effect/Dose"), system of superconducting magnets ("ITER/Field/Tokamak/Ferromagnetic), cooling system ("ITER/Cooling/Reactor/Barrier") and vacuum system ("Design/ITER/Vacuum/Procurement"). DVEI also continues its work with cyclotrons.

Figure 26. DVEI Research Community, 2012

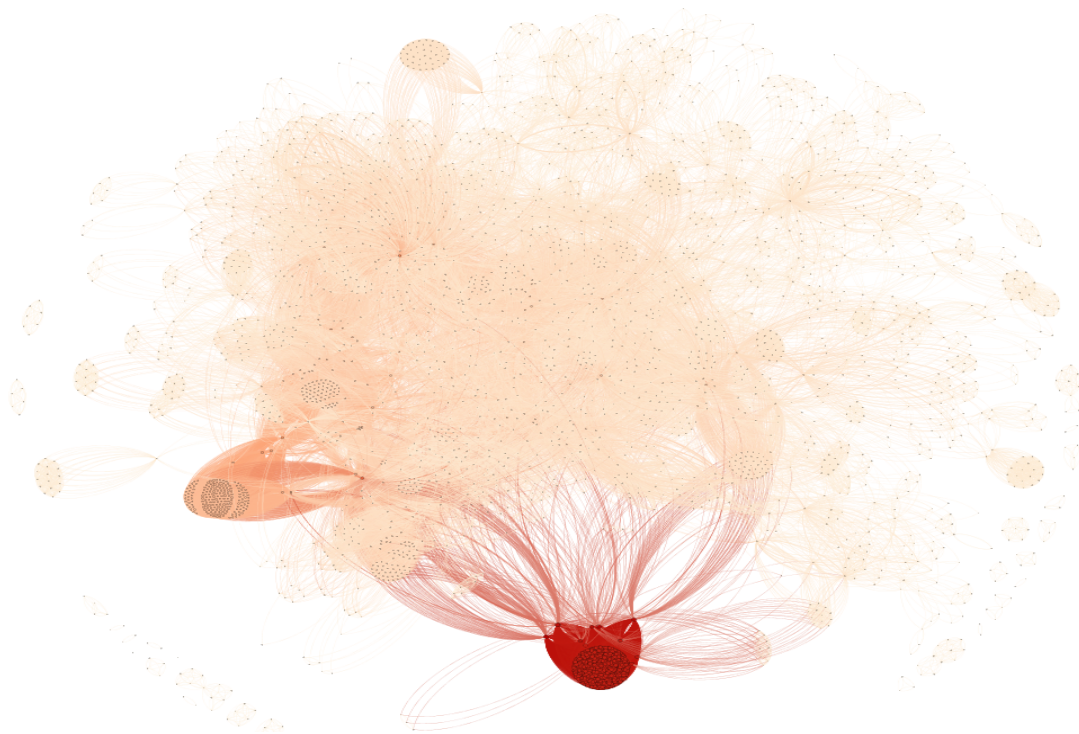
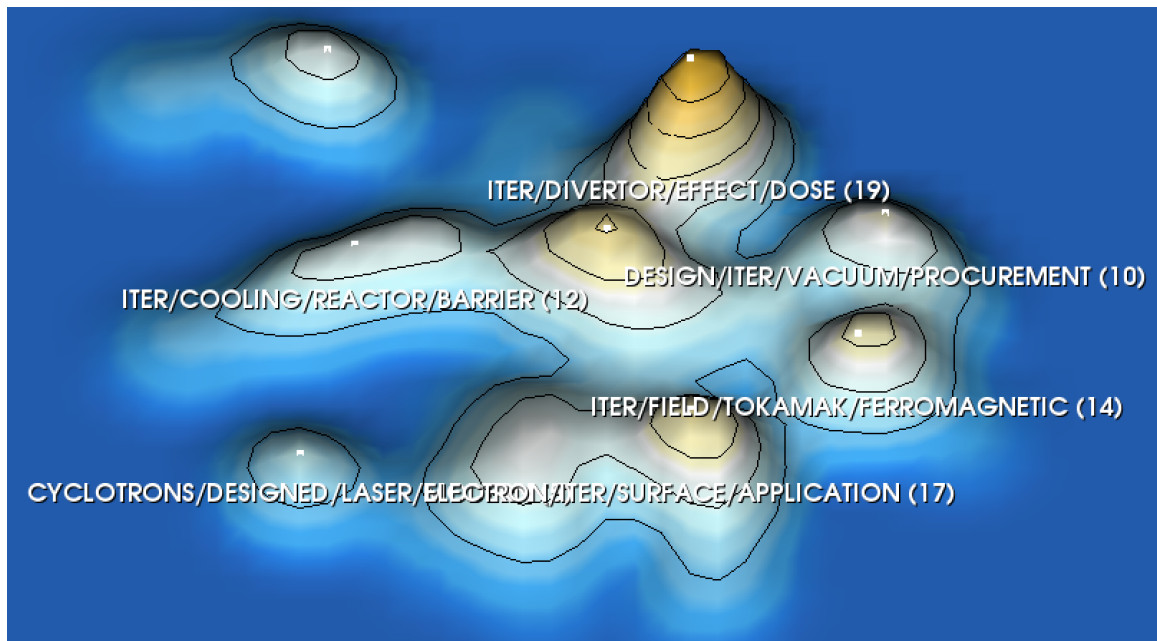


Figure 27. Latent Semantic Topology of DVEI Research, 2008-2012

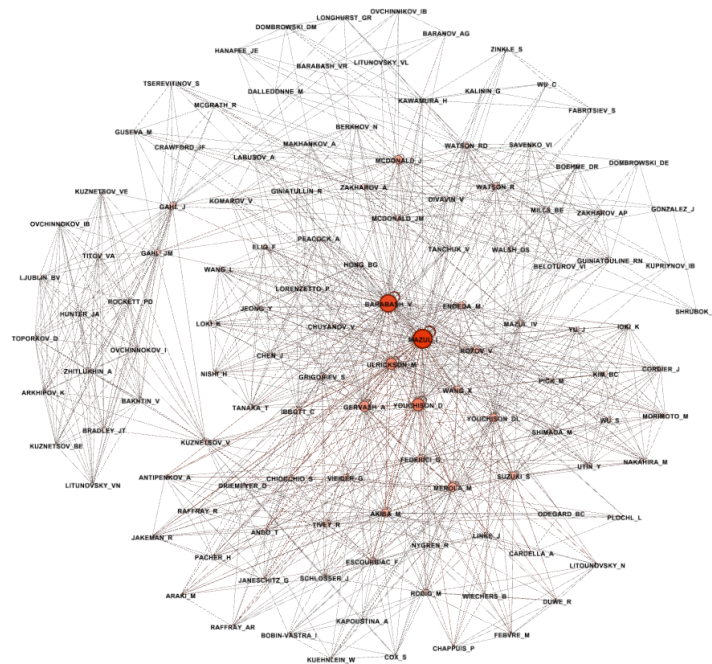


Sandia and DVEI began their collaborations in the early 1990s with a series of explorations the plasma physics of tokamak disruptions. After a short interruption in the late 1990s, this collaboration resumed as Sandia and the Efremov Institute began collaborating on the multi-national ITER project's plasma facing components. As might be expected, this work dominates the joint scientific publication co-authorship network between Sandia and the Efremov Institute (see Figure 28 below).

In terms of its evolution, this network appears to be just one step removed from the simple graph seen in the Lawrence Livermore-VNIITF relationship. There is one giant component in the network and ten distinct communities within this network. The Sandia and Efremov Institute scientists with the greatest prominence in this network are Dennis Youchison for Sandia and Igor Mazul for the Efremov Institute. J.M. McDonald at Sandia appears to have

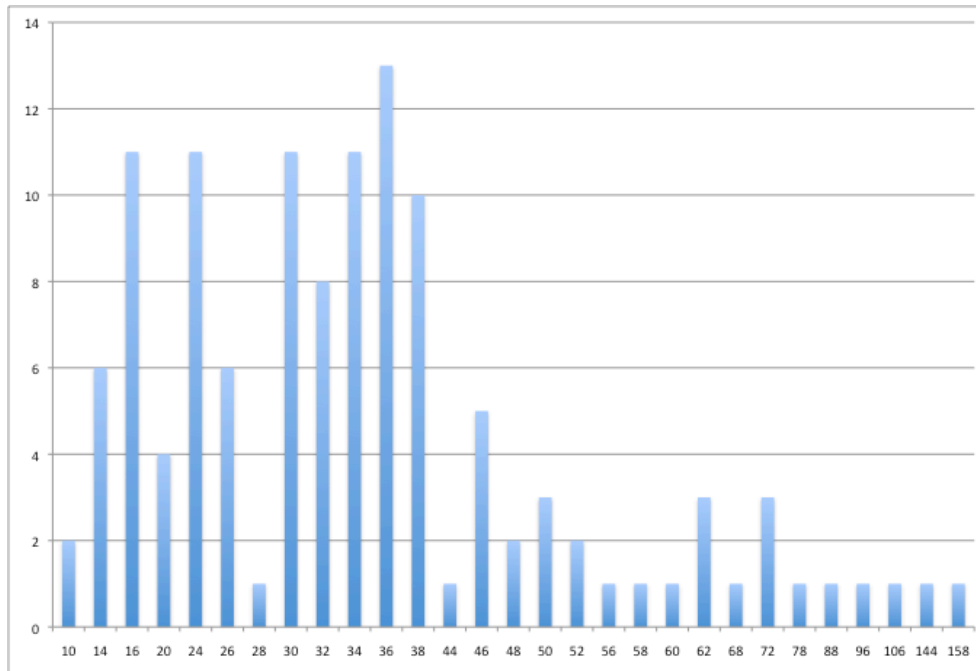
played a critical role in linking the earlier disruption focused activity to the later ITER focused activity.

Figure 28. Sandia-Efremov Institute Joint Scientific Publication Coauthorship Network (1977-2008)



Inspection of the degree distribution contained in Figure 29 shows that this distribution appears to follow the power law decay that is typically expected in scientific co-authorship network. This structure yields a network with a small diameter (i.e. the maximum shortest path length between any two pairs of nodes in the graph). Thus, this network demonstrates *small world* properties.

Figure 29. Sandia-Efremov Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008)



The subgraphs containing the collaborating Sandia and the Efremov Institute scientists appear to evolve via a preferential attachment mechanism. Efremov scientists had an average of only 37 unique co-author linkages when initiating their collaborations. By comparison, their Sandia colleagues possessed 51 such observed links. Like in earlier cases where preferential attachment can be observed, Russian scientists from Efremov should have natural affinity for their Sandia colleagues.

Unlike the past three relationships explored in this section, there is not much of a tacit knowledge gradient in this case. Sandia researchers still enter into these interactions with stronger publication records – an average of four publications per author – than their Efremov colleagues. While Sandia researchers have marginally more experience in publishing technical articles, this difference largely disappears at one standard deviation from the mean. As such,

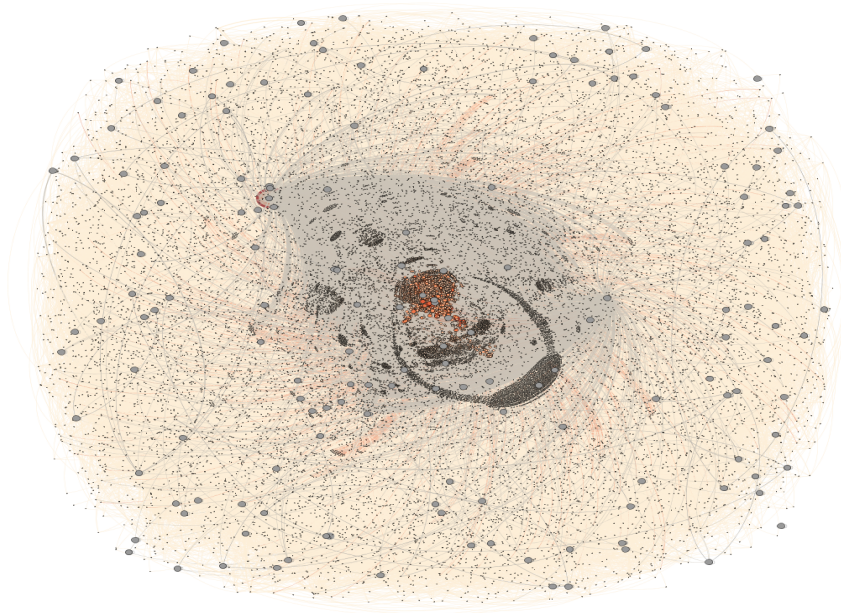
the most experienced researchers at each institute enter into these international collaborations at roughly the same level of publication experience.

P.N. Lebedev Physical Institute

The Russian Academy of Sciences' P.N. Lebedev Physical Institute (PLPI) in Moscow is one of the largest and oldest Russian scientific research centers. Founded in 1934 by S.I. Vavilov, Lebedev Institute scientists are noted for discoveries such as the Vavilov–Cherenkov effect, the phase-stability principle as well as the scientific basis for controlled thermonuclear fusion. PLPI scientists Andrei Sakharov and Igor Tamm are credited with designing the first Soviet thermonuclear weapons.

Static network analysis of PLPI's research community in 2012 shows a network neighborhood composed of 25,329 researchers linked by 18,062,149 co-

Figure 30. PLPI Research Community, 2012

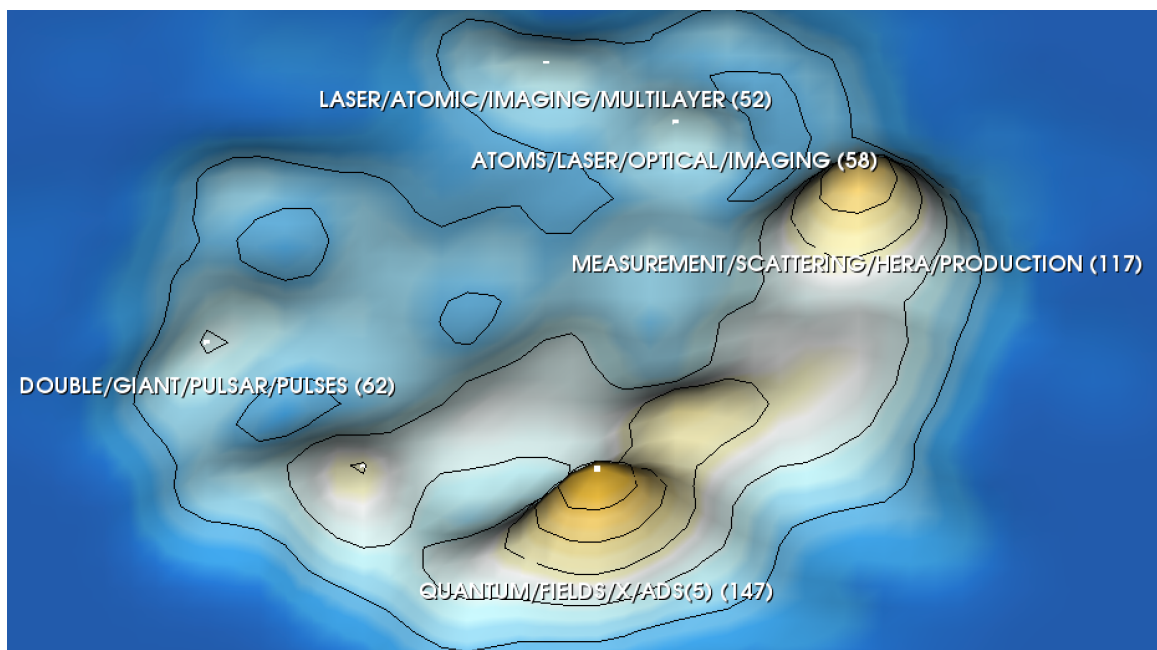


authorship relationships into 100 network components that can be segregated into 177 distinct communities. The open research community at PLPI is larger

than any other Russian research institute considered in this study. This community is multi-polar and densely connected (as can be seen in Figure 30).

Today, PLPI is known for its work in astrophysics of black holes and pulsars (observed in the “Double/Giant/Pulsar/Pulses” research peak in Figure 31 below) and particle physics (“Measurement/Scattering/HERA/Production”), including research into dark matter and string theory. PLPI has developed a strong reputation for its experimental work in plasma diagnostics (“Atoms/Lasers/Optical/Imaging” and “Laser/Atomic/Imaging/Multilayer”). This competency makes PLPI a sought after partner in the conduct of nuclear fusion relevant research.

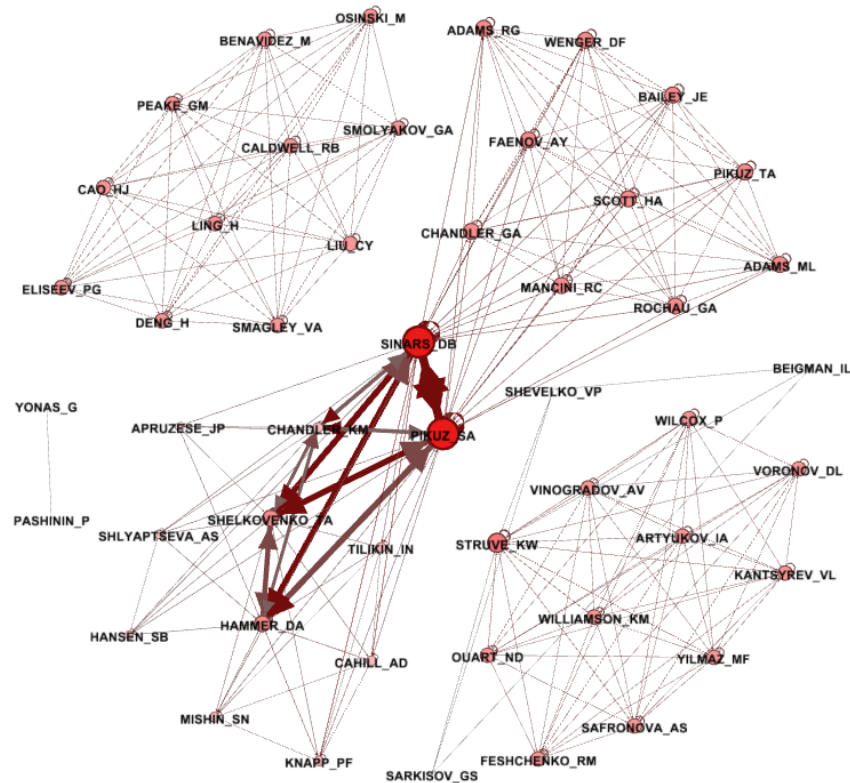
Figure 31. Latent Semantic Topology of Research at PLPI, 2008-2012



Sandia’s interactions with the Lebedev Institute date back to the mid-1970s and the period of détente between the United States and the Soviet Union in the midst of the Cold War. The focus of this initial collaboration centered on inertial confinement fusion topics – a topic that was closely held at the time. As

the political situation collapsed, this scientific interaction ceased as well. The relationship was reinvigorated thirty years later with different figures, but focused on the same technology set.

