



NATIONAL ENDOWMENT FOR THE

Humanities

OFFICE OF DIGITAL HUMANITIES

Narrative Section of a Successful Application

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Office of Digital Humanities application guidelines at <http://www.neh.gov/grants/guidelines/IATDH.html> for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title: Network Analysis for the Humanities

Institution: University of California, Los Angeles

Project Director: Timothy Tangherlini

Grant Program: Institutes for Advanced Topics in the Digital Humanities

1. Project Narrative

A. Proposal Overview

We propose an Institute in Advanced Topics in Digital Humanities focusing on techniques for the discovery, visualization and analysis of networks for the Humanities. Networks in this context are broadly defined to include both external networks (networks of production, networks of circulation, networks of influence, and networks of reception) and internal networks (networks of characters, networks of texts, networks of language) in Humanities data. The institute will consist of two main events: a ten day summer workshop in June 2010, and a three day research symposium in June 2011. It will be directed by Timothy Tangherlini (UCLA) and co-directed by Russel Caflisch (UCLA). Both events will be housed at NSF's Institute for Pure and Applied Mathematics (IPAM) on UCLA's campus.

During the ten day workshop in June 2010, participants will be introduced to the most recent approaches and tools for the discovery, visualization and analysis of networks; they will explore how these techniques can be applied to Humanities corpora; and they will gain experience in applying these techniques to Humanities data. Training in the use of software for network discovery/analysis, as well as workshops on how to prepare data for analysis, how to analyze network data, and the type of questions that can be posed/answered given a network model of a Humanities corpus, coupled to the theoretical basis for this work from an Applied Mathematics/Computer Science (AM/CS) perspective will be the main focus of this part of the institute. Presentations geared specifically to a Humanities audience will be held by fifteen to eighteen invited specialists in network discovery, visualization and analysis from the AM/CS community. Hands-on sessions exploring these computational techniques will be led by these presenters along with UCLA faculty and technology staff in UCLA computer labs. On alternate days, there will be a series of "success story" presentations by Humanities scholars to further illustrate how network discovery and analysis techniques can be applied to Humanities problems. Finally, the workshop will include several public lectures by prominent AM/CS and Humanities scholars on broad theoretical questions in network analysis for the Humanities. The last two days of the workshop will focus on preliminary collaborative work by workshop participants and faculty. During the three day symposium in 2011, participants will be invited back to UCLA where they will present the results of their collaborative work on network analysis that they started during the workshop in 2010. Participation in the institute will be open for up to thirty scholars from the Humanities. We will set aside four to five slots specifically for advanced graduate students in the Humanities.

This two year initiative aims to bridge the considerable gap that currently exists between the Humanities on the one hand, and Applied Mathematics and Computer Sciences on the other hand. In bridging this gap, the institute participants will explore what is possible in the study of networks, and learn how to apply these techniques to complex Humanities datasets. They will also meet and begin exchanging ideas with some of the leading scientists in the AM/CS community who work in the area of networks. A successful institute will have several tangible outcomes. These outcomes include developing a network of Humanities scholars conversant in techniques for advanced network discovery and analysis; fostering collaborations between Humanities scholars

and AM/CS scholars; and collaborative publications addressing aspects of the application of computational techniques to the study of networks in the Humanities. Scholars trained in this NEH institute will serve as a potential core group for an NSF-sponsored program at IPAM emphasizing the AM/CS side of this complex equation.

B. Background and Significance

In recent years, attention has been drawn in both the academic and popular press to the ubiquity of networks in everyday life, from communications networks to investment networks to power transmission networks to social networks (Barabási 2002; Buchanan 2002; Watts 2003; Kleinberg 2002 and 2007). As a result of this increasing awareness, the study of the different types of networks that link us together and the analysis of the structure of those networks has risen to greater and greater prominence not only in the mathematical, computational, and social sciences but also in the Humanities. This increased attention to networks in the Humanities can be traced in part to an abiding interest among Humanities scholars in understanding the shifting terrain of artistic expression, production and reception, coupled to a keen interest in understanding how a work of art—be it a novel, painting, a film or a poem—circulates and creates meanings for ever-changing communities. In cultural studies, the need for understanding the structure and dynamics of networks ranges from recognizing the impact of the internet on human expression and community formation in general, to the examination of the influence that virtual social networking tools such as Facebook or virtual recommendation tools such as Netflix have on communities of production and reception. Despite this increasing awareness of the importance of networks for theoretical advances in the Humanities, there is a considerable divide between recognizing in the broadest strokes the existence of these complex, dynamic systems and the very hard work of the consistent application of rigorous theoretically sound methods to the study of networks. Part of this gap can be attributed to the complexity of these networks—simply developing detailed, data-dependent descriptions of these networks is difficult, let alone analyzing these descriptions. Computational tools for the discovery and analysis of networks offer the promise of bridging this gap; unfortunately, many of these tools are as complex to work with as the underlying data itself. A main goal of this institute is to teach a group of Humanities scholars some of the most accessible of these techniques.

