Abstracts of Student Presentations

Implementation of eGovernment Components of the European Union’s Information Society Initiative: A Social Informatics Perspective on Inequality in Outcomes

Madelyn Sanfilippo

Unequal access to information has significant social and political consequences, and is itself a consequence of sociotechnical systems born of social, cultural, economic, and institutional context. Information is unequally distributed both within communities and between communities. While many factors that shape information inequality shift subtly over time, due to historical distributions and path dependent natures, changes to sociotechnical systems and infrastructural support for ICTs are often intentionally changed through access initiatives prescribed in policies. Furthermore, these policies also often seek to address issues of access in other ways, such as through information and digital literacy campaigns. As a result evaluation of the policies that define access initiatives provide a useful mechanism to understand how successful efforts to increase access are and how context impacts implementation and use in such ways as to produce unequal outcomes.

The research presented is drawn from dissertation research that examines what policy dimensions, particularly as relate to ICTs and digital infrastructure, lead to unequal outcomes in access initiatives, using a comparative design to examine differences in outcomes from the Information Society initiative across the European Union. While the dissertation will look at differences in outcomes across three policy areas—eGovernment, broadband development, and information literacy—this presentation will emphasize how policy and technology interact to impact eGovernment outcomes across the EU. The presentation will characterize: how Eurostat data can be used to conceptualize access, patterns of inequality in eGovernment services and access across the EU, and which social and technical implementation factors that influence policy outcomes, drawn from the full analysis of policies within the initiative, are supported by eGovernment data. The presentation will conclude with how these results will be incorporated into the development of models that will explain differences in policy outcomes, despite the uniform applicability of EU policies across national and regional borders.

Examination of community-based and context-based information diffusion in Twitter

Zheng Gao

Twitter has become a rich source of people’s opinions about a variety of topics, such as their daily life, and current issues. Twitter’s retweeting and mentioning mechanisms enable users to disseminate information broadly. In this study, we investigate the effects of community-based and context-based features on the users’ information adoption and diffusion patterns in Twitter. Community-based features capture how the adoption of a hashtag by users within the target user’s community and users outside that community influences the target user’s selection of this hashtag. Context-based features measure the influence of other users’ adoption of hashtags that are semantically similar with a hashtag on the target user’s adoption of this hashtag. We find the community-based features enhance the prediction of users’ hashtag adoption and diffusion. However, the further exploration of context-based features is needed.
Understanding Scientific Collaboration from the Perspective of Collaborators and their Network Structures

Chenwei Zhang

Scientific collaboration is one of the key factors triggering innovations; many scholars have investigated it from various aspects with both qualitative and quantitative methods, such as the formation and evolution of research teams, characteristics of coauthorship, and evaluation of multi-author articles. Collaboration forms a kind of social networks that scholars interact with each other to collaboratively explore and solve problems in one domain or cross different domains. However, previous researches on scientific collaboration focus on either scholar’s behavioral studies, or quantitative measurements of scholars’ impact from the coauthorship network structures. Actually, it is important to consider both the covariate effects from authors’ own attributes and the social network effects to gain a holistic understanding on scientific collaboration. In this work, we analyzed a coauthorship network in the field of information retrieval (IR) using Exponential Random Graph Models (ERGMs) to see how authors’ attributes, such as productivity, popularity, and their research interests, and the related homophily effects, interacting with the social network effects, such as transitivity, and preferential attachment, to affect the authors’ scientific collaboration.

Our work differs from previous studies because of the comprehensive consideration of both covariate effects of authors’ attributes and social network structure features to understand research collaboration, rather than examining each feature in isolation. Meanwhile, ERGMs allow us to calculate the possibilities that two authors might collaborate resulted from the effects of their own attributes, homophily on these attributes, transitivity, and their preferential attachment. In addition, our quantitative research is based on a relatively large network.

The findings are informative: an author tends to collaborate more with those having more nonfirst-authored collaborative publications, but reluctant to coauthor with others having similar level of single-authored publication; an author’s tendency to form new cooperation with his/her coauthors’ collaborators, or those who share research interest, is strong. An author also tends to co-work more with those who have more previous collaborators. In the future, we will focus on the topic diversity of coauthors and their collaboration tendency. We will expand the features. We will also apply ERGMs to the author citation networks and overlay with the coauthorship network to study whether the impact will drive the scientific collaboration.

Semantic Measures in Historiographical Research

Vincent Malic

Historians create structure in our collective understanding of past events. When writing narrative history, they establish connections between significant events, the contexts in which these events occurred, and the people who participated in them. Most of the knowledge created by historians still exists in the form of unstructured texts. This poster presents an exploration of the applicability of the concept of semantic relatedness to historiographical research. Computational methods are applied on texts about the history of the Roman Empire to identify named entities and build word-word relatedness matrices. These matrices are then analyzed to reveal larger thematic structures that characterize a particular historian’s view of
historical events. This analysis demonstrates the value of semantic relatedness for the unsupervised detection of common themes and unusual outliers in the narrative texts historians have created to store our collective knowledge of the past.

Data-Driven Knowledge Discovery in the Alzheimer’s Disease Literature

Satoshi Tsutsui

Scientific knowledge, usually embedded in unstructured texts, grows exponentially these days that researchers can no longer read all the related literature (Hunter & Cohen, 2006). The knowledge is composed of individually connected knowledge units called entities, and their relationships. These entities include papers, authors, keywords, topics, and more specific domain entities such as disease, drug, and genes in the case of medical fields. Their relationships are the connections between two entities. For example, the relationship between drug and target can be binding. An efficient way to represent knowledge is to use knowledge graph where nodes illustrate entities and edges indicates relationships between entities. Therefore, it is important to automatically extract, organize scientific knowledge into knowledge graphs, and then analyze them to find new hypotheses that ultimately lead to knowledge discovery.

We focus on Alzheimer’s disease (AD) because of its high social impacts. It is estimated that 5.3 million Americans have AD in 2015, and the number is increasing (Association, 2015). The US government in 2015 is expected to spend $153 billion for Medicare and Medicaid for people with AD and other dementia (Association, 2015). Despite the high prevalence and the cost, there are many open questions about AD which are still not fully understood (Rizzieri, 2012). However, the current information available on AD is overwhelming. Researchers contribute a huge amount of knowledge (e.g., papers, patient information, brain images, and gene datasets) available in public, but no researcher can keep up with all of them. For example, if you search “Alzheimer” in PubMed, you will find 81,300 articles and 1,499 books. Who can read all this literature? Integrating published knowledge into a knowledge graph and analyzing it becomes an efficient way to discover new hypotheses that help these overwhelmed domain experts.

Here, we report the preliminary research toward knowledge discovery in AD. We collect 81,300 articles related to AD from PubMed, extract major Medical Subject Headings (MeSH) descriptors, and construct a term co-occurrence graph. We visualize the backbone of the graph using the pathfinder approach (Schvaneveldt, 1990). The visualization outlines the major topics and their relationships in the AD literature. This could be particularly helpful for newcomers in the AD field. The relationships between major topics and other topics, especially the topics that connect two major topics, could be insightful for domain experts to propose new hypotheses. Moreover, we detect major topics using Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003). The results bear many similarities with the visualization.

There are three major challenges for the future research. The first challenge is how to construct a knowledge graph that better represents heterogeneity and semantics of scientific papers. Heterogeneity means different types of entities. Semantics are meaningful interactions of entities. The second challenge is how to deal with dynamics when analyzing knowledge graphs. Dynamics means how a graph changes over time. The third challenge is how to analyze the structure of the graph. Structure means how a complex network is composed of.