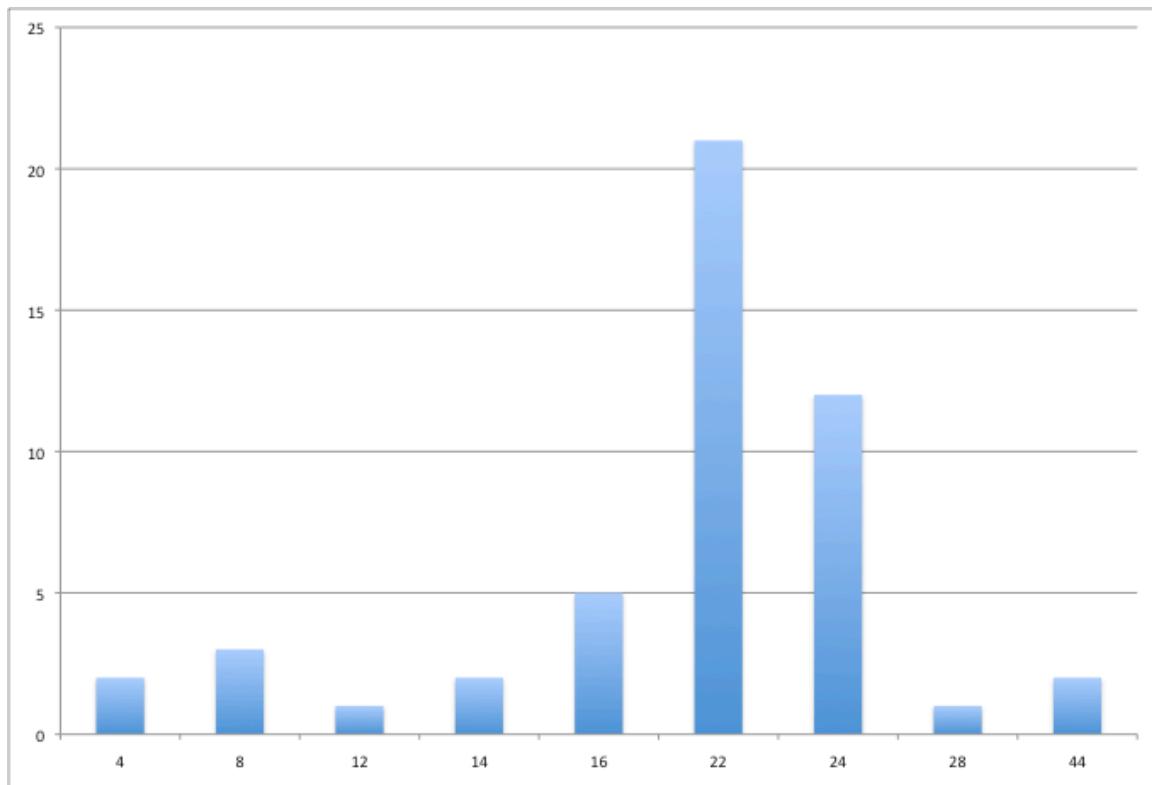
Figure 32. Sandia-Lebedev Joint Scientific Publication Coauthorship Network (1977-2008)



This newer collaboration is the center of the joint scientific publication coauthorship network between the Lebedev Institute and Sandia (see Figure 32 above). There are four giant components in the network and fifteen distinct communities within this network. Daniel Sinars of Sandia and Sergey A. Pikuz of the Lebedev Institute are the central figures in the joint scientific collaboration network between the two institutions. Individuals from external institutions do not appear to play much of role in this coauthorship network. Like many of the other degree distributions for these coauthorship networks, the Sandia-Lebedev

Institute network degree distribution is single peaked (Figure 33) and does not

Figure 33. Sandia-Lebedev Joint Scientific Publication Coauthorship Network Degree Distribution (1977-2008)



seem to follow the expected power law distribution.

The subgraphs containing the collaborating Sandia and the P.N. Lebedev Institute scientists show a strong preferential attachment mechanism. P.N. Lebedev Institute scientists had an average of only 41 unique co-author linkages when initiating their collaborations. By comparison, their Sandia colleagues possessed 204 such observed links. Like in earlier cases where preferential attachment can be observed, P.N. Lebedev Institute should seek out their Sandia colleagues for research collaborations in the absence of government interference.

This relationship exhibits a tacit knowledge gradient in the opposite direction of all the other dyads explored in this chapter. Lebedev researchers average double the publication experience of their Sandia colleagues at the

initiation of these collaborations. Thus, given the knowledge model proposed in Chapter Two, this relationship is the collaboration most likely to generate a productivity enhancement for the U.S. national laboratory partner in the relationship.

Educated Expectations

The knowledge economics perspective and this set of network case studies suggest the presence of more granular regularities that should be empirically observed in this alternative network study of international research collaborations than the fairly broad hypothesis regarding the productivity of internal and external collaborations in given contexts at research institutes as proposed at the end of Chapter 2. In addition to the atomistic and bilateral collaboration contexts described earlier, this set of network case studies display the importance of large-scale, multi-national scientific research activities in the collaboration activities that partner Russian research institutes and US national laboratories.

Within the context of the model proposed in the preceding chapter, the individual researchers who participate in these large-scale, multi-national scientific research activities face reduced costs of maintaining large numbers of research partners because these costs are partially borne by an external organization. This reduced cost allows for the development of massive scale collaborations that generate significant improvements in individual researcher productivity. Unfortunately, the gains of tacit knowledge in these large-scale, multi-national scientific research activities should be local and limited as well. Adding additional Russian or US coauthors to individual researchers' network

neighborhoods in these environments should have a negligible productivity impact as a result.

These network case studies also reveal some secondary patterns yielding specific expectations for who should benefit most from these collaborations. First, given the specification of the process of knowledge acquisition in Chapter 2, it should be expected that the least well-connected and least experienced researchers should benefit the most from engaging in research collaborations with better-connected and more experienced colleagues. As such, the greatest gains in research productivity should be seen in institutions that participate in bilateral collaborations where on-average low-connected researchers with low experience participate in relationships where there are large differences in average degree and average publication experience at collaboration initiation.

IV. Collaborating to Ruin? Measured Productivity Benefits of US-Russian National Laboratory Collaborations

Putting Expectations to the Test

The three preceding chapters developed the perspective that atomistic and large-scale multi-national research collaborations involving researchers at Russian research institutes and US national laboratories take place in environments where researchers are enabled to make optimizing decisions concerning their network neighborhoods. As such, individual researchers in these contexts can adjust the composition of their network neighborhoods amongst coherent types of researchers in keeping with the relative gains and maintenance costs associated with maintaining those linkages. Thus, if this perspective is accurate, the productivity gain from adding an additional researcher of a generic type should be positive or even close to zero – if the researcher has been able to fully optimize their network. This marginal productivity associated with an additional researcher from a given researcher type should be seen when all researcher level collaborations are considered in

aggregate as well. In fact, these small and positive or near zero marginal productivity gains should be symmetric (i.e. observed in both collaborating institutions settings).

“Bilateral” collaborations, however, distort this neighborhood type selection process by introducing collaboration partners by outside direction instead of research need. Thus, it is likely that researchers at one of the two institutions participating in an international research partnership between institutions possess a greater number of research partners than would otherwise be optimal. As such, these researchers should experience negative returns associated with additions of any researchers of this type to their network neighborhoods. In aggregate, the institutions of these researchers who possess these suboptimal neighborhood mixes should demonstrate negative returns associated with adding additional coauthors from the partner institution.

As such, the policy context likely dictates who benefits from these collaborations. Given the policy objectives behind these collaborative activities associated with the research relationships between Russian research institutes and US national laboratories, it is expected that US national laboratories will experience negative productivity impacts associated with these relationships. Conversely, Russian research institutes should experience positive productivity impacts.

Prior to diving into the empirical analysis, the proceeding sections describe the data collection, processing steps taken and challenges encountered in developing the dataset used in this study in some detail. In general, it is useful to understand the inherent limitations of using international publication data in this manner. However, this detail is included primarily to aid research policy

analysts in using this and similar scientometric techniques in assessing international collaborations.

Constructing Institutional Publication Record Sets

The empirical analysis of this study begins with the attempt to construct an appropriate corpus of scientific publications associated with the D.V. Efremov Institute, the Kurchatov Institute, Los Alamos, Lawrence Livermore, the P.N. Lebedev Physical Institute, Sandia, VNIIEF and VNIITF. The corpus of scientific publication records in this study are derived from publication database records contained in Thomson ISI Web of Science covering the period 1977-2012. These extracted records were processed via custom Perl scripting (included in Appendix I) to form flat files in ASCII text format in which each line of text represents a single publication record. These flat files were then sorted and filtered for duplicate records.

The simplicity of this process masks the underlying challenges associated with generating an appropriate document record set. Inadequate mapping of research (and researchers) to institutes is a common issue that bibliometric techniques of research performance assessment must overcome (Abramo, D'Angelo and Caprasecca, 2009). This issue emerges from a lack of consistency between professional journals in how institutional affiliations are tracked, if at all. As a result, relying on institutional affiliation alone to create an accurate corpus of publications for a given institution will consistently underestimate the number of publications produced by that institution. The practical result of

ignoring this issue is that any analytic estimates derived from econometric study will be less accurate.⁶

To avoid this systematic bias, the set of documents drawn from data extractions using only institutional affiliations was augmented with the addition of records derived from authors with that known institutional affiliation. When a definitive author-institution linkage was observed, that author's name is used to add additional documents to a given institution's corpus of publications.⁷ Establishing that link from examination of publication database records alone requires either direct citation of the individual's institutional affiliation in the database record or unique identification of the individual as part of a single laboratory team producing a journal article. Unfortunately, many early database records do not contain enough information for them to be useful for this kind of validation.

As the analyte of interest to this study is the publication productivity of pairs of researchers at specific national laboratories and research institutes, these additions to the corpus also must be filtered for researcher movement. While important for empirical measurement with precision, this filtration did not omit many publications from inclusion. National laboratories typically do not experience the same amount of researcher transition observed at many

⁶ Institutional attribution has improved over time within large collections of past scientific publications. One of the key factors improving this attribution capability is the inclusion of greater numbers of past conference publications in available collections of scientific publication records such as Thomson ISI Web of Science. Similarly, the move to make unclassified governmental research more available to a broader audience via entities like DOE's Office of Science and Technology Information (OSTI) strengthens the ability of research policy analysts to understand institutional attribution as well.

⁷ It should be noted that this process is unable to control for errors generated by the practice of some Soviet authors in the pre-1991 time frame (primarily from 1985-1991) to only cite the institution of one of the authors. This practice leads to the inaccurate assignment of researchers to institutions to which they do not belong.

universities. In the United States, this low transition can be partially attributed to the pay differential that exists between the contractor operated national laboratories and many state-run universities. Outside the United States, the mobility of significant researchers at national laboratories is often circumscribed by governmental fiat because of security concerns. These same researchers at these facilities also often do not have an incentive to move because they often receive prestige and special pecuniary incentives their colleagues are not granted.

The lack of a single naming convention across journals presented an even greater challenge for making accurate additions to the body of documents under analysis (Newman, 2001). This deviation created a distinct possibility of over-counting due to inaccurate researcher attribution in either a single initial and surname regime (which is more common in older records) and a two initial and surname regime (which is currently the standard). In the single initial and surname author naming convention, there is a significantly higher likelihood of double counting when compared with the two initial naming standard. This likelihood is increased when the publications are drawn from countries that do not follow European naming conventions and thus, common names are often mistakenly specified as family names.

To control for this issue, attention was given to the task of technical field attribution to discern researchers with the same name working in different technical fields. The rationale behind this activity is the likelihood that two researchers in a given narrow technical field share the same initials and surname

is small enough to be negligible in most cases.⁸ This control method only fails when two individuals with the same initials both participate in research applicable to the same specialized technical field. While rare, these failures did occur, because research interests occasionally are shared across pairs of scientists who have familial ties and share common initials.

At this point, individuals were discerned from one another using a technique derived from network scientific observations concerning subgraph stability. While authors' linkages to other researchers evolve with time, in the same time period, individuals will display stability in their research relationships. This stability comes from the fact that authors do not attract coauthors by chance, but acquire coauthors that are known to their current set of coauthors (i.e. *triadic closure*). As such, an author's collaboration pairs are useful for understanding which publications from a given name are appropriate to add to the corpus of institutional publications (Wasserman and Faust, 1994).

There are only two observed factors that occasionally cause the extraction of records by institutional affiliation to attribute significantly more publications to a given institution than it actually produced. Some special edition journal publications report the special edition's editors and their institutions as having coauthored the articles within the edition. Likewise, some researchers whose travel is funded to go to international conferences by the sponsors of a given conference are known to cite that sponsor as contributing to a conference publication that it did not participate in - outside of providing the means for the

⁸ Researchers from countries where the order of the surname and common name is reversed from the European naming convention often have their names inaccurately recorded in the publication databases. As such, this method is considerably less useful in examining relationships involving researchers from China than it is for considering relationships with Russian individuals.

paper to be presented. In the document corpus that is the focus of this document, these records tend to be from Russian authors who cite the Institute of Electronics and Electrical Engineers (IEEE) as a coauthoring institution. If not controlled for, this practice can overstate collaboratively produced publications by an order of magnitude in some post-1991 era years.

As can be seen in the above discussion of concerns about bias in corpus construction, one of the primary complicating factors in this process is that the assignment of a given researcher to a given institution is difficult to automate. To summarize, a publication to author mapping exists in these database records. Likewise, an institution to publication mapping also exists in newer records. However, there are no mappings of authors to institutions in these records for older publications involving researchers from multiple institutions for the entire time span of interest to this inquiry. This factor creates a situation in which direct inspection of the publication itself is often necessary. It also dramatically increases the time necessary to conduct a thorough inquiry.

Organization and Data Processing

To enable empirical analysis, these institutional record sets must be converted into data organized into panels representing co-author dyads. Two panels are associated with each co-author dyad. This representation is consistent with the idea that a link between coauthors actually consists of two directed ties (i.e. one in which the first author of the pair affects the second author in the pair and another in which the second author in the pair influences the first author). While the direction of influence changes in these co-authorship dyad panels, each panel possesses the same productivity values.

Due to the large size of the co-authorship networks between these eight institutions, this conversion primarily must take place in an automated fashion. Generation of the entire set of co-authorship pairs in these networks is a breadth-first search process and is both time and compute time intensive. For this study, the conversion of these institutional record sets to panel data was distributed across multiple Linux compute nodes to shrink processing time. These results must be filtered to control for trivial dyads (i.e. author *loops*). The conversion script used in this study is included in Appendix 1.

Operationalization of Variables

Exploring international collaboration productivity impacts in this context required operationalization of each of the variables proposed in the theoretical model contained in Chapter Two. Generating the dependent variable for this analysis simply required generating joint publication counts associated with pairs of researchers. However, the proposed model also demanded consideration of how to measure tacit knowledge - a dilemma that has stymied much empirical exploration in this area. Measuring the network neighborhood variables of interest (i.e. time length of collaborations, number of other researchers that a given researcher is linked to at any one point in time and researcher type partitions by country) simply had to avoid double-counting issues and institutional attribution issues in comparison. Institutional and research field connectivity metrics must be calculated to control for researcher prominence both within the institution to which they are affiliated and the research field to which they contribute.

Measuring knowledge directly is impossible. However, one can indirectly observe outward indicators associated with each of the types of knowledge (Foray, 2004). This study takes advantage of this attribute by using publication experience as an indicator of tacit knowledge for each author participating in a collaborating dyad, which is viewed as an unobservable variable within the context of this empirical analysis (Wooldridge 2010).

The number of publications a given author has published is a relevant measure of tacit knowledge. Collective tacit knowledge of a socially constructed process, such as publishing scientific articles, increases with iterated experience. Publication experience represents the number of iterations a researcher has gone through with the publication process. As such, it reflects a given author's tacit knowledge concerning which research questions to pursue, how to perform research relevant to these questions and the limitations of any conclusions that can be drawn from the work.

This indicator has notable flaws. Publication experience also does not reflect how many failed attempts a given researcher has conducted. Likewise, it also does not measure time spent with a master as an apprentice - the conventional means by which tacit knowledge has been transferred. The publication experience variable in this study does not segment publications by journal, impact factor or tier for example. This unitary treatment of publications implies the presence of a single uniform quality standard for publication that does not exist.

While individual researchers vary in their tacit knowledge, they also vary in their capabilities to draw insight from prior explicit knowledge, perform research and draw conclusions from empirical analysis. Uncontrolled, this

individual level heterogeneity could threaten the validity of any empirical results reached by this study. Thus, this study uses average researcher citations for both dyad participants in a given year to control for this particular source of productivity variance.

As this study is an alternative network analysis, the primary unit of analysis is the co-authorship dyad. In this empirical set up, it is important to identify when a collaborative tie may exist between two co-authors. It is routinely acknowledged that collaborative research relationships, like other kinds of social linkages decay without maintenance. However, many network science oriented treatments of scientific communities act as if once a linkage is forged between two researchers that it is permanent and permanently useful. Indeed, this concept is central to the idea that scientific communities in the various disciplines are all linked such that there are no isolates (e.g., scientific disciplines constitute giant components where there is at least one cycle that connects every member of the community to every other member of the community).

While this idea of permanence appears to be at odds with the observation of researcher behavior, there is far from a settled answer concerning how long an average research collaboration lasts without an outward symbol of productivity in the open science community. This study, for instance, makes use of a “temporary” collaboration variable (*tcollab*) based on an average five-year period around the publication generated by a co-author pair to describe the time when two researchers possessed a collaborative tie. This period is based on the notion that the submission to publication phase for a typical technical publication can take anywhere from 12 to 18 months. Once the work is published, the coauthors

can expect an equivalent period where they are asked to present their findings and field reactions to their work. Given the annual basic research funding cycle, it seems reasonable to expect there is six months to a year at the outset actually performing work and a similar amount of time on the other side of the publication planning for follow-on work.

It is doubtful this constructed variable represents an authoritative statement on collaboration length in this context. Such a statement would have to come from an intensive bibliometric analysis that is outside the scope or interest of this paper to perform. However, this construction is more consistent with a conception of research collaboration that communicates information that may depreciate rapidly. This conception of the information being communicated by these networks as high-value, but of short temporal relevance correlates with the short horizon of utility for technology specific information possessed by some researchers (Foray, 2004).