The recent increase in attention to networks might imply that networks are a new phenomenon—or a newly discovered object of study. Yet understanding networks has always been a key part of the Humanities; the problem has simply not been overtly stated as such. Internal to any work of art—a novel, for instance—one finds a series of networks: the social network of characters, as well as the more abstract network of language. Similarly, one finds in any work of art traces of networks of inspiration and influence. These networks of influence tie works of art together not only within literary periods, but also across periods, across linguistic boundaries, and across cultures. External to any work of art one can readily identify several other networks: networks of production linking authors to publishing houses to distributors and to consumers; networks of reception linking readers/viewers to each other and to the works of art; and networks of influence linking artists to each other (to name but three external networks). All of these networks are, of course, interrelated and in many respects mutually constitutive.

Despite the ubiquity of networks underlying Humanities study, little attention has been focused on understanding these networks in a consistent and scalable fashion. Humanities scholars have, for many years, been trained in the art of “close reading,” exploring a limited number of works from a particular cultural area, a particular period, a particular authorship or a particular movement. Indeed, in most cases, we as Humanities scholars have been trained to read canonically; even when we break out of these limits, transgressing the bounds of the canon, the scope of such transgression is necessarily limited by externalities such as time limitations and the accessibility of lesser known or forgotten works. It would be highly unusual for a scholar to be able to read and have significant command of more than a thousand works. This intensely focused approach has, on the one hand, led to several centuries of significant advances in understanding artistic production and reception. On the other hand, it has also been necessarily limiting, resulting in possible lacuna in our understanding of the development of humanistic production across cultures and through the ages. With the rapid development of very large digital collections in the Humanities, and the simultaneous development of very fast, yet inexpensive, computers, the time is ripe to engage in very large scale “distant reading” (to borrow a term from Franco Moretti), of hundreds, thousands or even tens of thousands of works. In this environment, we can discover and interrogate all of the different networks that inform and lend coherence (or incoherence) to a corpus. We can potentially determine the validity of our genre classifications, and even begin to interrogate the reach of influence, not only of authors or artistic movements, but also of publishing houses and distribution systems. We might also be able to discover why some works of art become “classics,” while other works disappear.

From a networks perspective, the Humanities are a particularly rich field of inquiry. Characters in novels form their own social network, with some characters forming positive bonds to each other, and other characters forming negative bonds. Frequently, over the course of a novel, these bonds change dramatically, with some links disappearing, and others strengthening. One could easily imagine that investigations of nineteenth century British literature, such those undertaken as part of the University of Virginia’s NINES project, might be expanded through the application of the network analysis techniques explored in this workshop. Are there network patterns of characters that characterize certain literary movements, authorships or genres? Do some internal network patterns appeal more to a nineteenth century readership than a twentieth century readership, as measured by editions sold? What is the relationship between social networks in fiction and in the contemporaneous historical period in which those works were written? The Icelandic sagas, from the late medieval period, for instance, are largely predicated on concepts of feud yet, as has been well established, the structure of feud is not uniform across the saga corpus. An analysis of feud networks in the saga would provide fascinating insight into shifting conceptions of feud and resolution in late medieval Europe. Projects such as Perseus and Monk have offered Humanities scholars a remarkable platform for computationally based text analysis including text classifiers and statistical tools related to word relationships; adding a network model of language will provide additional gains in our understanding of very large corpora (all of Classical literature, for instance). Even for a smaller corpus, such as Gertrude Stein’s writings, a network model (as opposed to a collocation or concordance model) might provide additional insight into this author’s creativity. Music corpora present their own challenges,

but the linkages of influence between pieces, and within pieces, can be productively studied from a network model. Models of external networks are also a promising area of enquiry. It is safe to say that the entire field of folklore is predicated on understanding communication in social networks as well as the complex manner in which these networks intersect and inform each other. In the case of the Child ballads, for instance, one has networks of ballad singers, networks of collectors and scholars, and networks of publishers as well as the complex network relationship between texts and melodies. A final example can be drawn from literary history, where considerations of networks of influence, production, distribution and reception have often been difficult for Humanities scholars to incorporate fully into their work. Network analysis might allow us to develop a model for understanding the complex and dynamic relationships between playwrights, theaters, and audience during the two century heyday of the Commedia dell'arte and the influence that these relationships may have had on each other, and the art form itself.