When compared to how difficult it is to disambiguate author-institutional ties, the construction of the network neighborhood variables is relatively straightforward. Despite this relative ease of construction, there are two issues, which can generate measurement errors if not properly screened. During the period of this study, an important transition in naming conventions occurs and several different initial-surname strings may refer to the same author. Partitioning this network neighborhood into Russian and US country origin co-authors suffers from the same author-institution assignment problem discussed in the preceding section.

The naming problem presents a significant issue for degree counting for scientific authors. Network scientists such as Newman (2001) have attempted to

understand this problem by conducting side-by-side statistical examinations of single initial citations against dual initial citations across a set of scientific fields. These comparisons have often focused on whether the degree distributions across the set of authors retained key network properties (e.g., conformity to a power law distribution, giant component size) and have found little difference.

This paper pools both one initial and double initial author references. This decision was deemed prudent after it was observed that the double initial standard became dominant in international scientific journals in the early 1990s - the beginning of the relevant period for considering most US-Russian scientific interactions. Notably, Russian publications adhered to the superior double initial standard for authors - even in the 1980s - because their formal record keeping always included patronymics to distinguish between individuals with similar last names. Thus, any measurement error generated by pooling both author references is likely to impact only US authors who authored papers in the 1980s.

To minimize even this measurement error, the counting script used to measure the coauthor connections of a given author undertakes a number of steps to minimize double counting. A set of all coauthors is constructed, sorted and exact duplicates are eliminated. Collisions between an author's name and others within this set are also dropped.

Despite these steps, some measurement error from the naming scheme is likely to remain. Female researchers in the United States are now more likely to be cited in a three initial format than with the now common two initial format. Alphabet differences generate transliteration errors, because there are multiple letter interpretations of some Cyrillic characters (e.g., the IA and IOY ligatures). Early optical character recognition programs also commonly misinterpreted

some character strings (e.g., even today IVI usually is machine interpreted as an M if the original work was produced on a typewriter). While optical character recognition programs have improved dramatically, such errors are still typical for machine interpretation of typewritten documents.

To bypass the author-institution assignment problem, the network neighborhood partitions discriminating Russian and US co-authors from co-authors of other affiliations were constructed using Russian and US reprint authors. This practice likely undercounts the number of both Russian and US researchers in the network neighborhood of most researchers. Reprint authors tend to be researchers who have attained a position of seniority within either the Russian research institute or the US national laboratory they are affiliated with such that they may be viewed as principal investigators on research projects.

Despite the fact that these partitions under-represent the total number of either Russian or US researchers that a given researcher participating in these international collaborations has in their total network neighborhoods, the choice of reprint authors as a discriminating heuristic may better represent the actual ties that exist between individual researchers in many settings. It is often the case that research teams perform work in what can be thought of as a star network configuration. The principal investigator (typically the reprint author) is the central node of this configuration that passes directions to researchers who perform tasks to support the broader research team effort. The other researchers who participate in a research team configured in this manner will pass information and results back to the principal investigator. If the team is dispersed by technical specialty and geography, it is likely that individual

researchers working on the same teams may not actually be acquainted with one another.

The absence of typical research and development input information cited in Chapter 2 means that a proxy is required in order to understand institutional effort in open science across this set of research institutions. Oddly enough, such a proxy has been established in the research and development economic literature - simple publication counts at the institutional level (Griliches, Hall and Pakes, 1991). In this context, however, using publication counts as a proxy for institutional effort introduces a source of endogeneity into the empirical analysis.

However, developing an alternative proxy measure may seem trivial if viewed simply from the US perspective. Indeed, it is true for the US national laboratories under discussion in this paper that institutional labor force estimates can be made for each of the US national laboratories over the past thirty years. However, generating comparable Russian laboratory labor force figures requires recognition that any figure generated will be imperfect and only be available over a shorter time period than truly optimal.

The key source of imperfection is that three of the Russian research institutions often did not participate in the unmediated generation of open science prior to the dissolution of the Soviet Union. When scientists at Soviet era research institutions did publish, they published in general Soviet scientific journals where the names of the authors were published, but their institution was not. The exceptions in this consideration are the Kurchatov Institute and the P.N. Lebedev Institute, which has routinely produced items for publication under both regimes.

Even as the Soviet Union was becoming more open under *glasnost* and *perestroika* and the scientific work of Soviet era nuclear weapons laboratories began to filter out, the actual open publication output at these facilities was obscured by discrepancies in how the scientists referred to their institutions. Many scientists just used the name of the city, instead of a specific facility. When researchers cited the name of a facility, it was often in a way that was non-standard - an added source of confusion. In some cases, laboratories with multiple sites, such as Kurchatov, ended up “rebranding” some of their distant sites as their own laboratories (such as Troitsk).

The Russian institutional effort figures used in this research have been corrected as much as practical for this latter set of problems. However, significant deficits still remain in the pre-1990 era institutional effort counts because of the mediated publication process at that time. Unfortunately, the only way to improve upon these figures is to add to the pre-1990 figures by taking the set of known scientists at these institutions in the post-1990 time frame and adding their output in general Soviet journals to the constructed institutional publication frequencies. Unfortunately, such a step is not defensible because of over-counting that may occur due to a lack of granularity concerning how Russian scientists migrated from institute to institute under the previous regime and a lack of knowledge about scientists who emerged from their professional training during this period. Fortunately, because most collaborations between US national laboratories and Russian laboratories began in the post-1990 period, the inability to correct for this issue only has a limited impact on the empirical examination conducted in this study.

Controlling for researcher prominence in the empirical analysis involves constructing snapshots of the institutional research community and the prominent research fields identified in the earlier chapters (i.e. high energy particle physics and nuclear fusion). For consistency sake, these snapshots are constructed using five-year scenes (similar to the “temporary” collaboration variable discussed earlier in this section) generated from the co-authorship relationships revealed in publications involving a particular institution or research field. The adjacency tables associated with these co-authorship relationship snapshots were processed to generate eigenvector centrality values associated with each community researcher and field participant for each year under consideration.

The name of each dataset variable, its mean, minimum and maximum values and description are contained in Table 2 below.

Table 2. Data Table

| <i>Variable</i> | <i>Mean</i> | <i>Min</i> | <i>Max</i> | <i>Description</i> |
|----------------------|-------------|------------|------------|---|
| <i>reln</i> | 5.205644 | 1 | 9 | <i>Institutional dyad relationship</i> |
| <i>dyad_authors</i> | . | . | . | <i>Author pair engaging in joint production</i> |
| <i>dyad_num</i> | 5021.533 | 1 | 10173 | <i>Dyad number</i> |
| <i>dyad_reln_num</i> | 197.2369 | 2 | 673 | <i>Coauthor pair number in the institutional dyad</i> |
| <i>dyad_type</i> | . | . | . | <i>Internal collaboration or external collaboration</i> |

| | | | | | |
|---------------------|-----------|--------|----------|--|--|
| <i>instit_link</i> | . | . | . | | <i>Institutional affiliation of authors linked in this dyad</i> |
| <i>auth1</i> | . | . | . | | <i>Name of the author of focus in an author pair engaging in joint production</i> |
| <i>institut_1</i> | 4.44179 | 1 | 8 | | <i>Institutional affiliation of the author of focus</i> |
| <i>institut_2</i> | 4.515684 | 1 | 8 | | <i>Institutional affiliation of the influencing author</i> |
| <i>jpubs</i> | 0.5398507 | 0 | 50 | | <i>Publications generated by each coauthor pair in a given year</i> |
| <i>jcites</i> | 35.14199 | 0 | 2134 | | <i>Citations associated with the coauthor pair's production</i> |
| <i>spubx1</i> | 1.190604 | 0.05 | 18.05 | | <i>Publication experience associated with the author of focus within a coauthor pair</i> |
| <i>spubx1sq</i> | 3.890565 | 0.0025 | 325.8025 | | <i>Cumulative citations associated with the author of focus within a coauthor pair</i> |
| <i>tspubx2</i> | 0.8432117 | 0 | 36.1 | | <i>tcollab*spubx2</i> |
| <i>xcitx1</i> | 21.04951 | 0 | 671.5 | | <i>Average citations for author of focus</i> |
| <i>citx1</i> | 737.9362 | 0 | 14315 | | <i>Cumulative citations associated with the author of focus within a coauthor pair</i> |
| <i>citx2</i> | 619.5802 | 0 | 14315 | | <i>Accumulated citations associated with the author influencing the author of focus</i> |
| <i>tscitx2</i> | 10.23256 | 0 | 357.875 | | <i>tcollab*(author of influence citation history divided by 40)</i> |
| <i>tcollab</i> | 0.2633895 | 0 | 1 | | <i>Period the co-authorship pair was present</i> |
| <i>ctry</i> | 0.6798352 | 0 | 1 | | <i>Country affiliation of the author of focus' institute</i> |
| <i>neighborhood</i> | 104.7652 | 0 | 5736 | | <i>Number of unique coauthors associated with the author of focus</i> |

| | | | | |
|---------------------|-----------|------|-----------|---|
| <i>IRussian</i> | 5.915561 | 0 | 155 | Log of Russian reprint authors in the author of focus' network neighborhood +1 |
| <i>IUS</i> | 41.36165 | 0 | 1003 | Log of US affiliated reprint authors in the author of focus' network neighborhood+1 |
| <i>instit_labor</i> | 424.1601 | 1 | 1383 | Number of reprint authors associated with the author of focus' institute |
| <i>instit_pubs</i> | 1222.424 | 2 | 5467 | Total publications associated with the author of focus' research institute |
| <i>efficiency</i> | 2.735417 | 1 | 47 | Instit_pubs/instit_labor lagged by one period |
| <i>efficiencysq</i> | 11.15169 | 1 | 2177.77 | Square of Instit_pubs/instit_labor lagged by one period |
| <i>qgp1</i> | 0.0539264 | 0 | 1 | Participation in quark gluon plasma research community |
| <i>ecqgp1</i> | 0.0078979 | 0 | 0.8642226 | Author of focus' eigenvector centrality in quark gluon plasma research |
| <i>nfus1</i> | 0.0845273 | 0 | 1 | Participation of the author of focus in the nuclear fusion research community |
| <i>ecnfus1</i> | 0.0030868 | 0 | 0.9989891 | Author of focus' eigenvector centrality in nuclear fusion research |
| <i>ecinst1</i> | 0.0547133 | 0 | 1 | Author of focus' eigenvector centrality within their institution |
| <i>time</i> | 12.9615 | 1 | 36 | Time period |
| <i>timesq</i> | 245.271 | 1 | 1296 | Time^2 |
| <i>year</i> | 1999.994 | 1977 | 2012 | Year |

Empirical Analysis

Examination of the hypotheses articulated earlier in this document requires specification of three separate empirical models to explore at the pooled US level. The base model (Equation 11) focuses on individual dynamics. It examines joint research generation ($jpubs$) as a function of an author of focus' network neighborhood composition ($Russian$, US), tacit knowledge regarding research publication ($spx1$, $spx1sq$), an influencing author's publication experience ($spx2$) during the period the two may have worked together ($tcollab$), institutional researcher productivity ($efficiency$), the institutional affiliation of the author of focus ($institut_1$), lagged the institutional affiliation of the influencing author ($institut_2$) and time trend($time$). The individual heterogeneity is captured with v_i .

$$(11) \quad \begin{aligned} jpubs_{it} = & \beta_0 + \beta_1 lRussia_{it} + \beta_2 lUS_{it} + \beta_3 spx1_{it} + \beta_4 spx1sq_{it} \\ & + \beta_5 (tcollab_{it} * spx2_{jt}) + \beta_6 efficiency_{i,t-1} + \beta_7 institut_1_i + \beta_8 institut_2_j \\ & + \beta_9 time_{it} + v_i + u_{it} \end{aligned}$$

This previous model contains one nonlinear element. With respect to tacit knowledge, this modeling reflects the notion that the value of a given individual's tacit knowledge in a given technology domain may rapidly appreciate as research-level tools enter commercial usage. If not adequately refreshed through the acquisition of new training or collaboration partners, the value of a given individual's tacit knowledge will depreciate rapidly in many technical research fields due to the obsolescence of enabling technologies and the emergence of new standards (Foray 2004).

The second model adds institutional connectivity to this mix (Equation 12). This model uses eigenvector centrality on an institutional basis (*ecinst1*) to describe the author of focus' position within their institution's research community. As can be seen in the institutional collaboration network mappings displayed in Chapter Three, these institutional research communities are composed of multiple component-level subgraphs pursuing disparate research programs. As such, multiple individuals within the institutional research community will possess the same centrality values.

$$(12) \quad \begin{aligned} jpubs_{it} = & \beta_0 + \beta_1 lRussia_{it} + \beta_2 lUS_{it} + \beta_3 spubx1_{it} + \beta_4 spubx1sq_{it} \\ & + \beta_5 (tcollab_{it} * spubx2_{jt}) + \beta_6 efficiency_{i,t-1} + \beta_7 institut_1_i + \beta_8 institut_2_j \\ & + \beta_9 time_{it} + \beta_{10} ecinst1_{it} + v_i + u_{it} \end{aligned}$$

The third and final model introduces nonlinear elements with respect to institutional efficiency and time. This modeling of institutional efficiency reflects the notion that the introduction of new research facilities within national laboratories and research institutes should trace out the shape of the long run average cost curve. The treatment of time in this model is consistent with the idea that researcher lifecycles should be characterized by a phase in which the individual researcher accumulates human capital at the initial stages of their publication career at a new institution and may exhibit declining productivity as ties to former university colleagues fade. This phase should be followed by a period in which the researcher has accumulated enough new linkages in their new environment to grow in their research productivity.

As importantly, this model also attempts to control for research field (Equation 13). Rather than rely upon impressions of scientific and technology fashions at these research institutions, the two research fields used in this model

are drawn from the peak research interest areas displayed in the latent semantic map research topologies displayed in Chapter Three. The two most prominent research areas depicted in these maps are nuclear fusion related research (*nfus1*) and particle physics research into quark gluon plasmas (*qgp1*). It should be noted that each one of these research fields is composed of distinct and occasionally disjoint research activities (i.e. nuclear fusion research can be decomposed into research into inertial confinement fusion and magnetic confinement fusion, etc.) that are considered to all be one research field for the purposes of this empirical analysis.

$$\begin{aligned}
 (13) \quad jpubs_{it} = & \beta_0 + \beta_1 lRussia_{it} + \beta_2 lUS_{it} + \beta_3 spubx1_{it} + \beta_4 spubx1sq_{it} \\
 & + \beta_5 (tcollab_{it} * spubx2_{jt}) + \beta_6 efficiency_{i,t-1} + \beta_7 efficiencysq_{i,t-1} + \beta_8 ecinst1_{it} \\
 & + \beta_{9,10} institut_1_i + \beta_{11-17} institut_2_j + \beta_{18} nfus1_{it} + \beta_{19} qgp1_{it} + \beta_{20} time_{it} \\
 & + \beta_{21} timesq_{it} + v_i + u_{it}
 \end{aligned}$$

Empirical analysis of these models requires employing statistical techniques that are appropriate for considering count data organized by collaboration pair publication history. Focusing on count data suggests that analysis employ either a Poisson or negative binomial regression model. In addition, the publication histories associated with each member of the collaboration pair imply that a technique appropriate for examining unbalanced panels be employed.

While Poisson regression techniques are traditionally used in considering count data, there are some fairly strong assumptions that must be met for Poisson regression estimates to be accurate. Chief among these assumptions is the requirement for equidispersion - the observed mean must be roughly equal to the observed variance. Correction techniques must be applied, up to and

including change to the negative binomial distribution, if the data exhibits overdispersion, e.g., the observed variation in the data is significantly greater than the mean. Similarly, if the data exhibited underdispersion requires appeal to generalized event count models.

As with most examinations of researcher publication history panels, there are a large number of zeroes in the dataset. This characteristic alone is often observed to suppress the mean relative to the variance (Winkelmann, 2008). As such, the data used in this examination was unlikely to have the equidispersion characteristics necessary for examination via Poisson regression. To test for overdispersion, a series of likelihood ratio tests were conducted regarding whether the overdispersion parameter (referred to as α) was equal to zero across this set of models. In each case, overdispersion was shown to be greater than zero. Given these test results, negative binomial panel regression analysis was selected as the starting point for empirical examination.

These likelihood ratio tests also generated AIC and BIC statistics. Uniformly, these statistics validate the choice of the negative binomial context over the Poisson context for these models. In addition, the third model described above (Equation 13) received the lowest AIC and BIC rating of the three models signaling it is likely the best of the three models.

Given these findings, it is important to specify how unobserved heterogeneity is correlated with the observed explanatory variables. Broadly speaking, there are two primary choices in this specification. The model can either contain a random effects specification in which there is zero correlation between unobserved effects and the explanatory variables or it can contain a fixed effect specification in which the unobserved effects possess a relationship

with the explanatory variables. Following Wooldridge (2010), a Hausman specification test was conducted on the third model in the negative binomial panel regression context, to determine which of the two specifications better modeled the unobserved heterogeneity in the data. This test indicated that a random effects context was not appropriate (chi square value = 752.11).

One of the challenges in applying this type of analysis consists of overcoming significant serial correlation in this dataset. There are two key sources of this serial correlation problem. First, the data set includes time-cumulative variables (*spx1*, *spx1sq*, *spx2*). Additionally, other variables (*Russian*, *US*, *tcollab*, *ecinst1*) had to be constructed using a moving five-year snapshot of network relationships to avoid unreasonable fragmentation of network structure. Tests for the presence of autocorrelation within this panel data (as per Wooldridge, 2002) revealed a significant serial correlation problem across all models on all relevant data subsets.

These panels should also contain considerable heteroscedasticity. Increased institutional researcher productivity over time, the explosion in research journals, and the growth in large multi-national research collaborations that generate published research should all lead to greater variance over time. Repeated adapted Wald tests for heteroscedasticity (Reyna, 2007) validated this expectation. Each of the models contains significant groupwise heteroscedasticity.

The influence author's publication experience variable is likely to be an endogenous covariate. To specifically examine the endogeneity issue, this empirical examination conducted a Hausman specification test with the influencing author's scaled publication experience (*tspx2*) as the potential

endogenous covariate. This specification test revealed that indeed, *tspubx2* was potentially endogenous.

To reduce this endogeneity, three instruments were selected. The key instrumental variable in this approach is the cumulative citation level associated with the influencing author's work (*tscitx2*) when the two researchers were deemed to be working together. The other two instruments chosen were *time* and *tcollab*. As required by this technique, this instrumental variable was found to be significant in each of the first stage regressions during the process of model selection (Wooldridge, 2010). The best set of instruments was selected on the basis of explained variance (R^2).

At present, there is no straightforward way to control for all of these issues simultaneously. In this context, the best alternative is to carry out parameters estimations via either a moments based method or generalized estimating equation approach (Trivedi, 2010). To specifically account for the serial correlation and heteroscedasticity issues in a count data context, this examination also used a population averaged negative binomial general estimating equation approach to panel data analysis using log-linked explanatory variables. This approach was selected because a second series of Hausman specification tests were conducted that implied that if the serial correlation and heteroscedasticity issues could be remedied, the endogeneity issues would be reduced as well.

This general estimating equation method allowed for within-group serial correlation to be accounted via a use of an appropriate correlation matrix. In the case of the data under consideration in this study, the observed serial correlation in the panels appeared to follow an AR(1) process. Heteroscedasticity is

controlled for by the use of clustered robust errors. The quasi-likelihood information criterion or QIC (Cui, 2007) was used for model selection. This approach falls short of a complete solution because of its inability to specifically address endogenous covariates. A sensitivity analysis associated with this approach is presented in Table 4.

Results

To fully appreciate the contribution of these variables to national laboratory researcher productivity, it is necessary to generate models in multiple contexts. The highest-level context in this study is the country level. Given the observation of institutional and research field heterogeneity in this context, this consideration is followed by examination of individual US national laboratory contexts and an examination of individual research fields.

The pooled US results are contained in Table 3 below. These results include all collaborations that US researchers participated in – both internal to their own laboratories and their collaborations with their Russian laboratory counterparts. It is notable that the results for the four models in the pooled context largely possess similar signs within the core model (as represented by Equation 11) where estimates are deemed to be significant. The contribution of publication experience to joint publication generation appears to follow an inverted parabola, signaling the presence of diminishing returns to tacit knowledge stock past an optimal point. Both the personal attributes and institutional affiliation of an influencing coauthor have a significant impact on the generation of joint publications. Individual research productivity appears to

react to improvements in institutional researcher productivity parabolically – only generating returns past a critical point.

Table 3. Pooled GEE Models - US, SNL, LLNL, LANL⁹

| | US GEE | | SNL GEE | | LLNL GEE2 | | LANL GEE | |
|--|------------|----------|-------------|----------|--------------|----------|-------------|----------|
| <i>IRussia</i> | 0.549*** | (29.06) | 0.163*** | (5.01) | 0.294*** | (7.09) | 0.922*** | (19.97) |
| <i>IUS</i> | 0.184*** | (12.02) | 0.395*** | (13.27) | 0.287*** | (6.43) | -0.000199 | (-0.01) |
| <i>spxbx1</i> | 0.305*** | (17.03) | 0.380*** | (8.66) | 0.393*** | (5.73) | 0.530*** | (20.26) |
| <i>spxbx1sq</i> | -0.0234*** | (-11.47) | -0.0508*** | (-7.41) | -0.0309*** | (-3.71) | -0.0396*** | (-15.07) |
| <i>tspbx2</i> | 0.204*** | (77.14) | 0.286*** | (56.98) | 0.199*** | (22.83) | 0.158*** | (46.89) |
| <i>efficiency</i> | -0.139* | (-2.27) | -0.0479 | (-0.66) | -0.149** | (-2.89) | -0.392*** | (-15.30) |
| <i>efficiencysq</i> | 0.00283* | (2.31) | 0.00367 | (1.34) | 0.00267* | (2.46) | 0.00803*** | (13.76) |
| <i>ecinst1</i> | 0.642*** | (7.61) | 0.517* | (2.18) | -0.541* | (-2.48) | 0.359*** | (4.22) |
| <i>nfus1</i> | 0.281*** | (8.74) | -0.189*** | (-3.47) | -0.117 | (-0.74) | 0.632*** | (16.50) |
| <i>qgp1</i> | 0.498*** | (20.77) | -0.731*** | (-6.34) | 0.405*** | (8.09) | 0.510*** | (18.26) |
| <i>time</i> | -0.0428*** | (-7.30) | 0.0768*** | (6.39) | -0.0586* | (-2.55) | -0.103*** | (-13.55) |
| <i>timesq</i> | 0.000253 | (1.52) | -0.00336*** | (-10.18) | -0.000794 | (-1.19) | 0.00209*** | (10.20) |
| <i>_cons</i> | -4.697*** | (-21.84) | -5.017*** | (-31.18) | -3.317*** | (-12.72) | -4.643*** | (-33.00) |
| <i>N</i> | 125011 | | 60483 | | 9407 | | 55121 | |
| <i>t statistics in parentheses</i> | | | | | | | | |
| * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$ | | | | | | | | |

Despite these similarities in sign, there are some notable differences in relationships between these variables across these three contexts. For example, the relationship between the Russian network neighborhood variable and the US neighborhood variable is inconsistent across the panel. In the pooled model and at LANL, this relationship signals a greater marginal contribution associated with an additional Russian network neighborhood member than an additional US researcher. At SNL, this relationship is reversed.

Similarly, there is variation across the set of laboratories in the relationship between the individual researcher's life cycle and their joint productivity. In the US pooled model and the LANL specific results a parabola

⁹ Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

best models this relationship. This relationship would be consistent with individual researchers investing in human capital at the beginnings of their careers – when they have low productivity -- and as these individuals accumulate access to greater knowledge via the cultivation of colleagues to combine with a critical level of tacit knowledge, their productivity explodes.

At LLNL and SNL, this relationship is inverted. In this context, it is notable that both of these institutions bill themselves as engineering laboratories. This relationship is more in consistent with individuals at these laboratories being productive in generating basic research early in their careers, but exiting basic research production as their stock of tacit knowledge declines in value.

A comparison across techniques is shown in Table 4. The first comparator to the general estimating equation approach is a fixed effect negative binomial panel regression model as suggested by the initial Hausman test. The other comparison model is a pseudo-instrumental variables approach within the general estimating equation approach. To carry out this approach, the negative binomial model estimated as the initial stage of the 2SLS instrumental variables approach was used to generate predicted values for the endogenous variable (*tspubx2_p*). These predictions were then used as a replacement for *tspubx2* in the conventional GEE model.

Table 4. Sensitivity Analysis for the US Model - NB (FE), GEE, "IV" GEE¹⁰

| | <i>Pooled_ NB (FE)</i> | | <i>Pooled GEE</i> | | <i>Pooled_IV GEE</i> | |
|--|----------------------------|----------|-----------------------|----------|--------------------------|----------|
| <i>IRussia</i> | -0.0227 | (-1.45) | 0.549*** | (29.06) | 0.0936*** | (6.94) |
| <i>IUS</i> | 0.465*** | (22.61) | 0.184*** | (12.02) | 0.171*** | (11.43) |
| <i>spx1</i> | -0.00795 | (-0.55) | 0.305*** | (17.03) | 0.193*** | (13.47) |
| <i>spx1sq</i> | 0.000808 | (0.73) | -0.0234*** | (-11.47) | -0.00962*** | (-6.55) |
| <i>spx2_p</i> | 1.470*** | (41.87) | | | 1.050*** | (45.79) |
| <i>efficiency</i> | -0.420*** | (-25.58) | -0.139* | (-2.27) | -0.00646 | (-0.18) |
| <i>efficiencysq</i> | 0.00994*** | (23.76) | 0.00283* | (2.31) | 0.00132 | (1.92) |
| <i>ecinst1</i> | -0.174*** | (-5.56) | 0.642*** | (7.61) | 0.462*** | (5.37) |
| <i>nfus1</i> | -0.441*** | (-14.33) | 0.281*** | (8.74) | -0.0992*** | (-3.69) |
| <i>qgp1</i> | 0.249*** | (20.40) | 0.498*** | (20.77) | 0.460*** | (23.54) |
| <i>time</i> | -0.164*** | (-33.00) | -0.0428*** | (-7.30) | -0.105*** | (-24.46) |
| <i>timesq</i> | 0.000590*** | (5.56) | 0.000253 | (1.52) | 0.000210 | (1.63) |
| <i>spx2</i> | | | 0.204*** | (77.14) | | |
| <i>_cons</i> | -19.26*** | (-11.30) | -4.697*** | (-21.84) | -15.85*** | (-52.20) |
| <i>N</i> | 121421 | | 125011 | | 125011 | |
| <i>t statistics in parentheses</i> | | | | | | |
| * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$ | | | | | | |

This comparison shows broad agreement between the pseudo-instrumental variables general estimating equation technique and the conventional general estimating equation approach with respect to the signs of calculated parameter estimates. While there is parameter deviation between the two models (especially with respect to the dummy control variables), the two models show the same relative phenomenon shapes (i.e. downward facing parabola for publication experience, parabolas for the relationship between researcher productivity at the institutional level and time trend). The parameters estimated via the negative binomial fixed effects model often differ in both sign and magnitude from the conventional general estimating equation approach. As such, this approach does not even depict the same relationship in the data as the

¹⁰ Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

two other approaches in the case of the publication returns to publication experience.

For the purpose of broader comparison, the results of the pseudo-instrumental variables general estimating equation technique are presented in Table 5. With respect to our variables of key interest in this study there are some notable sign changes. The Russian neighborhood variable goes from being small and positive with significance in the original GEE model in the SNL context to being insignificant. Likewise, in the LANL context, the US network neighborhood variable parameter goes from being negative and significant to being positive and significant.

Table 5 "IV" GEE Models - Pooled US, SNL, LLNL, LANL¹¹

| | <i>Pooled_IV</i> | | <i>SNL_IV_GEE</i> | | <i>LLNL_IV_GEE</i> | | <i>LANL_IV_GEE</i> | |
|--|------------------|----------|-------------------|---------|--------------------|----------|--------------------|----------|
| <i>IRussia</i> | 0.0936*** | (6.94) | -0.0442 | (-1.83) | 0.0234 | (0.67) | 0.248*** | (7.65) |
| <i>IUS</i> | 0.171*** | (11.43) | 0.0511* | (1.99) | 0.311*** | (7.32) | 0.203*** | (9.54) |
| <i>spubx1</i> | 0.193*** | (13.47) | 0.123* | (2.27) | 0.288*** | (4.85) | 0.299*** | (16.53) |
| <i>spubx1sq</i> | -0.00962*** | (-6.55) | -0.00903 | (-0.88) | -0.0167* | (-2.26) | -0.0189*** | (-10.65) |
| <i>tspubx2_p</i> | 1.050*** | (45.79) | 2.096*** | (7.40) | 0.878*** | (15.57) | 0.891*** | (36.28) |
| <i>efficiency</i> | -0.00646 | (-0.18) | -0.0334 | (-0.45) | -0.298*** | (-4.92) | -0.166*** | (-6.80) |
| <i>efficiencysq</i> | 0.00132 | (1.92) | 0.00548 | (1.91) | 0.00642** | (5.27) | 0.00427*** | (8.19) |
| <i>ecinst1</i> | 0.462*** | (5.37) | 0.644* | (2.28) | -0.0183 | (-0.06) | 0.148 | (1.80) |
| <i>nfus1</i> | -0.0992*** | (-3.69) | -0.101 | (-1.96) | -0.338** | (-2.60) | 0.0980** | (3.08) |
| <i>qgp1</i> | 0.460*** | (23.54) | -0.590*** | (-5.32) | 0.208*** | (5.75) | 0.433*** | (17.93) |
| <i>time</i> | -0.105*** | (-24.46) | -0.098*** | (-4.31) | -0.170*** | (-10.03) | -0.124*** | (-23.92) |
| <i>timesq</i> | 0.000210 | (1.63) | -0.0018*** | (-7.26) | 0.00172*** | (3.87) | 0.00116*** | (7.83) |
| <i>_cons</i> | -15.85*** | (-2.20) | -30.90*** | (-7.57) | -15.78*** | (-15.50) | -11.60*** | (-46.65) |
| <i>N</i> | 125011 | | 60483 | | 9407 | | 55121 | |
| <i>t statistics in parentheses</i> | | | | | | | | |
| * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$ | | | | | | | | |

The pooled Russian models are presented in Table 6. Many of the trends observed in the pooled US model appear in the Russian context as well. The

¹¹ Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

signs and significance of personal attributes and institutional affiliation of an influencing coauthor, relative contribution of a Russian reprint author in a researcher's network neighborhood to joint productivity with respect to a US reprint author, and the general shape of the contribution of an individual author's tacit knowledge to joint productivity are all similar. The most notable deviation between the two sets of models concerns the response of joint productivity to institutional efficiency in the pooled Russian model. This shape is driven by the response of the non-nuclear weapons Russian research institutes in the dataset. KIAE, VNIIEF, and VNIITF all share the same shape as in the US context.

Table 6 Pooled GEE Models - Russia, KIAE, VNIIEF and VNIITF¹²

| | <i>Russia</i> GEE | | <i>KIAE</i> GEE | | <i>NW</i> GEE | |
|------------------------------------|----------------------|-----------------------|--------------------|----------|------------------|----------|
| <i>lRussia</i> | 0.480*** | (18.46) | 0.642*** | (14.31) | 0.545*** | (35.82) |
| <i>lUS</i> | 0.358*** | (21.31) | 0.268*** | (10.37) | 0.175*** | (16.05) |
| <i>spubx1</i> | 0.624*** | (6.83) | 1.935*** | (8.39) | 0.309*** | (16.93) |
| <i>spubx1sq</i> | -0.0991*** | (-5.12) | -0.412*** | (-6.38) | -0.0243*** | (-11.08) |
| <i>tspubx2</i> | 0.292*** | (40.97) | 0.294*** | (25.56) | 0.226*** | (89.66) |
| <i>l_instit_efficiency</i> | 0.517*** | (4.33) | -0.793 | (-1.95) | -0.0937** | (-2.68) |
| <i>l_instit_efficiency_sq</i> | -0.0676*** | (-3.36) | 0.0319 | (0.42) | 0.00188** | (2.73) |
| <i>ecinst1</i> | 1.545*** | (22.32) | -4.063*** | (-5.55) | 0.852*** | (15.11) |
| <i>nfus1</i> | -0.222** | (-2.93) | 0.0733 | (0.32) | 0.255*** | (8.37) |
| <i>qgp1</i> | -0.177 | (-0.73) | -5.323*** | (-41.48) | 0.540*** | (23.36) |
| <i>Time</i> | 0.00641 | (0.57) | 0.0247 | (1.01) | -0.0278*** | (-5.29) |
| <i>Timesq</i> | -0.00159*** | (-4.55) | -0.00222** | (-3.19) | -0.000212 | (-1.38) |
| <i>_cons</i> | -4.887*** | (-24.93) | -2.461*** | (-5.31) | -3.478*** | (-28.32) |
| <i>N</i> | 58457 | | 23106 | | 183468 | |
| <i>t statistics in parentheses</i> | | | | | | |
| <i>* p<0.05</i> | <i>** p<0.01</i> | <i>*** p<0.001</i> | | | | |

¹² Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

It is clear from these results that there is institutional-level heterogeneity with respect to joint research production in this dataset. This heterogeneity may emerge from institutional structure differences as well as researcher life cycle differences across national laboratories. By inspection, it is also clear that a significant degree of heterogeneity to joint research productivity exists with respect to research field. Given the discussion offered in Chapter 3, this heterogeneity comes as little surprise. The two primary topics of research focus for many of the research institutions discussed in this study possess unique network structures that deviate from the atomistic collaboration model. As previously indicated, nuclear fusion research has often taken place in a more “bilateral” collaborative fashion. As such, two key research field contexts – nuclear fusion and all other pursuits – will be compared in the following analysis tables.

The relative contributions of factors influencing joint research production performance of US national laboratory research collaborations in the nuclear fusion milieu are documented in Table 7. The most notable trend that can be observed is the estimated relationship between Russian reprint authors and joint research productivity in this context. At LLNL, this relationship is estimated as being negative. It is of negligible significance at SNL. At LANL, it is positive. In the pooled context, the larger number of observations associated with the nuclear fusion activities at LANL is responsible for the signs attributed to both Russian and US network neighborhood members in the pooled context.

Two other observations can be made. As indicated previously, there is often a parabolic shape to the productivity returns associated with increases in institutional researcher productivity. Unsurprisingly, each of the institutions

that compose this data set display this same relationship. It should be noted that this observed relationship is contrary to the relationship observed at the aggregate institutional level for both LLNL and SNL. Both of these institutions possess on-site “big science” nuclear fusion linked experimental facilities and as such display the negative returns to publications as efficiency increases through the development of new facilities. Once these facilities enter into operation, there are positive returns associated with the new infrastructure. The useful shelf life of those facilities also appears to drive the behavior of the researcher’s life cycle as observed through the two time trend variables.

Despite this observed relationship, it is also the case that there is a strongly negative association between researcher prominence as measured via eigenvector centrality within the nuclear fusion research field. This relationship reflects the relatively high mobility of nuclear fusion researchers in the US

Table 7. Nuclear Fusion - US, LLNL, SNL and LANL GEE Models¹³

| | US GEE | | LLNL GEE | | SNL GEE | | LANL GEE | |
|--|-----------|---------|-----------|---------|-----------|---------|----------|---------|
| <i>IRussia</i> | 0.660*** | (15.06) | -0.612*** | (-4.02) | 0.0938 | (1.65) | 1.475*** | (14.73) |
| <i>IUS</i> | 0.165*** | (5.10) | 0.525*** | (4.51) | 0.633*** | (6.14) | -0.30*** | (-4.49) |
| <i>spubx1</i> | 0.542** | (3.02) | 1.042*** | (3.93) | 0.374*** | (3.72) | 2.984*** | (8.26) |
| <i>spubx1sq</i> | -0.0458* | (-2.02) | -0.056*** | (-4.14) | -0.039*** | (-4.19) | -0.35*** | (-6.52) |
| <i>tspubx2</i> | 0.237*** | (26.34) | 0.216*** | (12.26) | 0.274*** | (23.67) | 0.153*** | (10.00) |
| <i>efficiency</i> | -1.180 | (-1.36) | -2.678 | (-0.78) | -0.442 | (-0.32) | -2.573 | (-1.14) |
| <i>efficiencysq</i> | 0.245 | (1.76) | 0.350 | (0.67) | 0.0374 | (0.16) | 0.585 | (1.63) |
| <i>ecinst1</i> | -0.876*** | (-3.52) | -0.601 | (-1.78) | -0.468* | (-2.40) | 0.935 | (0.98) |
| <i>ecnfus1</i> | -0.567** | (-3.24) | -2.264*** | (-3.81) | -5.509** | (-3.14) | -0.811** | (-3.12) |
| <i>qgp1</i> | -0.355*** | (-3.89) | -0.429 | (-0.77) | -0.447** | (-3.16) | -0.283 | (-1.48) |
| <i>time</i> | -0.117*** | (-4.80) | 0.128 | (1.61) | 0.095*** | (3.40) | -0.30*** | (-6.23) |
| <i>timesq</i> | 0.000600 | (1.07) | -0.0076** | (-3.28) | -0.004*** | (-5.27) | 0.00256* | (2.46) |
| <i>_cons</i> | -2.256 | (-1.73) | -0.0970 | (-0.02) | -5.714** | (-2.78) | -0.588 | (-0.17) |
| <i>N</i> | 12204 | | 831 | | 6684 | | 4689 | |
| <i>t statistics in parentheses</i> | | | | | | | | |
| * <i>p</i> <0.05 ** <i>p</i> <0.01 *** <i>p</i> <0.001 | | | | | | | | |

¹³ Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

national laboratories from one facility to another. It is often the case that researchers who play prominent roles in the development of one facility will be invited to participate in the construction (or maintenance) of another such facility. In the case of magnetic confinement fusion work, this path actually leads to work on ITER tokamak facility in France – outside the US national laboratory community.