Over the past decade, there has been a sea change in applied mathematics and computing science in the understanding of networks. This change was presaged by the convergence of developments in understanding the mathematics of networks and graphs with the availability of increasingly large, complex datasets and sufficiently fast computers. Consequently, social scientists, applied mathematicians and computer scientists have been able to develop sophisticated techniques for the discovery, description, visualization and analysis of networks. The goal of this institute is to bring leading figures in applied mathematics and computer science together with Humanities scholars who would like to learn how to apply these techniques to the increasingly large and complex datasets now available in the Humanities. This institute will stand as a small, yet vital step, in the process of addressing the very large questions presented above.

C. Institutional Profile

UCLA is the largest research university in California, and is one of two flagship campuses of the University of California system. The proposed Institute in Advanced Topics will make use of two important centers at UCLA: NSF's Institute for Pure and Applied Mathematics and UCLA's Center for Digital Humanities. The resources of these two institutes combined will ensure that we are able to attract a wide range of scholars who are interested in advanced topics in computation for the Humanities while, at the same time, assuring that the scientists from the AM/CS communities are leaders in their fields. One of the main hurdles to addressing computation in the Humanities is high level and credible access to the research network(s) of the AM/CS communities: IPAM can provide us with that connection.

IPAM is one of seven NSF funded national institutes focusing on the mathematical sciences. It has distinguished itself by its broad interdisciplinary focus and was recently lauded by the NSF review committee for this approach. It is widely recognized internationally as one of the leading institutes for the interdisciplinary study of the application of mathematical models to solve difficult problems in other fields. The willingness of IPAM researchers to cross disciplinary boundaries, often in dramatic ways, is a hallmark of the Institute; it is also one of the reasons that Caflisch, the current director of IPAM, along with the scientific advisory board of IPAM, are eager to assist in hosting the proposed NEH Institute for Advanced Topics in Digital Humanities.

IPAM is an excellent venue for the proposed institute for several reasons. First, the institute has its own building centrally located on the UCLA campus. The building houses not only a series of well-equipped conference and lecture rooms, but it also has several “commons” areas that become the focal point of informal discussions. Since this proposed institute is scheduled for a gap in IPAM’s otherwise full calendar, it may be possible to offer all participants office space in the IPAM building. Computing support is very good, from wireless connectivity to access to the various high performance computing resources available at UCLA, including the Hoffman2 cluster, housed at UCLA’s Academic Technology Services. Finally, IPAM has significant experience in hosting hands-on tutorials as part of their programs, and we will be able to take advantage of their expertise in this area as we develop the hands-on sessions for this institute.

While IPAM provides the institute with access to the AM/CS community, CDH offers a high-visibility link to the Digital Humanities community. Eight years ago, the Division of Humanities recognized the need for a research-oriented center for understanding the emerging field of Digital Humanities. The Center, currently under the joint faculty leadership of Willeke Wendrich and Todd Presner, has been successful in attracting funding for a wide range of sophisticated projects in Digital Humanities. Currently, the Center is hosting a Mellon-sponsored project on “What is(n’t) the Digital Humanities,” and has recently received funding from the Keck Foundation to develop an undergraduate curriculum in digital cultural mapping. Consequently, we have access to a well-established network of Humanities scholars on whom to draw for possible participants. Coupled to CDH’s close connection to HASTAC, we expect to be able to attract a broad range of scholars to participate in this institute.

D. Curriculum and Work plan

The institute has two parts—a main workshop held in June 2010, and a follow-up symposium in June 2011. The main workshop is deliberately designed to run over the course of two weeks, to allow participants an adequate opportunity to interact with each other, to experiment with software and data, and to learn from the faculty and other participants. The Humanities faculty have been selected on the basis of their current work in computation and the humanities, while the AM/CS faculty have been selected on the basis of their interest in developing and applying tools for the analysis of complex Humanities data, be this in the form of artistic corpora (such as a collection of novels, an archive of music recordings, a record of correspondence, or collections of poetry), historical datasets (such as publishing records), or a combination of these. Furthermore, the AM/CS participants have been selected based on their ability to present extremely technical topics in a manner that is consonant with the generally non-technical background of most Humanities scholars. We understand that this is a key part of the success of this institute: there is little point in holding an institute the content of which is inaccessible to its participants.