It is necessary to inspect Table 8 to understand how anomalous this set of estimates is within the broader dataset. This table describes the estimated

Table 8. Rest of Research Fields - US, LLNL, SNL, LANL¹⁴

| | US | | LLNL | | SNL | | LANL | |
|------------------------------------|------------|----------|-------------------|----------|------------|---------------------|------------|----------|
| <i>IRussia</i> | 0.459*** | (19.00) | 0.264*** | (4.26) | 0.158*** | (4.55) | 0.863*** | (14.51) |
| <i>IUS</i> | 0.282*** | (14.69) | 0.215*** | (4.60) | 0.373*** | (12.16) | 0.0892* | (2.00) |
| <i>spubx1</i> | 0.224*** | (8.67) | 0.458*** | (6.10) | 0.411*** | (8.15) | 0.446*** | (12.09) |
| <i>spubx1sq</i> | -0.0151*** | (-4.93) | -0.0353*** | (-5.21) | -0.0566*** | (-6.56) | -0.0318*** | (-9.05) |
| <i>tspubx2</i> | 0.239*** | (76.00) | 0.231*** | (19.06) | 0.288*** | (58.66) | 0.186*** | (39.04) |
| <i>efficiency</i> | 0.0366 | (0.97) | -0.00182 | (-0.02) | -0.0200 | (-0.26) | -0.274*** | (-7.62) |
| <i>efficiencysq</i> | -0.00086 | (-1.09) | -0.000142 | (-0.08) | 0.00261 | (0.90) | 0.00555*** | (7.13) |
| <i>ecinst1</i> | 0.467*** | (3.52) | -0.542 | (-1.58) | 0.690** | (2.61) | -0.0465 | (-0.39) |
| <i>time</i> | -0.0127 | (-1.55) | 0.0117 | (0.37) | 0.0780*** | (6.05) | -0.107*** | (-9.70) |
| <i>timesq</i> | -0.00050* | (-2.26) | -0.0028** | (-3.14) | -0.0033*** | (-9.47) | 0.00223*** | (7.89) |
| <i>_cons</i> | -4.809*** | (-27.23) | -3.623*** | (-10.44) | -4.985*** | (-29.17) | -4.304*** | (-21.29) |
| N | 103544 | | 6940 | | 53389 | | 43215 | |
| <i>t statistics in parentheses</i> | | | | | | | | |
| * <i>p</i> <0.05 | | | ** <i>p</i> <0.01 | | | *** <i>p</i> <0.001 | | |

empirical relationships observed over the span of the rest of fields of inquiry. In these areas, the conventional relationship between Russian members of researchers' network neighborhoods and joint production activity appears to hold at both LANL and LLNL. Ultimately, these “rest of fields” collaborations

¹⁴ Table omits the institutional dummy variables used to control for institutional effects within the third empirical model for the sake of presentation clarity.

are responsible for coloring the results observed in the pooled US model presented earlier in this section.

Findings

In all, the chapters prior to this empirical examination proposed one central hypothesis. In atomistic collaboration, researcher network neighborhood composition will be the result of individual researchers balancing marginal research productivity associated with the participation of an additional researcher to their network neighborhood against the cost of maintaining a link with that researcher. If types of researchers have common link maintenance cost characteristics, individual researchers' network neighborhood compositions will reflect these common characteristics. Individual researchers will have greater numbers of low maintenance cost collaboration partners and relatively fewer higher maintenance cost collaboration partners. Given the estimated contribution of the Russian and US reprint authors to joint productivity in the pooled models and the relationship of these estimates to one another, this hypothesis appears to be supported.

When intervention external to the normal scientific process occurs, as in the case of "bilateral collaborations," this mix of researcher types in individual researchers' network neighborhoods should be altered from this atomistic state. It is likely that researchers who participate in these bilateral collaborations will have greater or less numbers of one of the researcher types than each would prefer in the alternative state. This deviation from the optimal mix should result in diminished productivity and thus, lower formation of tacit knowledge for participating researchers. Again, the estimated contributions of the Russian and

US reprint authors to joint productivity in the nuclear fusion and the relationship of these estimates to one another suggests this hypothesis is also supported.

The higher link maintenance cost that appears to be associated with maintaining Russian network neighborhood members should not come as much of a surprise. Communication is often difficult with Russian scientists for US national laboratory researchers because of the significant language barrier. Preparation of research papers can take twice as long because all submissions from the non-English speaking researchers often have to be edited for content prior to submission. Research collaborators are unlikely to know how to use the same equipment because they are submerged in different enabling technology states. Coordination of work often requires expensive and time-consuming face-to-face negotiations. In this context, it is clear that if a given national laboratory researcher wishes to add a Russian scientist to their network neighborhood, that scientist has to pass a perceived benefit hurdle that the US national laboratory researcher may not apply to their local colleagues.

It is clear from this consideration that participation in international technical collaborations is a step that some researchers take on the way to a permanent change in status within their research facility that fundamentally lowers their research productivity. For example, participation in international collaborations is often a positive signal to research facility management of the researcher's suitability for management. Participating in international high energy physics research collaborations as a key member often involves coordinating tens of researchers at multiple research establishments with large budgets under challenging deadlines. Likewise, if a successful researcher is not on a path to a career in management, a good track record in managing

international cooperation can result in the researcher being offered a senior advisory post that also takes the researcher away from the bench top. Similarly, failure to perform well in international research collaboration could end a researcher's career at the laboratory. Each of these factors results in a negative association between a researcher's institutional eigenvector centrality and their joint research production.

So, which side loses more?

As measured, Russian researchers experience productivity suppression in more institutional relationships than their US colleagues. As such, one takeaway from this particular observation is that there may be too many US national laboratory researchers working with Russian research institute scientists in collaborative research endeavors. If so, this level of engagement with US scientists may be actually hindering Russian research institute researchers from developing the kind of relationships with their local colleagues that contribute to the formation of high-performing research clusters. Thus, from a Russian perspective, gradual disengagement may be a prudent path toward redevelopment of the Russian scientific and research community.

Such a movement may already be underway. Recent technology foresight studies on nanotechnology development in Russia funded by the Russian government do not possess the same outward looking, collaboration seeking tone of previous work. The current studies appear to stress indigenous development (Karasev and Edelkina, 2013).

In the relationships where US national laboratory researchers experience publication suppression related to their Russian colleagues, however, the suppression tends to be more focused. Indeed, if net effects are compared, the

degree of suppression in these unfavorable relationships is significant enough that it overwhelms the marginal differences observed across all of the other relationships. Thus, it is also reasonable for US research policy decision-makers to also support a policy of gradual disengagement in these areas.

Conclusions

This examination suggests that individual researchers and by-and-large institutions in bilateral collaboration areas do not appear to benefit from enhanced open science research productivity from these linkages with Russian research institute partners in this context. Verification of these results in other domains is problematic. The set of all government reports generated by joint Russian-US collaborations is not available for comparative analysis. Difference in the implementation of patent regimes between Russia and the United States has frustrated serious empirical attempts to understand if the lack of improvement to publication frequency is offset by increases in the government's holdings of productive intellectual property.

It is also unknown if these national laboratory institutions derive enough national benefits from these interactions to continue pursuit of these relationships. It is clear that if these collaborations were entered into primarily as a means to reduce nuclear danger, they have proved to be a qualified success. After the collapse of the Soviet Union, two nation-states have demonstrated nuclear weapons capability - Pakistan and North Korea. Yet another country, Iran is perceived to be on the brink of a nuclear weapons capability. However, the observed open role of Russian weapons expertise in fueling this horizontal

proliferation appears to have been meager at best. Indeed, North Korea and Pakistan are both viewed as clients of China, not Russia.

Despite this qualified political success, the merit of continuing these collaborations as a means of controlling nuclear danger appears to be out-dated. Russia, using infusions of export revenue from the foreign sale of oil, has recapitalized part of its formerly grand scientific complex. Wages for Russian scientists are comparable to their Western compatriots. Russia is exhibiting the results of a corporatist research policy - namely Russian research institutions that serve state aims closely receive funding at the expense of their basic research oriented colleagues.

There is some anecdotal evidence that suggests the crossover benefits may be significant in scope. VNIIEF, Kurchatov and Sandia collaborated on the redevelopment of the Z-facility (the so-called ZR facility) opened in 2007 at Sandia. Part of this redevelopment involved the adoption of linear transformer drivers (LTDs) pioneered at VNIIEF to replace Marx generators within the facility. The adoption of LTDs allowed the Z-facility to carry out orders of magnitude more inertial confinement events per day. This change was significant enough to push inertial confinement fusion plants (Z-IFE) into the discussion again along side magnetic confinement fusion facilities as the possible fusion power plant of choice in the future. Fittingly enough, however, the intellectual property rights for the LTD are currently under dispute.

V. Concluding Remarks

The Changing Mission Mix

The perceived success of collaborative scientific endeavors between US and Russian research institutes in the wake of the Cold War has led to questions about how US national laboratories' missions should be repurposed in the US national innovation system (Scotchmer, 2004). The decline in perception of the Soviet Union as a nuclear threat is viewed as a factor reducing demand for nuclear weapons science, and thus, "national security" oriented work at the sites. This perceived change in demand is viewed to have hastened the growth of open science as well as technology development activities similar to those that are a part of industrial laboratory work at these institutions as the percentage of US national security oriented work declined. However, the increased role of the open science and industrial development models did not come without

competing implications for the knowledge competencies at the root of the laboratories' success.

Neither open science research nor commercial technology transfer were new to these US research facilities at the end of the Cold War. As can be seen in the gradual rise in published journal articles at these institutions, open science had already been growing at these laboratories - a trend that dates back to the US governmental push for renewable energy research during the energy crises of the 1970s. As can be seen in the R&D 100 awards prior to 1977, DOE/NNSA laboratories were engaged in commercial transfer oriented technology development well before the end of the Cold War. However, while the proprietary research model gained momentum from the push for technology transfer following the Soviet Union's collapse, the poor early record of US national laboratories in technology commercialization denied this model enough momentum to be a significant factor at many US national laboratories (Jaffe, Fogarty and Banks, 1998).

With the emergence of basic research as a greater component of US nuclear weapons laboratory work came greater acceptance of a research and development model that challenged the dominance of more appropriation-oriented research and development models at the laboratories. This development created a dynamic tension within these institutions that mirrored the outside uncertainty about what role national laboratories should play in the innovation ecosystem, a result that should be expected when these competing research models nest in a common institution (Aghion, David and Foray, 2009).

The open science model as described by Foray (2004) is a model in which individual researchers are rewarded for their other contributions to the

organization (e.g., performing teaching duties at a university or managing a technical team at a national laboratory) rather than the generation of knowledge. This form of reward system allows researchers to earn a steady income - an outcome that would not be assured with a discovery based reward system (Dasgupta and David, 1994). The incentive to generate new knowledge comes in the form of generated reputation effects that give the individual researcher who makes discoveries priority in the assignment of patronage in the form of grant money. Thus, individual researchers under the open science model have an incentive to publish their results promptly so they can attain control over their own research program.

This incentive to immediately disclose discoveries that is so much a part of the open science model is different from incentives under a typical national benefits/security model. While it is impossible to specify a generalized national benefits/security model owing to the diversity of national priorities by nations with national laboratories (Foray 2004), there is at least one pertinent observation that can be made from the US experience. Under the national benefits/security model in the US national laboratories, researchers receive comparable pay to their colleagues in industrial laboratories. This rate of pay is higher than what university colleagues working in similar fields receive, ostensibly for teaching their fields to incoming students and attracting grants to the university. This remuneration standard was put into place to serve three purposes:

- Attract elite quality researchers to work on technical challenges affecting national capabilities,
- Provide high income security to promote technical risk-taking, and

- Promote the emergence of long-term capability growth rather than just myopic refinement and combination of existing technology - long a focus of the proprietary research model as implemented by modern industrial laboratories.

This higher rate of remuneration serves another function as well - compensation for foregone prestige for discoveries and technical advancements. In US facilities that possess a dominant national benefits/security orientation, technical developments and discoveries are not immediately disclosed - an area of commonality with respect to the proprietary research model (Aghion et al., 2009). If the national laboratory conducted its investigations at the behest of another government entity, the technical development may never be disclosed publicly - unless that entity desires the disclosure. Even if the discovery is funded in a publicly disclosed way, there are multiple hurdles that must be passed prior to disclosure including reviews for export control and classified matter. Such hurdles are often either not present or underdeveloped in university settings - regardless of whether the institution is public or private (Mowery, 2004).

As Foray (Foray, 2004) notes, both the open science and national benefits models are descriptions of research incentives that exist at either end of the research spectrum for public institutions that receive funding from national sources. Any national laboratory is likely to have characteristics of both models. Understanding which model is dominant at any particular time, however, requires the development of some indirect measures and some fairly strong assumptions.

Both of these incentive structures clearly have their implications for innovation readiness and appropriability within the laboratory. Those organizations within the national laboratories that have open science type incentives within them (e.g., performance metrics like refereed journal publications per year, conference presentations given, etc.) help the national laboratories as a whole to remain innovation-ready in key competency areas by maintaining connections to networks of individuals outside the walls of the national laboratory and its supporting university/industrial partnerships. However, these organizations threaten the ability of the laboratory or the nation to appropriate the full benefit of the development because a significant portion of the appropriability advantage is lost by the immediacy of the disclosure. Conversely, areas within the laboratory that hew more closely to a national benefits model will have difficulty remaining on top of technical developments due to an absence of linkages with scientists outside the national laboratory. These organizations, however, are masters at maximizing national appropriability by limiting the loss of information.

This difference in connectivity between the two models suggests that these two ideal types have different implications for the evolution of laboratory competencies. To demonstrate these implications, a laboratory competency could be thought of as a stock of relevant knowledge for technology development. Under the open science model, the laboratory's stock of knowledge is linked to numerous other such competencies at other research facilities - linkages that help the laboratory readily adopt and accept research results generated at other facilities, but also are avenues for knowledge loss. Any momentary increase in the open science oriented laboratory's stock of knowledge raises the level of

knowledge for all organizations with the same skill set, but with diffusion related lag. By comparison, the national benefit oriented stock of knowledge has fewer outside linkages, but more linkages within the research institution. As such, these local developments of knowledge are likely to remain concealed for a longer period of time, but as new knowledge is also filtered, the knowledge within this stock is at risk of becoming stagnant.

As might be discerned from this previous discussion, competition between incentive structures has significant consequences for mission maintenance for US national laboratories. If the majority of the laboratory is guided by open science principles, the laboratory is likely to become the equivalent of a nationally run research university with less capability to perform sensitive or nationally oriented technology development projects. Conversely, if the majority of the laboratory is oriented toward national benefits, the laboratory limits its ability to remain a cutting edge research and development institution. Clearly, a balance must be maintained between the two perspectives to ensure the national laboratories remain viable parts of the US national innovation system that do not just replicate functions served by other constituents within that system (Aghion et al., 2009).

Networks as the Analytic Substrate

In the previous chapters, tacit knowledge has been established as a key asset of a research and development organization. As any asset, its development must be monitored and its transfer guarded to secure long run competitive advantage for a research and development organization. However, the protection of tacit knowledge is unlikely to be like any other asset protection. The

difficulty with respect to protecting tacit knowledge is that it is largely unobservable.

This unobservability presents any research and development institution with an interesting dilemma. R&D organizations can make a strategic choice to either codify that knowledge into manuals, software programs and mentoring plans or allow the knowledge to stay tacit - embodied in their human employees. The problem with the codification option is that it is easy for firms to lose “trade secret” protection if information security practice are not strong enough and the information is leaked from within the institution. For open research institutions, the application of strong information security practices runs counter to their organizational culture and will be expensive to implement and enforce. The difficulty with allowing the knowledge to remain tacit is that employees are mobile and limiting turnover -- no matter how costly -- will become an expensive organizational priority.

This concern with turnover should also be matched with a concern about which collaboration partners’ interface with key researchers and scientists. As the transfer of tacit knowledge outside the firm may only be accomplished through repeated contact via joint work, these connections with external firms become natural foci of analytic concern. Any transfer of tacit knowledge from one individual to another is not a general phenomena experienced by everyone within a given topical field in this scenario.

Similarly, once an individual acquiring tacit knowledge goes back to his own facility or the collaboration ends, the acquisition should continue to have an impact. In fact, this impact should be magnified. Once the collaboration ends, the recipient of the knowledge should transfer what they have learned to

members of their research groups via repeated contacts on local projects. The result should be an uptick in productivity or research quality that can be observed for that local set of researchers, if not the entire research and development institution where the noise associated with the aggregation of multiple research groups is likely to obscure the impact of any such transfer. Like the initial transfer of tacit knowledge across institutional boundaries, this within-group transfer is not diffuse.

Knowledge and its Public Good Character

Following Foray (Foray 2004), knowledge developed from research funded within the national laboratory system produces an ambiguous good. This ambiguity is different from that observed with respect to knowledge generated at a university in that the knowledge produced at a national laboratory has two dimensions. In the university context, the explicit knowledge developed is released to the global public domain in the majority of cases and thus, creates a global public good. The tacit knowledge created within the university researcher as a result of their research activities is a private good for which they can appropriate future benefits.

In the national laboratory case, the “goods” classification of generated knowledge is subject to further subdivision as to what kinds of public goods these institutions generate. These entities generate national public goods (e.g. national defense and economic competitiveness) as well as global public goods (e.g. systems for monitoring epidemics). Given that all (or the vast majority) of funding at most national laboratories is their nation’s public through the governmental instrument, it can be argued that this balance should favor the

development of national public goods over global public goods to avoid the inefficiencies associated with free-riding agents. As the open science model is linked to the generation of global public goods, that organization structure should be present within these institutions, but the dominant model driving decision-making within these institutions should be national welfare driven.

The dominant decision making model driving these research institutions appears to be an open science model. Facilities built for national defense purposes are being integrated into a larger global research program with little understanding how that integration benefits the national welfare of the United States. For example, many of Los Alamos National Laboratory's unique physics facilities built to develop the kind of scientific understanding that would allow for confidence in nuclear stockpile reliability are billed as high energy physics to be part of a "global research infrastructure" in their talks at professional conferences.

The results presented in the preceding three chapters call into question the wisdom of US national laboratories continuing to engage in collaborations with Russian research institutes without a change in either research policymaker situational awareness or organizational knowledge management philosophy. Examination of US nuclear weapons laboratories collaborations with similar entities in Russia appear to indicate that two of these laboratories may be eroding their institutional tacit knowledge base by participating in research efforts with administratively dictated research partners undertaken to fulfill policy goals instead of optimizing national laboratory research production (and related tacit knowledge acquisition) by choosing research partners via other means, such as via preferential attachment. The results from the nuclear fusion research field

examination suggests that continued Russian engagement in nuclear fusion work at LLNL may not be prudent, given the observed negative impact on LLNL researcher productivity in the field.

The divergence of observed impact of Russian research network neighborhood participants when compared to US network neighborhood members from the theoretically expected relationship suggests that in some cases US tacit knowledge is transferred in the context of these interactions while gaining little in return. This transfer of tacit knowledge through repeated interaction should yield asymmetric benefits to the Russian collaboration participants in the form of increased publication production and higher research quality. One can even make the observation that US national laboratories have played a critical role in reviving the Russian scientific establishment's ties with the leading international scientific community in the wake of the Cold War. If the results of Trajtenberg with respect to citation weighted patents can be applied to citation-weighted research publications (Trajtenberg 1990), it would appear that these collaborations are increasing the net impact of the Russian scientific research program as well.

This increase in the net impact of the Russian scientific program as a result of interactions with US national laboratory researchers is not good news for US innovation policy in either the national security sphere or the sphere of economic competitiveness. National innovation policy in both of these arenas is based on the idea of innate US scientific advantage such that the US can compete its adversaries into obsolescence. While it is debatable if policy based on this assumption is sound from a net social benefits perspective, it is clear that any action that persistently reduces this gap damages the applicability of this policy

primitive and diminishes any benefits from actions that are falsely premised on this assumption.

Collaboration Choices

As indicated in the previous chapter, there are certain kinds of collaboration links that are more likely to lead to benefits for the national laboratory engaging in them. All such linkages have either a neutral or positive knowledge gradient when viewed from the perspective of the U.S. national laboratory engaging in the collaboration pair. For example, linkages between individuals who are themselves highly connected are the ties that dictate the future direction of research programs in technical fields. Likewise, linkages between junior US researchers should be encouraged to form collaborative links with more senior researchers from other nation's national laboratories.

Similarly, there are collaborations that tend only to result in costs for US national laboratories. Senior US researcher to junior Russian researcher linkages would appear to disproportionately benefit the Russian participant via the transfer of tacit knowledge from the senior US researcher. If a junior Russian participant is actually performing work that could be performed in the United States for less cost by a future national laboratory employee, there is also a dynamic cost to this relationship in the form of an erosion of the pool of young researchers capable of performing advanced research in a real world setting.

Likewise, linkages to scientists who marginally participate in the conduct of a large high energy physics collaborations - acting primarily as country "observers" - are depleting as well. These individuals typically contribute little to the larger collaboration as many of them have only shallow publication

records in the technical area and are primarily sitting on these collaborations as science and technology monitors for their countries - who often only supply token financial support for the research being performed. Outside of the problem associated with the acquisition of cutting edge knowledge with only minimal investment, the issue here is that marginal participation gives these “observers” access to the breadth of world class researchers in a given field - vastly simplifying the acquisition of science and technology information paid for by the scientific establishments of other countries. Once this researcher access is established, it becomes a simpler matter to acquire tacit knowledge from individuals with unique insights and experiences by establishing repeated contacts with them.

Improving Appropriability

In light of the preceding observations, it would be logical to deduce that improving appropriability demands simply choosing collaborations that are likely to produce tacit knowledge benefits over those that should result in loss. If these thoughtful choices are aggregated, it may very well be the case that national laboratories may grow their research competencies through international collaboration rather than diminish them. There are other alternatives, however.

Oddly enough, one of the best ways to assure the national laboratories gain more benefits from collaborations is to attract true technical collaboration on issues of clear atomistic research interest to individuals within the laboratories through the possession of open research facilities with unique tools. As indicated previously in this examination, national laboratories have been able to invest in

unique facilities that are beyond the reach of most universities and corporations. It is these facilities that can become the attractive force laboratories need to attract unmanaged collaboration from leading researchers.

There are clear benefits to this strategy for national laboratories in becoming magnets for innovative research in emerging technical areas. National laboratory researchers running the unique equipment will gain the benefit of the broadest exposure to researchers outside the laboratory using a diversity of approaches to solve technical problems. This broad exposure to diverse researchers will constantly refresh the technical competency of the resident employees working at these unique facilities, keeping them at the cutting edge of work in the emergent technical area. It is clear such a strategy minimizes the cost of human capital maintenance - an important factor for national laboratories that attempt to hold a broad range of unique equities in an equally broad set of technical pursuits.

There is a dynamic benefit to this process as well. Once attracted to the facilities, researchers from other locations will have an incentive to stay in the area, contract to the laboratory, and potentially, even become permanent members of the national laboratory workforce. Similarly, students with an interest in an emergent technical field will be more likely to attend a university either in close proximity to the unique facilities that are needed to perform advanced studies or with a university that possesses close ties to the national laboratory that runs these facilities. As they graduate from their studies, these individuals will become a workforce that is ready to go to work in the national laboratories as technologists, if not leading researchers in their own right. The clear benefit of each of these situations is that the national laboratory running the

unique facility will have ready access to a replacement labor force as its own laboratory employees in the technical area turn over.

If implemented appropriately, this strategy would actually assist the national laboratories in their conduct of work in the support of national interests. If the same cadre of employees worked at both a unique nanotechnology research facility and a national security oriented micro-systems facility, it is clear the employees would bring their new knowledge to their national security work. Thus, the national security oriented work of the laboratories would benefit from the acquisition of the latest techniques being used in the nanotechnology research facility.

Appendix 1: Pre-processing codes

Post-extraction processing of the publication record corpus from Thomson ISI Web of Knowledge sources was performed using the following Perl and UNIX shell scripts. This set of scripts was used to reformat often non-standard Web of Knowledge publication records to a fielded, single line per record format.

The line-record format is most suitable for flat-file manipulation. The flat-file record storage method was selected over database storage due to its superior replicability for other researchers, the lack of a speed premium in extraction and the linear scalability of flat files as the numbers of records increase. What the choice of flat-files gives up in extraction speed it more than make up for in customizability. Fielded line-records in such files are easily manipulated using simple UNIX utilities like `grep`, `wc -l`, and `sort` and are less subject to variance in SQL implementation across database types.

As documented in Chapter 4, duplicate records are a significant problem in dealing with publication records. Whereas such records often can persist in database formats due to the presence of unique keys, such duplicates can be easily washed out of flat-files through the judicious use of `sort` and `uniq`.

Discerning researcher clusters among large sets of coauthor relationships is easier to conduct with a visually oriented approach than with some sort of text-search algorithm. However, any attempt at visualization requires reformatting the data into a format that enables network representation. In this case, the network representation format involved breaking author data down into collaboration pairs and the frequency of their occurrence in the corpus.

Each of the Unix shell scripts contained in this section was developed first in the Mac OS X (10.5.6) environment for use in the Mac implementation of BSD UNIX in Darwin 9.6.0. All scripts were later ported for use on a server using Red Hat Linux. The Perl scripts were built in or updated to version 5.8.8.

Reformatting Scripts

Wrapper script:

WoS_process.pl

```
grep -vf ~/WoS_filter ~/Downloads/$1 | tr 'a-z' 'A-Z' > WoS_1
~/WoS_1.pl < WoS_1 > WoS_2
~/WoS_2.pl < WoS_2 > WoS_3
grep \ | AF\ | WoS_3 | cut -d"|" -f1,2,5,6,7,8,9,10,11,12,13 > WoS_4
grep -v \ | AF\ | WoS_3 | cut -d"|" -f1,2,3,4,5,6,7,8,9,10,11 | grep -v ^$ > WoS_5
cat WoS_4 WoS_5 > WoS_fin
rm WoS_?
```

Subroutines:

WoS_1.pl

```
#!/usr/bin/perl -w
while ($_=<STDIN>) {
s/^AU/\ | AU\ | /;
s/^AF/\ | AF\ | /;
s/\n/\;\n/;
s/^TI/\ | TI\ | /;
s/^SO/\ | SO\ | /;
s/^PD/\ | PD\ | /;
```

```

s/^PY/\ |PY\ | /;
chomp;
print;
}
WoS_2.pl
#!/usr/bin/perl -w
while ($_=<STDIN>) {
s/\;\ \ \ / \;/g;
s/\ \ \ / \ /g;
s/, \ / _/g;
s/\;\ \ / \ /g;
s/\ \ |AU\ \ | \nAU\ \ | /g;
print;
}

```

Visualization Enabling Scripts

To visualize the networks represented in the text of these files, co-authorship relationships must be extracted from the corpus of publication records and put into a form that visualization software will recognize. The primary goal of the following scripts is to develop a list of all collaboration pairs represented in the corpus. These scripts are used to format data from Thomson ISI Web of Science for processing via Gephi. The pseudocode for the layout algorithm used in generating the graphs in Chapter 3 is included for completeness.

The latent semantic analysis visualizations in this document may be reproduced by processing institutional publication sets through the reformatting scripts in the earlier subsections. These publication sets must then be ingested into LDRDView. To create a visualized latent semantic graph from these institutional datasets, one must first decide how to cluster the records. For topical clustering, record clustering must be done by title. For enabling technology clustering, the clustering must be done by abstract. Choosing fewer dimensions will develop cleaner graphs. The absolute ceiling on document set records is about 35,000 records. Graph creation at the upper end of this document set can require significant time periods (36 to 48 hours). Animations may be produced by developing a screen shot based movie in QuickTime while examining higher dimensional representations.

Adjacency map development:

WoS_map

control part

```
cut -d"|" -f2 $1 > list
```

```
~/WoS_cx_1.pl < list > intermed
```

```
sort -u intermed | grep -vf map.filter > intermed_1
```

dynamic part

```
echo "for AUTONE in `~/WoS_3.pl < intermed_1`" > dynamap.$1
```

```
echo "do" >> dynamap.$1
```

```
echo "for AUTTWO in `~/WoS_3.pl < intermed_1`" >> dynamap.$1
```

```
echo "do" >> dynamap.$1
```

```
echo 'PAIR=`grep ${AUTONE} list | grep ${AUTTWO} | wc -l`' >> dynamap.$1
```



```

echo 'if [ ${PAIR} -gt 0 ]; then' >> dynamap.$1
echo 'echo "${AUTONE}|${AUTTWO}|${PAIR}|$2"' >> author.map' >>
dynamap.$1
echo "fi" >> dynamap.$1
echo "done" >> dynamap.$1
echo "done" >> dynamap.$1
chmod 777 ~/dynamap.$1
~/dynamap.$1
WoS_3.pl
#!/usr/bin/perl -w
while($_=<STDIN>) {
    chomp;
    print "$_ ";
}
WoS_cx_1.pl
#!/usr/bin/perl -w
while ($_=<STDIN>) {
    s/\\. \ / /g;
    s/\\. $ / /;
    s/\\;/\\n/g;
    s/^$ / /;
    print;
}

```

Fruchterman-Reingold layout algorithm (Fruchterman and Reingold, 1991)

```

area := W * L; { W and L are the width and length of the frame }
G := (V, E); { the vertices are assigned random initial positions }
k :=  $\sqrt{\text{area}/|V|}$ ;
function fe(z) := begin return x2/k end;
function fr(z) := begin return k2/z end;

for i := 1 to iterations do begin
  { calculate repulsive forces }
  for v in V do begin
    { each vertex has two vectors: .pos and .disp }
    v.disp := 0;
    for u in V do
      if (u # v) then begin
        { Δ is short hand for the difference }
        { vector between the positions of the two vertices }
        Δ := v.pos - u.pos;
        v.disp := v.disp + ( Δ / | Δ | ) * fr(| Δ |)
      end
    end
  end

  { calculate attractive forces }
  for e in E do begin
    { each edge is an ordered pair of vertices .v and .u }
    Δ := e.v.pos - e.u.pos
    e.v.disp := e.v.disp - ( Δ / | Δ | ) * fe(| Δ |);
    e.u.disp := e.u.disp + ( Δ / | Δ | ) * fe(| Δ |)
  end

  { limit the maximum displacement to the temperature t }
  { and then prevent from being displaced outside frame }
  for v in V do begin
    v.pos := v.pos + ( v.disp / |v.disp| ) * min ( v.disp, t );
    v.pos.x := min(W/2, max(-W/2, v.pos.x));
    v.pos.y := min(L/2, max(-L/2, v.pos.y))
  end

  { reduce the temperature as the layout approaches a better configuration }
  t := cool(t)
end
end

```

Automated Panel Dataset Construction

take authors from \$1

cut -d"|" -f2 \$1 > list

~/WoS_cx_1.pl < list > intermed

sort -u intermed | grep -vf map.filter > intermed_1

```

# develop dynamic file
echo "for AUTHORA in `~/WoS_3.pl < intermed_1`" > joint_prod.$1
echo "do" >> joint_prod.$1
echo 'ONYX=`grep ${AUTHORA} WoS_fin_tot_$2 | wc -l | sed \'\'s/^[
\t]*//\'\'` >> joint_prod.$1
echo 'wait' >> joint_prod.$1
echo 'AGATE=`grep ${AUTHORA} WoS_fin_tot_$3 | wc -l | sed \'\'s/^[
\t]*//\'\'` >> joint_prod.$1
echo 'wait' >> joint_prod.$1
echo 'if [ ${ONYX} -gt ${AGATE} ]; then' >> joint_prod.$1
echo 'INSTITUTA=$2' >> joint_prod.$1
echo "else" >> joint_prod.$1
echo 'INSTITUTA=$3' >> joint_prod.$1
echo "fi" >> joint_prod.$1
echo 'if [ "${INSTITUTA}" == "SNL" ]; then' >> joint_prod.$1
echo "CTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "LANL" ]; then' >> joint_prod.$1
echo "CTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "LLNL" ]; then' >> joint_prod.$1
echo "CTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "DVEI" ]; then' >> joint_prod.$1
echo "CTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "KIAE" ]; then' >> joint_prod.$1
echo "CTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "PLPI" ]; then' >> joint_prod.$1

```

```

echo "CTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "VNIIEF" ]; then' >> joint_prod.$1
echo "CTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "VNIITF" ]; then' >> joint_prod.$1
echo "CTRY=0" >> joint_prod.$1
echo "fi" >> joint_prod.$1
echo "for AUTHORB in `~/WoS_3.pl < intermed_1`" >> joint_prod.$1
echo "do" >> joint_prod.$1
echo 'ORYX=`grep ${AUTHORA} WoS_fin_tot_$2 | wc -l | sed \'\'s/^\[
\t]*//\'\'` >> joint_prod.$1
echo 'AGARE=`grep ${AUTHORA} WoS_fin_tot_$3 | wc -l | sed \'\'s/^\[
\t]*//\'\'` >> joint_prod.$1
echo 'if [ ${ORYX} -gt ${AGARE} ]; then' >> joint_prod.$1
echo 'INSTITUTB=$2' >> joint_prod.$1
echo "else" >> joint_prod.$1
echo 'INSTITUTB=$3' >> joint_prod.$1
echo "fi" >> joint_prod.$1
echo 'if [ "${INSTITUTB}" == "SNL" ]; then' >> joint_prod.$1
echo "BCTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTB}" == "LANL" ]; then' >> joint_prod.$1
echo "BCTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTB}" == "LLNL" ]; then' >> joint_prod.$1
echo "BCTRY=1" >> joint_prod.$1
echo 'elif [ "${INSTITUTB}" == "DVEI" ]; then' >> joint_prod.$1
echo "BCTRY=0" >> joint_prod.$1

```

```

echo 'elif [ "${INSTITUTB}" == "KIAE" ]; then' >> joint_prod.$1
echo "BCTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTB}" == "PLPI" ]; then' >> joint_prod.$1
echo "BCTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTAB}" == "VNIIEF" ]; then' >> joint_prod.$1
echo "BCTRY=0" >> joint_prod.$1
echo 'elif [ "${INSTITUTB}" == "VNIITF" ]; then' >> joint_prod.$1
echo "BCTRY=0" >> joint_prod.$1
echo "fi" >> joint_prod.$1
echo 'rm ${AUTHORA}_${INSTITUTA}' >> joint_prod.$1
echo "for YEAR in 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
2005 2006 2007 2008 2009 2010 2011 2012" >> joint_prod.$1
echo "do" >> joint_prod.$1
echo 'grep PY\ | ${YEAR}\; WoS_fin_tot_${INSTITUTA} | grep ${AUTHORA}
>> ${AUTHORA}_${INSTITUTA}' >> joint_prod.$1
echo 'grep PY\ | ${YEAR}\; WoS_fin_tot_${INSTITUTA} | grep ${AUTHORB} >>
${AUTHORB}_${INSTITUTA}' >> joint_prod.$1
echo 'JPUBS=`grep PY\ | ${YEAR}\; ${AUTHORA}_${INSTITUTA} | grep
${AUTHORB} | wc -l | sed \"s/^[\t]*/ /\ \"` >> joint_prod.$1
echo "wait" >> joint_prod.$1
echo 'grep PY\ | ${YEAR}\; ${AUTHORA}_${INSTITUTA} | grep ${AUTHORB}
| cut -d"|" -f12 | cut -d";" -f1 > ${AUTHORA}_jcite' >> joint_prod.$1
# echo 'JCIT=`wc -l ${AUTHORA}_jcite`' >> joint_prod.$1
# echo 'if [ "${JCIT}" -gt "0" ]; then' >> joint_prod.$1

```

```

# echo 'JCITES=`cat ${AUTHORA}_jcite | awk '\{SUM+=\$1} END {print
SUM}\`"' >> joint_prod.$1

# echo 'else' >> joint_prod.$1

# echo 'JCITES=0' >> joint_prod.$1

# echo 'fi' >> joint_prod.$1

# echo "wait" >> joint_prod.$1

echo 'APUBX=`wc -l ${AUTHORA}_${INSTITUTA} | sed '\`s/^[ \t]*//\' | cut
-d" " -f1`' >> joint_prod.$1

echo "wait" >> joint_prod.$1

echo 'BPUBX=`wc -l ${AUTHORB}_${INSTITUTA} | sed '\`s/^[ \t]*//\' | cut -
d" " -f1`' >> joint_prod.$1

echo "wait" >> joint_prod.$1

echo 'grep PY\ | ${YEAR}\; ${AUTHORA}_${INSTITUTA} | cut -d"|" -f12 | cut -
d";" -f1 > ${AUTHORA}_cite' >> joint_prod.$1

# echo 'ACIT=`wc -l ${AUTHORA}_cite`' >> joint_prod.$1

# echo 'if [ "${ACIT}" -gt "0" ]; then' >> joint_prod.$1

# echo 'ACITX=`cat ${AUTHORA}_cite | awk '\{SUM+=\$1} END {print
SUM}\`"' >> joint_prod.$1

# echo 'else' >> joint_prod.$1

# echo 'ACITX=0' >> joint_prod.$1

# echo 'fi' >> joint_prod.$1

echo "wait" >> joint_prod.$1

echo 'grep PY\ | ${YEAR}\; ${AUTHORB}_${INSTITUTA} | cut -d"|" -f12 | cut -
d";" -f1 > ${AUTHORB}_cite' >> joint_prod.$1

# echo 'BCIT=`wc -l ${AUTHORB}_cite`' >> joint_prod.$1