The institute is designed so that there is significant amount of time for two types of important activity: (1) independent and group learning/experimentation with software on test datasets, so that lessons learned are not purely theoretical, but have an applied component to them; and (2) structured free-time for developing collaborative ideas. This latter component is concentrated during the second week of the institute by which time participants should have had time to learn and try some of the approaches discussed

during the first week. This scheduling also allows participants enough time to make some preliminary contacts and find potential collaborations with other participants and institute faculty. These structured free-time sessions will be allowed to develop organically within reasonable parameters—several topics for these breakout clusters are described below and in the appendix describing the curriculum. We expect that the contacts between scholars and preliminary collaborative experiments undertaken in this week will lay the foundation for collaboration in the ensuing year. These collaborations will form the basis for presentations at the shorter three day symposium in year two of the grant.

In broadest terms, the topics to be addressed in the Institute are: (a) the science of networks and networks in Humanistic inquiry (b) preparing and cleaning Humanities data for network analysis (c) internal networks in Humanistic data: networks of characters, networks of texts, networks of language (d) external networks in Humanistic data: networks of influence, networks of production, networks of reception.

Year One

In year one of the grant, the Institute will focus on the ten day workshop. The “normal” workshop day will be broken into a series of “sessions”. Morning sessions will run from 8:30-12:30, and afternoon sessions from 2:00-5:00 or 2:00-6:00pm. Evening lectures will run from 7:00-9:00pm. The morning sessions will consist of presentations by Institute faculty; these presentations will be structured as seminars rather than lectures. Using this format, participants will be encouraged to stop the presentation to explore topics in greater detail, and to receive clarification of points. Afternoon sessions are designed to be “hands-on”, with participants learning how to work with software on test data that has been properly formatted for particular exercises. On alternate workshop days (four in all), we will have one hour “success story” presentations by Humanities scholars who have used network analysis in their own work. In the second week, participants will be encouraged to work with the network analysis techniques learned during the first week using their own data. Except for Day 0, and the public lectures, all of the sessions will take place at IPAM.

The general outline of the workshop is described briefly below. A more substantive discussion of topics and presenters can be found in the curriculum appendix.

Week 1:

Day 0: Introduction to the Workshop / Institute

- a. General Work plan / Administrative matters
- b. Guest Lecture: Networks and the Humanities

Day 1: Networks for the Humanities—overview

- a. Morning Session – Complex Systems and the Humanities
- b. Afternoon Session – Understanding External Networks in the Humanities / Models of Network Formation, Growth and Decay
- c. Evening Session—Public Lecture
 - Jon Kleinberg (Cornell) or Duncan Watts (Columbia)

Day 2: Data Preparation for Humanities Network Analysis—

- a. Morning Session—Data preparation for network analysis
- b. Afternoon Session—Hands-on Session: Data formatting for network analysis