```

```

# echo 'if [ "${BCIT}" -gt "0" ]; then' >> joint_prod.$1

# echo 'BCITX=`cat ${AUTHORA}_cite | awk '\'{SUM+=\ $1} END {print
SUM}\`"' >> joint_prod.$1

# echo 'else' >> joint_prod.$1

# echo 'BCITX=0' >> joint_prod.$1

# echo 'fi' >> joint_prod.$1

# echo "wait" >> joint_prod.$1

# temporary collaboration

echo 'YEARIP=`expr ${YEAR} + 1`' >> joint_prod.$1
echo 'YEARIIP=`expr ${YEAR} + 2`' >> joint_prod.$1
echo 'YEARMIP=`expr ${YEAR} - 1`' >> joint_prod.$1
echo 'YEARMIIP=`expr ${YEAR} - 2`' >> joint_prod.$1

echo 'grep PY\ | ${YEARIP}\; WoS_fin_tot_${INSTITUTA} | grep ${AUTHORA}
> ${AUTHORA}_${INSTITUTA}.p1' >> joint_prod.$1

echo 'grep PY\ | ${YEARIIP}\; WoS_fin_tot_${INSTITUTA} | grep ${AUTHORA}
> ${AUTHORA}_${INSTITUTA}.p2' >> joint_prod.$1

echo 'JPUBII=`grep PY\ | ${YEARIIP}\; ${AUTHORA}_${INSTITUTA}.p1 | grep
${AUTHORB} | wc -l | sed '\`s/^[\ t]*/ /\`"' >> joint_prod.$1

echo 'JPUBI=`grep PY\ | ${YEARIP}\; ${AUTHORA}_${INSTITUTA}.p2 | grep
${AUTHORB} | wc -l | sed '\`s/^[\ t]*/ /\`"' >> joint_prod.$1

echo 'JPUBMII=`grep PY\ | ${YEARMIP}\; ${AUTHORA}_${INSTITUTA} | grep
${AUTHORB} | wc -l | sed '\`s/^[\ t]*/ /\`"' >> joint_prod.$1

echo 'JPUBMI=`grep PY\ | ${YEARMIP}\; ${AUTHORA}_${INSTITUTA} | grep
${AUTHORB} | wc -l | sed '\`s/^[\ t]*/ /\`"' >> joint_prod.$1

```

```

echo 'if [ "${JPUBII}" -gt "0" ] || [ "${JPUBI}" -gt "0" ] || [ "${JPUBS}" -gt "0" ] || [
"${JPUBMI}" -gt "0" ] || [ "${JPUBMII}" -gt "0" ]; then' >> joint_prod.$1
echo 'TCOLLAB=1' >> joint_prod.$1
echo "else" >> joint_prod.$1
echo 'TCOLLAB=0' >> joint_prod.$1
echo "fi" >> joint_prod.$1
# network neighborhood
echo 'echo ${YEAR} > windows.txt' >> joint_prod.$1
echo 'echo ${YEARMIIP} >> windows.txt' >> joint_prod.$1
echo 'echo ${YEARMIP} >> windows.txt' >> joint_prod.$1
echo 'echo ${YEARIP} >> windows.txt' >> joint_prod.$1
echo 'echo ${YEARIIP} >> windows.txt' >> joint_prod.$1
echo 'cat ${AUTHORA}_${INSTITUTA} ${AUTHORA}_${INSTITUTA}.p1
${AUTHORA}_${INSTITUTA}.p2 > ${AUTHORA}_${INSTITUTA}_super' >>
joint_prod.$1
echo 'cut -d"|" -f2,14 ${AUTHORA}_${INSTITUTA}_super > WoS_fin_s1' >>
joint_prod.$1
echo 'grep -f windows.txt WoS_fin_s1 | cut -d"|" -f1 > coaut_t' >> joint_prod.$1
echo '~ / WoS_cx_1.pl < coaut_t > authors_t1' >> joint_prod.$1
echo 'grep -v ${AUTHORA} authors_t1 > authors_t' >> joint_prod.$1
echo 'ATCONNEX=`grep -v ${AUTHORA} authors_t | sort -u | wc -l | sed
\'s/^[\t]*//\' >> joint_prod.$1
# echo 'sed "s/^/$2\|/" authors_t > authors_g1.${YEAR}' >> joint_prod.$1
# echo 'sort -t"|" -k2 authors_g1.${YEAR} | sort -u > authors_g.${YEAR}' >>
joint_prod.$1

```



```

echo 'RUSCONNEX=`grep -f authors_t rus_authors | sort -u | wc -l | sed
\'s/^[\t]*//\' >> joint_prod.$1
echo 'USCONNEX=`grep -f authors_t us_authors | sort -u | wc -l | sed \'s/^[\
\t]*//\' >> joint_prod.$1
#institutional fields
echo 'if [ "${INSTITUTA}" == "SNL" ]; then' >> joint_prod.$1
echo 'INSTIT=SANDIA_record_SNL' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "LANL" ]; then' >> joint_prod.$1
echo 'INSTIT=ALAMOS_record_LANL' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "LLNL" ]; then' >> joint_prod.$1
echo 'INSTIT=LAWRENCE_record_LLNL' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "DVEI" ]; then' >> joint_prod.$1
echo 'INSTIT=EFREMOV_record_DVEI' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "KIAE" ]; then' >> joint_prod.$1
echo 'INSTIT=KURCHATOV_record_KIAE' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "PLPI" ]; then' >> joint_prod.$1
echo 'INSTIT=LEBEDEV_record_PLPI' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "VNIIEF" ]; then' >> joint_prod.$1
echo 'INSTIT=SAROV_record_VNIIEF' >> joint_prod.$1
echo 'elif [ "${INSTITUTA}" == "VNIITF" ]; then' >> joint_prod.$1
echo 'INSTIT=SNEZHINSK_record_VNIITF' >> joint_prod.$1
echo "fi" >> joint_prod.$1
echo 'OPUBS=`grep ^${YEAR}\ | ${INSTIT} | cut -d"|" -f3` >> joint_prod.$1
echo 'OLBR=`grep ^${YEAR}\ | ${INSTIT} | cut -d"|" -f4` >> joint_prod.$1
echo 'COINST=`grep ^${YEAR}\ | ${INSTIT} | cut -d"|" -f5` >> joint_prod.$1

```

```

# hot topics

echo 'APFO=`grep ${YEAR}$ qgp_central.csv | grep ${AUTHORA} | wc -l | sed
\'s/^[\t]*//\'` >> joint_prod.$1

echo 'if [ ${APFO} -gt 0 ]; then' >> joint_prod.$1

echo 'AQGP=1' >> joint_prod.$1

echo 'else' >> joint_prod.$1

echo 'AQGP=0' >> joint_prod.$1

echo "fi" >> joint_prod.$1

echo "wait" >> joint_prod.$1

echo 'if [ "${AQGP}" -eq "1" ]; then' >> joint_prod.$1

echo 'AECQGP=`grep ${YEAR}$ qgp_central.csv | grep ${AUTHORA} | cut -d","
-f2` >> joint_prod.$1

echo "else" >> joint_prod.$1

echo 'AECQGP=0' >> joint_prod.$1

echo "fi" >> joint_prod.$1

echo 'ANFO=`grep ${YEAR}$ nf_central.csv | grep ${AUTHORA} | wc -l | sed
\'s/^[\t]*//\'` >> joint_prod.$1

echo 'if [ ${ANFO} -gt 0 ]; then' >> joint_prod.$1

echo 'ANFU=1' >> joint_prod.$1

echo 'else' >> joint_prod.$1

echo 'ANFU=0' >> joint_prod.$1

echo "fi" >> joint_prod.$1

echo 'if [ "${ANFU}" -eq "1" ]; then' >> joint_prod.$1

echo 'AECNF=`grep ${YEAR}$ nf_central.csv | grep ${AUTHORA} | cut -d"," -
f2` >> joint_prod.$1

```

```

echo "else" >> joint_prod.$1
echo 'AECNF=0' >> joint_prod.$1
echo "fi" >> joint_prod.$1
# institutional connectivity
echo 'AECINST=`grep ${AUTHORA}
~/Downloads/Coding_resources/${INSTITUTA}_${YEAR}_ec.csv | cut -d"," -
f3`' >> joint_prod.$1
echo 'echo "${AUTHORA}-${AUTHORB}|${CTRY}-${BCTRY}|${INSTITUTA}-
${INSTITUTB}|${AUTHORA}|${INSTITUTA}|${YEAR}|${JPUBS}|${JCITES}|$
{APUBX}|${ACITX}|${CTRY}|${BCITX}|${BPUBX}|${TCOLLAB}|${ATCONN
EX}|${RUSCONNEX}|${USCONNEX}|${OPUBS}|${OLBR}|${COINST}|${AQ
GP}|${AECQGP}|${ANFU}|${AECNF}|${AECINST}" >> $2-$3_dyad_record'
>> joint_prod.$1
echo "done" >> joint_prod.$1
echo "done" >> joint_prod.$1
echo "done" >> joint_prod.$1
chmod 777 ~/joint_prod.$1
~/joint_prod.$1 $1 $2 $3

```

Appendix 2: Stata do.file

This script displayed below can be used in conjunction with the CtR.dta file to replicate the empirical results tables in Chapter 4 in Stata 13. Note that running this .do file requires the installation of multiple user contributed Stata programs that are not apart of the standard Stata 13 program distribution. This set includes the ivreg2 routine for exploring variable endogeneity, qic for implementing the quasiliikelihood information criterion for GEE model selection, xtserial routine for implementing the Wooldridge serial correlation test, and xttest3 for assessing groupwise heteroscedasticity via a modified Wald test.

```
clear
```

```
set more off
```

```
use "/Users/*****/Downloads/CtR.dta", clear
```

```
xtset dyad_num time
```

```
log using codebook.txt, text
```

```
codebook reln dyad_authors dyad_num dyad_reln_num dyad_type instit_link
```

```
Internal_US auth1 institut_1 institut_2 jpubs jcites spubx1 spubx1sq tspubx2
```

```
xcitx1 citx1 citx2 tscitx2 tcollab ctry neighborhood Russian US instit_labor
```

```
instit_pubs efficiency efficiencysq qgp1 ecqgp1 nfus1 ecnfus1 ecinst1 time
```

```
timesq year, compact
```

```
log close
```

```
* test independent variable for stationarity
```

xtunitroot fisher jpubs, dfuller lags(1)

*** pooled US model**

*** prepare for likelihood ratio test for overdispersion**

**xtpoisson jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, re
estimates store Pooled_XTP**

**xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, re
estimates store Pooled_XTN**

**xtpoisson jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, re
estimates store Pooled_XTP3**

**xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, re
estimates store Pooled_XTN3**

**xtpoisson jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, re
estimates store Pooled_XTP4**

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, re  
estimates store Pooled_XTN4
```

```
xtpoisson jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, re  
estimates store Pooled_XTP4
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, re  
estimates store Pooled_XTN4
```

* prepare for Hausman specification test w.r.t. fixed/random effects (pooled,
LLNL, SNL, LANL)

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, re  
estimates store Pooled_XTNBR
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, fe difficult  
nonrtolerance  
estimates store Pooled_XTNBF
```

```
hausman Pooled_XTNBF Pooled_XTNBR
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, re  
estimates store Pooled_XTNBR1
```

```
xtnbreg jpubs jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency  
efficiencysq ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if  
institut_1==4, fe difficult nonrtolerance  
estimates store Pooled_XTNBF1
```

```
hausman Pooled_XTNBF1 Pooled_XTNBR1
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, re  
estimates store Pooled_XTNBR2
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, fe  
difficult nonrtolerance  
estimates store Pooled_XTNBF2
```

```
hausman Pooled_XTNBF2 Pooled_XTNBR2
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, re
```

estimates store Pooled_XTNBR3

xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, fe
difficult nonrtolerance
estimates store Pooled_XTNBF3

hausman Pooled_XTNBF3 Pooled_XTNBR3

* test for serial correlation

xtserial jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 nfus1 qgp1 time timesq if ctry==1
xtserial jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 nfus1 qgp1 time timesq if institut_1==4
xtserial jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 nfus1 qgp1 time timesq if institut_1==6
xtserial jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 nfus1 qgp1 time timesq if institut_1==3

* modified Wald groupwise heteroscedasticity test (Reyna 2007)

xtreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, fe
xttest3


```
xtreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, fe
xttest3
```

```
xtreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, fe
xttest3
```

```
xtreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, fe
xttest3
```

* Hausman specification test for endogeneity

```
ivreg2 jpubs lRussia lUS spubx1 spubx1sq efficiency efficiencysq ecinst1
nfus1 qgp1 time timesq (tspubx2= tcollab tscitx2 time) if ctry==1,
endog(tspubx2)
```

```
ivreg2 jpubs lRussia lUS spubx1 spubx1sq efficiency efficiencysq ecinst1
nfus1 qgp1 time timesq (tspubx2= tcollab tscitx2 time) if institut_1==4,
endog(tspubx2)
```

```
ivreg2 jpubs IRussia IUS spubx1 spubx1sq efficiency efficiencysq ecinst1
nfus1 qgp1 time timesq (tspubx2= tcollab tscitx2 time) if institut_1==6,
endog(tspubx2)
```

```
ivreg2 jpubs IRussia IUS spubx1 spubx1sq efficiency efficiencysq ecinst1
nfus1 qgp1 time timesq (tspubx2= tcollab tscitx2 time) if institut_1==3,
endog(tspubx2)
```

```
ivreg2 jpubs Russian US spubx1 spubx1sq tspubx2 ecinst1 efficiency nfus1
qgp1 (spubx2= tcollab tscitx2 time) if ctry==1, endog(spubx2)
cluster(dyad_num)
```

```
ivreg2 jpubs Russian US spubx1 spubx1sq tspubx2 ecinst1 efficiency nfus1
qgp1 (spubx2= tcollab tscitx2 time) if institut_1==4, endog(spubx2)
cluster(dyad_num)
```

```
ivreg2 jpubs Russian US spubx1 spubx1sq tspubx2 ecinst1 efficiency nfus1
qgp1 (spubx2= tcollab tscitx2 time) if institut_1==6, endog(spubx2)
cluster(dyad_num)
```

```
ivreg2 jpubs Russian US spubx1 spubx1sq tspubx2 ecinst1 efficiency nfus1
qgp1 (spubx2= tcollab tscitx2 time) if institut_1==3, endog(spubx2)
cluster(dyad_num)
```

* Table 3: Pooled US models

```

xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==0, pa corr (ar 1)
vce(robust)

```

```

estimates store Pooled_GEE

```

```

* qic jpubs tspubx2 spubx1 spubx1sq efficiency ecinst1 Russian US time
ecnfus1 qgp1 if nfus1==1 & institut_1==3, family(nbinomial) link(log) force
corr(ar 1)

```

```

xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, pa corr
(ar 1) vce(robust)

```

```

estimates store SNL_GEE2

```

```

xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, pa corr
(ar 1) vce(robust)

```

```

estimates store LLNL_GEE2

```

```

xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, pa corr
(ar 1) vce(robust)

```

```

estimates store LANL_GEE

```

```
esttab Pooled_GEE SNL_GEE2 LLNL_GEE2 LANL_GEE using
US_institutions_results.csv, title("Table GEE Models: US, LLNL, SNL and
LANL Models") wide mtitles
```

* Table 4: Sensitivity Analysis

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2_p efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, fe difficult
nonrtolerance
estimates store Pooled_NBREG
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, pa corr (ar 1)
vce(robust)
estimates store Pooled_GEE
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2_p efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, pa corr (ar 1)
vce(robust)
estimates store Pooled_IV_GEE
```

```
esttab Pooled_NBREG Pooled_GEE Pooled_IV_GEE using
US_sensitivity_results.csv, title("Table Pooled US Model: IV(1), GEE, GEE(IV)
Models") wide mtitles
```

* Table 5: Reported "IV" GEE

```
xtnbreg tpubx2 lRussia lUS spubx1 spubx1sq efficiency efficiencysq ecinst1  
i.institut_1 i.institut_2 nfus1 qgp1 time timesq tscitx2 tcollab, fe difficult  
nonrtolerance
```

```
predict tpubx2_p
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tpubx2_p efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==1, pa corr (ar 1)  
vce(robust)
```

```
estimates store Pooled_IV_GEE
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tpubx2_p efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==6, pa corr  
(ar 1) vce(robust)
```

```
estimates store SNL_IV_GEE2
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tpubx2_p efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==4, pa corr  
(ar 1) vce(robust)
```

```
estimates store LLNL_IV_GEE2
```

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tpubx2_p efficiency efficiencysq  
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==3, pa corr  
(ar 1) vce(robust)
```

estimates store LANL_IV_GEE

**esttab Pooled_IV_GEE SNL_IV_GEE2 LLNL_IV_GEE2 LANL_IV_GEE using
US_institutions_IV_GEE_results.csv, title("Table "IV" GEE Models: US, LLNL,
SNL and LANL Models") wide mtitles**

*** Table 6: Pooled Russian model**

**xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if ctry==0, pa corr (ar 1)
vce(robust)**

estimates store Pooled_Russia_GEE

**xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==2, pa corr
(ar 1) vce(robust)**

estimates store KIAE_GEE

**xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==7, pa corr
(ar 1) vce(robust)**

estimates store VNIIEF_GEE

```
xtnbreg jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if institut_1==8, pa corr
(ar 1) vce(robust)

estimates store VNIITF_GEE
```

```
esttab Pooled_Russia_GEE KIAE_GEE VNIIEF_GEE VNIITF_GEE using
RS_pool_results.csv, title("Table Pooled Russia Model: Pooled, KIAE, VNIIEF,
VNIITF Models") wide mtitles
```