Day 3: Network Visualization

- a. Morning Session—Representing Networks / The Power of Visualization

- b. Afternoon Session
 - a. Success story lecture 1: Lew Lancaster (Berkeley), “Seeing Dots: The SGER Program and a Network Visualization Approach to the Korean Buddhist Canon”
 - b. Hands-on session with network representation / visualization tools
- Day 4: Network Analysis 1: Social Networks inside and outside Humanities data
- a. Morning Session—Social Networks and Information Flow
 - b. Afternoon Session—Hands-on session with Humanities network data
 - c. Evening Session—Public Lecture (Royce Hall 314 / Reception Royce 306)
 - Franco Moretti (Stanford): Network Analysis and the Humanities
- Day 5: Network Analysis 2: Language as Networks
- a. Morning Session—Language Networks and Machine Learning
 - b. Afternoon Session
 - a. Success story lecture 2: Jonathan Berger (Stanford): Automatic Music Recognition and Networks of Influence
 - b. Hands-on session on language network tools
- Week 2:**
- Day 6: Network Analysis 3: Language as Network 2
- a. Morning Session—Narrative Chains and Topic Maps
 - b. Afternoon Session—Hands-on session with collocation networks, topic maps, LDA
 - c. Evening Session—Public Lecture (Royce Hall 314 / Reception Royce 306)
 - a. Ronald Coifman (Yale) – TBA
- Day 7: Network Analysis 4: Networks of texts
- a. Morning Session—Tools for network discovery in texts: Named entity detection, mapping character relations in texts and supervised learning
 - b. Afternoon Session
 - a. Success Story Lecture 3: John A. Walsh (Indiana): Networks and Network Visualization in Swinburne
 - b. Hands-on session: supervised learners (Naïve Bayes, SVM)
- Day 8: Network Analysis 5: Unsupervised Machine Learning and the Analysis of Networks
- a. Morning Session—Introduction to Unsupervised Learners
 - b. Afternoon Session
 - a. Success Story Lecture 4: Tim Tangherlini (UCLA), “Facebook for Ghosts: Machine Learning and Network Analysis in Folklore”
 - b. Hands-on session: WEKA learners / RapidMiner
- Day 9: Preliminary presentations 1
- Day 10: Preliminary presentations 2 / Charter for future collaborations
- a. Morning Session—Presentations
 - b. Afternoon Session—Wrap-up and Work plan for second year

Year Two

In year two of the Institute, we will hold a concluding research symposium. The symposium will offer an opportunity for participants to present their collaborative work on network analysis in the Humanities started during the first year of the institute. Topics

for the symposium will be driven entirely by the research produced by the participants and their collaborators during the year. The symposium will also have two keynote addresses in the evening, one by a prominent Humanities scholar and another by a prominent AM/CS scholar. By hosting this symposium at IPAM, we will also be able to attract audiences from applied mathematics and computer science, an important part of bridging the gap between our scholarly communities.

Day 1: Introduction / Presentations / Public Lecture

Day 2: Presentations / Public Lecture

Day 3: Presentations / Institute conclusion

Reading / Materials

A series of main texts will be assigned to participants prior to their arrival at UCLA. Two of the books, Duncan J. Watts' *Six Degrees* (2003) and Albert-László Barabási's *Linked* (2002) are quite accessible, and require little training. Also, participants will be asked to read Stanley Wasserman and Katherine Faust (1994) *Social Network Analysis: Methods and Applications*, and Matthew Jackson *Social and Economic Networks*. A final book, *Introduction to Information Retrieval*, by Christopher Manning, Prabhakar Raghavan and Hinrich Schütze is slightly less accessible, but provides an excellent overview of the techniques to be explored in this institute. Other readings for the course will be developed during the course of the year leading up to the workshops. These will include a series of topic specific articles, as well as user manuals for Network Workbench and RapidMiner, two of the main software packages that will be used in the workshop. Additional software, developed or used by the workshop faculty will also be made available along with any relevant documentation. An interactive institute website will allow participants to interact with institute faculty in the form of online discussion boards from January 2010 through the start of the summer workshop. The website will continue to serve as a virtual collaborative space for the time between the end of the workshop and the research symposium in 2011.

E. Participants

Participants for this workshop will be recruited via an open call across a number of organizations to which Humanities scholars interested in Digital Humanities and computational approaches to Humanities data look for news, events and support. Foremost among these networks are HASTAC (Humanities, Arts, Science and Technology Advanced Collaboratory) and ADHO (Alliance for Digital Humanities Organizations). We will make a similar call for participation through the MLA, as well as through our Mellon sponsored seminar on "What Is(n't) Digital Humanities." We do not intend to limit participation in this institute based on a pre-selected group of people. On the other hand, we do intend to screen applications for participation according to several criteria, as we want the institute to be an opportunity for fairly rapid acquisition of difficult skills. To that end, we will expect participants to have some background computation for the humanities, and bring with them—or at least have ready access to—a data set with which they want to work. Although we will make several test datasets available for learning purposes, the goal is that all participants will make preliminary steps toward applying some of the techniques presented in the first week and a half of the institute to their presentations in the last two days of the institute. Participants will be

made aware of the expectation that they will pursue networks research subsequent to the workshop that can be presented at the research symposium in the second year of the grant. We will reserve four to five spots for advanced graduate students in the Humanities, as we hope that techniques and theories learned at this institute can help them in their research and equip them for success in an increasingly competitive job market where expertise in Digital Humanities is seen as a “value added” qualification for any job candidate. We believe that our Institute will also form a strong complement to some of the most sophisticated groups in computing in the Humanities, including Project Monk at Illinois, The Nines Project at Virginia, and the Perseus Digital Library Project at Tufts.