* Table 7: Nuclear fusion models (GEE)

```
xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 ecnfus1 qgp1 time timesq if nfus1==1 &
institut_1 ==3, family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store LANL_nfus

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
ecnfus1 qgp1 time timesq if nfus1==1 & institut_1==3, family(nbinomial)
link(log) force corr(ar 1)
```

```
xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 ecnfus1 qgp1 time timesq if nfus1==1 & institut_1==4,
family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store LLNL_nfus
```

```

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
ecnfus1 qgp1 time timesq if nfus1==1 & institut_1==4, family(nbinomial)
link(log) force corr(ar 1)

```

```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_2 ecnfus1 qgp1 time timesq if nfus1==1 & institut_1==6,
family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store SNL_nfus

```

```

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
ecnfus1 qgp1 time timesq if nfus1==1 & institut_1==6, family(nbinomial)
link(log) force corr(ar 1)

```

```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_2 ecnfus1 qgp1 time timesq if nfus1==1 & ctry==1,
family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store pooled_US_nfus

```

```

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
ecnfus1 qgp1 time timesq if nfus1==1 & ctry==1, family(nbinomial) link(log)
force corr(ar 1)

```

```

esttab pooled_US_nfus LLNL_nfus SNL_nfus LANL_nfus using
nfus_results.csv, title("Table Nuclear Fusion: Pooled, LANL, LLNL, SNL
Models") wide mtitles

```

* Table 8: Rest of fields (GEE)


```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_2 nfus1 ecqgp1 time timesq if nfus1==0 & qgp1==0 &
institut_1==3, family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store LANL_row

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
time timesq if nfus1==0 & qgp1==0 & institut_1==3, family(nbinomial)
link(log) force corr(ar 1)

```

```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_2 nfus1 ecqgp1 time timesq if nfus1==0 & qgp1==0 &
institut_1==4, family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store LLNL_row

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
time timesq if nfus1==0 & qgp1==0 & institut_1==4, family(nbinomial)
link(log) force corr(ar 1)

```

```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if nfus1==0 & qgp1==0
& institut_1==6, family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store SNL_row

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
time timesq if nfus1==0 & qgp1==0 & institut_1==6, family(nbinomial)
link(log) force corr(ar 1)

```

```

xtgee jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq
ecinst1 i.institut_1 i.institut_2 nfus1 qgp1 time timesq if nfus1==0 & qgp1==0
& ctry==1, family(nbinomial) link(log) force corr(ar 1) vce(robust) nolog
estimates store pooled_US_row

qic jpubs lRussia lUS spubx1 spubx1sq tspubx2 efficiency efficiencysq ecinst1
time timesq if nfus1==0 & qgp1==0 & ctry==1, family(nbinomial) link(log)
force corr(ar 1)

esttab pooled_US_row LLNL_row SNL_row LANL_row using
row_Pooled_results.csv, title("Table All Other Fields: Pooled, LANL, LLNL,
SNL Models") wide mtitles

```

References

- Abdou, MA, Bertolini, E., Hancox, R., Kostenko, AI., Kulcinski, GL., Miller, RL et al (1986) "Fusion Reactor Design IV: Report on the 4th IAEA Technical Committee and Workshop, Yalta, USSR, May 26-June 6, 1986" *Nuclear Fusion*, 26, no. 10 (Oct.1986), p. 1377-1428.
- Abramo, G.; D' Angelo, CA; and Caprasecca, A. (2009) "Allocative Efficiency in Public Research Funding: Can Bibliometrics Help?" *Research Policy*, 38 (2009) p. 206-215.
- Aghion, P., David, P., and Foray, D. (2009) "Science, Technology and Innovation for Economic Growth: Linking Policy Research and practice in 'STIG systems,'" *Research Policy*, 38, no.4, p. 681-693.
- American Association for the Advancement of Science (2002) "Proposed Dept. of Homeland Security Would Include NIH, DOE and USDA R&D Programs," AAAS R&D Funding Update, June 21, 2002.
- Arrow, K. (1994) "Methodological Individualism and Social Knowledge," *American Economic Review* 84, no 2 p.1-9.
- Artsimovich, L.A. (1958) "Research on Controlled Thermonuclear Reactions in the USSR," *Proceedings of the Second United Nations International Conference on the Peaceful Uses of Atomic Energy*, Geneva, September 1-13, 1958, v. 31, p. 6-21.
- Baba, Y., Shichijo, N. and Sedita, S. (2009) "How Do Collaborations with Universities Affect Firms' Innovative Performance? The Role of Pasteur Scientists in the Advanced Materials Field," *Research Policy* 38, no. 5, p. 756-764.
- Barrat, A., Barthelemy, M. and Vespignani, A. (2008) *Dynamical Processes on Complex Networks* (Cambridge University Press: New York)
- Beck, M.D., Cupitt, R.T., Gahlaut, S. and Jones, S.A. (2003) *To Supply or to Deny: Comparing Nonproliferation Export Controls in Five Key Countries* (Kluwer Law International: The Hague, Netherlands)
- de Bellis, N. (2009) *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics* (Scarecrow Press; Lanham, MD)
- Bishop, A. (1958) *Project Sherwood - The US Program in Controlled Fusion* (Addison Wesley Publishing: Reading, MA)
- Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008) "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10), p. 1000.

- Bondarenko, B.D. (2001) "Role Played by OA Lavrent'ev in the Formulation of the Problem and Initiation of Research into Controlled Nuclear Fusion in the USSR," *Uspekhi Fizicheskikh Nauk*, vol. 171, no.8, pp. 844-851.
- Borner, K. (2000) "Extracting and Visualizing Semantic Structures in Retrieval Results for Browsing," *Proceedings of the Fifth ACM Conference on Digital Libraries*, June 2-7, 2000, San Antonio, Texas, p. 234-235.
- Boyack, K. and Rahal, N. (2004) "Evaluation of Laboratory Directed Research and Development (LDRD) Investment Areas at Sandia," *Technological Forecasting & Social Change* 72, no. 9., p. 1122-1136.
- Bromberg, J.L. (1983) *Fusion: Science, Politics and the Invention of a New Energy Source* (MIT Press: Cambridge, MA)
- Bush, V. (1949) *Science: The Endless Frontier* (Ayer Company Publishers, Inc.: North Stratford, NH)
- Cabellero, R. and Jaffe, A. (1993) "How High are the Giants' Shoulders? An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in Jaffe and Trajtenberg eds. *Patents, Citations and Innovations: A Window on the Knowledge Economy* (MIT Press: Cambridge, MA) p. 89-152.
- Clery, D. (2013) *A Piece of the Sun: The Quest for Fusion Energy* (Overlook Press: New York City)
- Cockburn, I. and Griliches, Z. (1988) "Industry effects and appropriability measures in the stock market's valuation of R&D and patents," *American Economic Review* 78 (May 1988), p. 419-423.
- Cockburn, I. and Henderson, R. (1998) "Absorptive capacity, coauthoring behavior, and the organization of research during drug discovery," *Journal of Industrial Economics*, 46 no.2, p. 157-183.
- Cohen, W. and Levinthal, D. (1989) "Innovation and learning: the two faces of R&D" *Economic Journal* 99, p.569-596.
- Collins, H. M. (2010) *Tacit and Explicit Knowledge* (University of Chicago Press: Chicago)
- Commission on Maintaining United States Nuclear Weapons Expertise (1999) Report to the National Defense Authorization Acts of 1997 and 1998 (General Printing Office: Washington, DC)
- Crane, D. (1972) *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (University of Chicago Press: Chicago)

- Cui, J. (2007) "QIC Program and Model Selection in GEE Analyses," *The Stata Journal* 7, no. 2, p. 209-220.
- Dasgupta and David (1994) "Toward a new economics of science," *Research Policy* 23, no. 5, p. 487-521.
- Dean, S.O. (2013) *Search for the Ultimate Energy Source: A History of the U.S. Fusion Energy Program* (Springer: New York)
- Dumais, S.T. (2007) "LSA and Information Retrieval: Getting Back to Basics," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *"Handbook of Latent Semantic Analysis,"* (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 293 - 321.
- Escribano, A., Fosfuri, A. and Tribo, J. (2009) "Managing external knowledge flows: the moderating role of absorptive capacity," *Research Policy* 38 (2009) p. 96-105.
- Fan HY, Leve EW, Scullin C, et al. (2005) "Surfactant-assisted synthesis of water-soluble and biocompatible semiconductor quantum dot micelles," *Nano Letters*, 5 no.4, p. 645-648.
- Foray, D. (2004) *The Economics of Knowledge*. (MIT Press: Cambridge, MA).
- Fowler, T.K. (1997) *The Fusion Quest* (Johns Hopkins University Press: Baltimore)
- Freeman, R. (2006) "Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?" in A. Jaffe and J. Lerner eds. *Innovation Policy and the Economy* v.6 (National Bureau of Economic Research: Washington, DC) p. 123-158.
- Fruchterman, T.M.J. and Reingold, E.M. (1991) "Graph Drawing by Force Directed Placement," *Software: Practice and Experience*, 21(11) (November 1991), pp. 1129-1164.
- Geim, A. and Kim, P. (2008) "Carbon Wonderland," *Scientific American*, vol. 298, no. 4 (01 April 2008), pp. 90-97.
- Geisler, E. (2000) *The Metrics of Science and Technology* (Quorum Books: Westport, CT)
- General Accounting Office (1994) "National Laboratories: Are Their R&D Activities Related to Commercial Product Development?" (GAO: Washington, DC)
- Goncharov, G.A. (2001) "The 50th Anniversary of the Beginning of Research in the USSR on the Potential Creation of a Nuclear Fusion Reactor," *Uspekhi Fizicheskikh Nauk*, vol. 171, no.8, pp. 894-901.

- Goyal, S. (2007) *Connections: An Introduction to the Economics of Networks* (Princeton University Press: Princeton, NJ)
- Griliches, Z. (1984) *R&D, Patents and Productivity* (University of Chicago Press: Chicago)
- Griliches, Z., Hall, B. and Pakes, A. (1991) "R&D, Patents and Market Value Revisited: Is There a Second (Technological Opportunity) Factor?" *Economics of Innovation and New Technology* v.1, p.183-202.
- Hansen, M. (1999) "The search transfer problem: The role of weak ties in sharing knowledge across organization subunits," *Administrative Science Quarterly* 44 (March 1999) p. 82-111.
- Hendrickson, S.M.L. (2009) "The Wrong Wright Stuff: Mapping Human Error in Aviation," (Doctoral dissertation). Acquired from author.
- Hermann, R. (1990) *Fusion: The Search for Endless Energy* (Cambridge University Press: New York)
- Hilbe, J.M. (2008) *Negative Binomial Regression* (Cambridge University Press: New York)
- Hu, X., Wiemer-Hastings, P. and McNamara, D.S. (2007) "Strengths, Limitations and Extensions of LSA," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 401 - 426.
- Jackson, M.O. (2008) *Social and Economic Networks* (Princeton University Press: Princeton, NJ)
- Jaffe, A. and Trajtenberg (2002) *Patents, Citations and Innovations: A Window on the Knowledge Economy* (MIT Press: Cambridge, MA)
- Jaffe, A., Fogarty, M. and Banks, B. (1998) "Evidence from the Patents and Patent Citations on the Impact of NASA and other Federal Labs on Commercial Innovation" in Jaffe and Trajtenberg eds. *Patents, Citations and Innovations: A Window on the Knowledge Economy* (MIT Press: Cambridge, MA) p. 89-152.
- Josephson, P.R. (2000) *Red Atom: Russia's Nuclear Power Program from Stalin to Today* (University of Pittsburgh Press: Pittsburgh, PA)
- Karasev, O. and Edelkina, A. (2013) "Sectoral Foresight Studies: Future Perspectives of Nanotechnologies," in Meissner, D., Gokhberg, L. and Sokolov, A. eds (2013) *Science, Technology and Innovation Policy for the Future: Potentials and Limits of Foresight Studies*, (Springer: New York), p.121-136.

Kintsch, W. (2007) "Meaning in Context," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 89 - 105.

Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ)

Landauer, T.K. (2007) "LSA as a Theory of Meaning," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 3 - 34.

Lazonick, W. (2005) "The Innovative Firm," in Fagerberg, J., Mowery, D.C., and Nelson, R.R. eds (2005) *Oxford Handbook of Innovation*, (Oxford University Press: Oxford), p. 29-55.

Lima, M. (2011) *Visual Complexity: Mapping Patterns of Information* (Princeton Architectural Press: New York)

Lundvall, B. (1998) "The Learning Economy: Challenges to Economic Theory and Policy in K. Nielsen and B. Johnson eds. *Institutions and Economic Change: New Perspectives on Markets, Firms, and Technology* (Edward Elgar Publishers: Cheltenham, UK)

Martin, D.I. and Berry, M.W.(2007) "Mathematical Foundations Behind LSA," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 35 - 56.

McNamara, D.S., Cai, Z., and Louwerse, M.M. (2007) "Optimizing LSA Measures of Cohesion," in Landauer, T.K., McNamara, D.S., Dennis, S. and Kintsch, W. eds (2007) *Handbook of Latent Semantic Analysis*, (Lawrence Erlbaum Associates, Inc.: Mahwah, NJ), p. 379 - 400.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002) "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, vol.298 (October 25, 2002), p. 824-827.

Mirowski, P. and Sent, E.M. (2002) *Science Bought and Sold: Essays in the Economics of Science* (University of Chicago Press: Chicago)

Mosher, J., Lewis, P. and Leahy, R. (1992) "Multiple Dipole Modeling and Localization from Spatiotemporal MEG Data," *IEEE Transactions on Biomedical Engineering* 39, no.6, p.541-557

Mowery, D. and Rosenberg, N.(1998) *Paths of Innovation: Technological Change in 20th-Century America* (Cambridge University Press: New York)

- Mowery, D., Nelson, R., Sampat, B. and Ziedonis, A. (2004) *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act* (Stanford Business Books: Palo Alto, CA)
- National Intellectual Property Rights Coordination Center (2012) "2012 U.S. Intellectual Property Enforcement Coordinator Joint Strategic Plan," (General Printing Office: Washington, DC).
- National Science Board (2008) *Science and Engineering Indicators, 2008* (National Science Foundation: Arlington, VA)
- National Science Board (2012) *Science and Engineering Indicators, 2012* (National Science Foundation: Arlington, VA)
- Newman, M., Barabasi, A. and Watts, D. (2006) *The Structure and Dynamics of Networks* (Princeton University Press: Princeton, NJ).
- Newman, M. (2001) "The structure of scientific collaboration networks," *Proceedings of the National Academy of Science*, 98 no. 2 (2001) p. 404-409.
- Ono, M. (2003) "Spherical Torus Development Path," Fusion Energy Science Advisory Council (FESAC) Development Path Meeting, General Atomics, Jan. 13-14, 2003. Accessed 5/26/2009 at http://fire.pppl.gov/fesac_dp_st_ono.pdf.
- Pakes, A. (1985) "On patents, R&D and the stock market's rate of return," *Journal of Public Economy*, no.93, p. 390-409.
- Polanyi, M. (1967) *The Tacit Dimension* (Anchor Books: New York).
- Puretskiy, A.A., Shutt, G.L., and Berry, M.W. (2010) "Survey of Text Visualization Techniques," in Berry, M.W. and Kogan, J, eds. (2010) *Text Mining: Applications and Theory* (John Wiley and Sons: West Sussex, UK) p. 107-127.
- Robinson-Avila, K. (2008) "Sandia Revamps, Relaunches Entrepreneurial Leave Program," *New Mexico Business Weekly*, Dec. 12, 2008. Accessed Aug. 10, 2010 at <http://albuquerque.bizjournals.com/albuquerque/stories/2008/12/15/story9/html>.
- Romer, D. (1996) *Advanced Macroeconomics* (McGraw-Hill: San Francisco)
- The Royal Society (2011) *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century*, RIS Policy Document 03/11 (March, 2011)
- Saxenian, A. (2006) *The New Argonauts: Regional Advantage in a Global Economy* (Harvard University Press: Cambridge, MA)
- Schiermeier, Q. (2007) "Russian Science: The Battle for Russia's Brains," *Nature* vol. 449 (October 3, 2007) p. 524-527.

- Scotchmer, S. (2004) *Innovation and Incentives* (MIT Press: Cambridge, MA).
- Shafranov, V.D. (2001) "The Initial Period in the History of Nuclear Fusion Research at the Kurchatov Institute," *Uspekhi Fizicheskikh Nauk*, vol. 171, no.8, pp. 894-901.
- Small, H. (2011) "Text Mining of Citations in Emerging Scientific Communities," Elsevier/Los Alamos National Laboratory Workshop on *Mapping and Measuring Scientific Output*, Santa Fe, May 10, 2011
- Sorenson, O. and Singh, J. (2007) "Science, social networks and spillovers" in *Industry and Innovation* 14 no. 2 (May 2007) p.219-238.
- Stokes, D. (1997) *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Institution Press: Washington, DC)
- Tomlinson, M. (1999) "The Learning Economy and Embodied Knowledge Flows in Great Britain," in *Journal of Evolutionary Economics* 9, p.431-451.
- Trajtenberg, M. (1990) "A Penny for Your Quotes: Patent Citations and the Value of Innovation," *RAND Journal of Economics* v.20, p.182-197.
- UNESCO (2010) *UNESCO Science Report 2010* (UNESCO: Paris)
- UNESCO (2005) *UNESCO Science Report 2005* (UNESCO: Paris)
- UNESCO (1998) *UNESCO Science Report 1998* (UNESCO: Paris)
- UNESCO (1996) *UNESCO Science Report 1996* (UNESCO: Paris)
- UNESCO (1993) *UNESCO Science Report 1993* (UNESCO: Paris)
- Van Arsdall, A. (2007) "Pulsed Power at Sandia: The First Forty Years," SAND2007-2984P (Sandia National Laboratories: Albuquerque, NM)
- Vega-Redondo, F. (2007) *Complex Social Networks* (Cambridge University Press: Cambridge, MA)
- Yegorov, I. (2009) "Post-Soviet Science: Difficulties in the Transformation of the R&D systems in Russia and Ukraine," *Research Policy* 38, no. 4 (May 2009), p. 600-609.
- Yonas, G. (1978) "Fusion Power with Particle Beams," *Scientific American* v. 239, no.5, p. 50-61.
- Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications* (Cambridge University Press; New York)

Winkelmann, R. (2008) *Econometric Analysis of Count Data, Fifth Edition* (Springer Verlag: Berlin)

Woods, J.P. (2012) Statement to the House Committee on Homeland Security Subcommittee on Counterterrorism and Intelligence, *Economic Espionage: A Foreign Intelligence Threat to American Jobs and Homeland Security*, Hearing, June 28, 2012.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press: Cambridge, MA)

Zhu, W. and Chen, C. (2007) "Storylines: Visual Exploration and Analysis in Latent Semantic Spaces," *Computers & Graphics* 31, no. 3, p. 338–349