The call for applications will be sent out in July 2009, with an application deadline of October 15, 2009. Selection will be made by November 15, 2009. The applications will be reviewed by Tangherlini, Cafilisch, Borovsky, Lancaster, Berger and Eliassi-Rad. A final roster of participants will be selected by December 1, 2009. These participants will then have access to an institute website, which will act as a virtual platform for initial study prior to the residential workshop. Participants will be expected to have read an initial reading list and worked with some preliminary exercises in data formatting and network discovery and visualization during these months, so that when they arrive at UCLA in June 2010, they will have had four to six months of preparation.

F. Impact and Evaluation

We expect that the institute will have significant impact in jump-starting computational work on Humanities networks. At a recent workshop sponsored by the NEH and held under the auspices of the Committee on Library and Information Resources, Greg Crane asked the provocative question, “Why haven’t there been any papers on clustering in the Humanities?” The participants in this institute will, at the very least, be well equipped to lay this observation to rest. We expect that the outcome(s) of this institute should resonate throughout the humanistic scholarly community, educating a new group of scholars in techniques for the discovery and analysis of networks in the ever increasing body of large humanities corpora. By design, the institute is also intended to foster meaningful and sophisticated interdisciplinary collaborations between humanities scholars and scholars from the AM/CS communities.

Participants in the institute will be asked to evaluate the institute on several occasions: on the first day, participants will be asked to briefly describe their goals for the two week workshop. At the end of each of the weeks of the workshop, the participants will be asked to evaluate both the theoretical and practical instruction, as well as the structure of the workshop. After six months, the participants will again be asked to evaluate their experiences with the institute, and the impact of the institute on their own research. The institute will continue to host a website, which will act as a repository for white papers and preprints produced by the participants and their scientific collaborators as a result of their participation in the institute. Finally, during the symposium in 2011, participants will be asked to provide a final evaluation of the institute. Presentations at this symposium will also help us gauge the impact of the institute on participants’ research.

IPAM has significant experience in evaluating workshops, symposia and summer schools and has pledged evaluation support as part of their role in the institute.

G. Staff, faculty and consultants

Because of the technical nature of this institute, and because of the close collaboration between scholars from the Humanities and AM/CS, we have decided to have two boards of advisors, a Scientific Panel and a Humanities Panel beyond the typical Director/co-Director structure. People who have committed to participating in these leadership capacities include Tangherlini (UCLA) as director and Caflisch (UCLA) as co-director as well as the members of the Scientific Advisory Panel and the Humanities Advisory Panel. The Scientific Advisory Panel is comprised of Mark Green (UCLA), Peter Jones (Yale), Ronald Coifman (Yale) and Tina Eliassi-Rad (Lawrence Livermore National Laboratory). The Humanities Advisory Panel is comprised of Lewis Lancaster (UC Berkeley), Gregory Crane (Tufts), Jonathan Berger (Stanford) and Zoe Borovsky (UCLA). Profiles of these key participants have been appended to this application.

A number of core faculty, lecturers and consultants have been contacted and have committed to participating in the institute. Given the large number of faculty, and the high demand that these scientists face, we recognize that some may not be able to participate as planned. Because of IPAM's extensive network and high profile, we expect to be able to find suitable replacements if any of the faculty, consultants or lecturers are forced to cancel. This may also lead to small modifications to the topics presented in the morning sessions. To date, we have received commitments from Yannet Interian (Google), David Blei (Princeton), Edoardo Airoldi (Harvard), Tony Davis (Comcast), Kendall Giles (VCU), Katy Börner (Indiana), Steven Zucker (Yale), David Smith (UMass), Fernando Diaz (Yahoo!), David Liben-Nowell (Carleton) and John A. Walsh (Indiana). We have received tentative commitments from Stanley Wasserman (Indiana), Filippo Menczer (Indiana), Matthew Jackson (Stanford) and Johan Walden (UC Berkeley). Finally, Franco Moretti (Stanford) and Ronald Coifman (Yale) have agreed to present plenary lectures.

H. Budget notes

IPAM will provide rooms, offices, and staff support for the Institute. This commitment constitutes the cost-share portion of the grant. The two largest NEH funded expenditures for the institute are related to transportation and housing for participants and faculty during the two institute events. Travel costs are based on an average expected cost for airfare and airport transfers. Lodging is based on an average of current UCLA rates at three local hotels. We have decided to provide refreshments during breaks—IPAM has, over its many years of holding various institutes of this nature, discovered that a great deal of collaborative work is started over informal conversations during breaks. By keeping the participants in the large, informal common areas of IPAM's building, we hope to have similar success. Finally, we have budgeted for lunch for the four days on which we have "Success Story" lectures—this is a purely pragmatic decision as these days are tightly scheduled, and we would like to keep participants from dispersing. The symposium likewise includes budget lines for coffee breaks. There are two lines for faculty support: one for Tangherlini (Director) and one for Borovsky. In addition to her service on the application review team, she will maintain the Institute interactive website, ready the computer laboratories for the workshop, and coordinate other technical aspects of the workshop.

1. Curriculum and Work plan

Detailed Curriculum for Network Analysis for the Humanities

Year One—Ten Day Workshop

Week 1:

Day 0: (Held at UCLA's Mays Landing retreat site/Malibu)

- a. Part One:
 - a. Introductions: this first day of the workshop is intended to allow participants and faculty an informal opportunity to meet and discuss their goals for the workshop, discuss their background, and introduce to the other participants the main corpus with which they work. Facilitated by Tangherlini (UCLA) and Caflisch (IPAM)
 - b. Part Two:
 - a. Introduction to Network Theory and Networks for the Humanities: Guest lecture. We will invite Duncan Watts, author of *Six Degrees*, or Jon Kleinberg, whose *Memetracker* project has great relevance to the Humanities, to hold the introductory lecture.
 - c. Evaluation 1: Expectations questionnaire

Day 1: Networks for the Humanities—overview

- a. Morning Session
 - a. Welcome (Tangherlini and Caflisch; Peccei)
 - b. Filippo Menczer (Indiana U): Models of Network Formation, Growth and Decay
 - c. Tina Eliassi-Rad (Lawrence Livermore National Laboratory): Gentle introduction to network analysis, graphs and analysis of graphs—concepts of nodes and edges. Introduction to directional graphs. Networks as graphs.
- b. Afternoon Session
 - a. Yannet Interian (Google): External networks in Humanities data – understanding social networks; scale free networks; power laws and distributions; small world problem; affiliation networks
 - b. Johan Walden (UC Berkeley) Information models and value in Humanities data
 - c. Edoardo Airoldi (Harvard) Influence, Affiliation and Taste: Understanding the Netflix Competition
- c. Evening Session—Public Lecture (Royce Hall 314 / Reception Royce 306)
 - a. Franco Moretti (Stanford): Network Analysis and the Humanities

Day 2: Data Preparation for Humanities Network Analysis—

- a. Morning Session
 - a. Data formats, data cleaning and data preparation for network analysis
- b. Afternoon Session
 - a. Hands-on Session: Data formatting for network analysis
 - i. Finding and acquiring data
 - ii. Data formats
 - iii. Tagging / using existing tags to help construct networks
 - iv. Experiment / hands-on workshop with small Humanities data-sets

Day 3: Network Visualization

- a. Morning Session
 - a. Steven Zucker / Edo Liberty (Yale): Network representation and visualization
 - b. W. Bradford Paley: Visualizing Networks in and of Humanities data
- b. Afternoon Session
 - a. Success story lecture 1: Lew Lancaster (Berkeley), “Seeing Dots: The SGER Program and a Network Visualization Approach to the Korean Buddhist Canon”
 - b. Hands-on workshop with network representation / visualization tools / Network Workbench and Pajek

Day 4: Network Analysis 1: Social Networks inside and outside Humanities data

- a. Morning Session
 - a. Stanley Wasserman (Indiana): Introduction to Social Network Analysis
 - b. Matthew Jackson (Stanford): Social and Economic Networks in the Humanities
 - c. David Liben-Nowell (Carleton): Tracing Information Flow on a Global Level: Chain Letters and Networks of Communication
- b. Afternoon Session
 - a. Hands-on workshop with Humanities network data: mapping network of characters in Jane Austen’s *Emma*
- c. Evening Session—Public Lecture (Royce Hall 314 / Reception Royce 306)
 - a. Jon Kleinberg (Cornell) or Duncan Watts (Columbia)

Day 5: Network Analysis 2: Language as Networks: Natural Language and Language Network Tools

- a. Morning Session
 - a. Christopher Manning (Stanford) Introduction to machine learning and natural language processing: Hubs, authorities, Markov chains and PageRank for the Humanities
 - b. Tony Davis (Comcast): Introduction to language network tools (semantic web; collocation)
- b. Afternoon Session
 - a. Success story lecture 2: Jonathan Berger (Stanford): Automatic Music Recognition and Networks of Influence
 - b. Applied hands-on workshop on language network tools / YAKS, Wordij, NodeXL/Integrating language data into network representations in Network Workbench
 - c. Evaluation 2: Interim evaluation

Week 2: During week two of the institute, presentations are held only in the morning. The afternoons are set aside for group work/experimentation, and preparation for the last two days of preliminary presentations and critiques. We meet briefly as a group each evening to debrief and discuss what we learned in the afternoon sessions.

Day 6: Network Analysis 3: Language as Network 2

- a. Morning Session
 - a. Dan Jurafsky (Stanford) Narrative Chains: Discovery of Narrative Patterns in large scale corpora

- b. David Blei (Princeton) Introduction to Topic Maps and Latent Dirichlet Allocation
- b. Afternoon Session
 - a. Hands-on workshop with collocation networks, topic maps, LDA
- c. Evening Session—Public Lecture (Royce Hall 314 / Reception Royce 306)
 - a. Ronald Coifman (Yale) – Information Cascades and the Study of Rumor

Day 7: Network Analysis 4: Networks of texts

- a. Morning Session
 - a. David Smith (UMass): Tools for network discovery in texts: Named entity detection, mapping character relations in texts
 - b. Fernando Diaz (Yahoo! Montreal): Supervised learning for texts as method to rapidly classify texts; inferred bibliometrics through unsupervised learning on large corpora
- b. Afternoon Session
 - a. Success Story Lecture 3: John A. Walsh (Indiana): Visualizing Networks in Swinburne
 - b. Introduction to supervised learners (Naïve Bayes, SVM) and RapidMiner software

Day 8: Network Analysis 5: Unsupervised Machine Learning for the analysis of networks

- a. Morning Session
 - a. Kendall Giles (VCU) Introduction to Unsupervised Learners / “Clustering” / Introduction to Markov chains and PageRank
- b. Afternoon Session
 - a. Success Story Lecture 4: Tim Tangherlini (UCLA), “Facebook for Ghosts: Machine Learning and Network Analysis in Folklore”
 - b. Training on WEKA learners / RapidMiner environment

Day 9: Preliminary presentations –

- a. This day is set aside for presentations of preliminary findings on the data sets that small groups of scholars have explored.
 - a. Each presentation team is given 30 minutes to present, and 30 minutes to respond to questions/critiques/suggestions from the other participants, institute faculty, and interested parties from UCLA’s Humanities faculty, and faculties in Computer Science and Applied Mathematics. We expect to do eight presentations on this day.

Day 10: Preliminary presentations and Charter for future collaborations

- a. Morning Session
 - a. The morning session will complete the preliminary presentations from the previous day, with up to four presentation teams presenting.
- b. Afternoon Session
 - a. Chart possibilities for future collaborations
 - b. Explain the format for the 2011 symposium
 - c. Evaluation

Year Two – Three Day Research Symposium

The program for this year is reserved entirely for participants to present their research on Network Analysis in the Humanities. There will be two keynote public lectures on the first two evenings of the symposium, one by a Humanities scholar, and another by a AM/CS scholar. In

the event that more than 20 participants are prepared to present their work, we will hold several concurrent sessions to accommodate all of these presentations. Presentations will be solicited with an April 15, 2011 deadline to allow for scheduling.

Day 1: Thursday

- a. Morning Session:
 - i. Introduction / Welcome (Caflisch /Tangherlini)
 - ii. Session 1 (2 presentations)
- b. Afternoon Session:
 - i. Session 2 (2 presentations)
 - ii. Session 3 (2 presentations)
- c. Evening Session: Public lecture

Day 2: Friday

- d. Morning Session:
 - i. Session 4 (2 presentations)
 - ii. Session 5 (2 presentations)
- e. Afternoon Session:
 - i. Session 6 (2 presentations)
 - ii. Session 7 (2 presentations)
- f. Evening Session: Public Lecture

Day 3: Saturday

- g. Morning Session:
 - i. Session 8 (2 presentations)
 - ii. Session 9 (2 presentations)
- h. Afternoon Session:
 - i. Session 10 (2 presentations)
 - ii. Conclusion
- i. Dinner for participants