

## ネットワークコンテキストに基づく 知的コラボレーションの支援

木實新一†

人々の知識の相違や重なりを考慮して効果的にコラボレーションを支援するためには、従来のような場所、時間、プレゼンス等のコンテキストだけでなく、ソーシャルネットワークを含むコミュニティ構造に関連したコンテキストを体系的に考慮する必要がある。本稿では、研究者の学術的なコラボレーションを支援するコミュニティマッピングツールについて述べ、複数のネットワーク構造を考慮したコンテキストウェアサービスの可能性を議論する。

## Supporting Intellectual Collaboration Using Network Context

Shin'ichi Konomi†

Understanding the differences and overlaps of users' knowledge is important in systems that support intellectual collaboration. Therefore, it is desirable to capture and use community-relevant contexts such as social network structures as well as conventional contexts such as location, time, and presence. This paper describes a community mapping tool that provides context-aware services based on multiple network structures to support academic collaboration among researchers, and discusses the possibilities of novel context-aware services for supporting intellectual collaboration.

### 1. はじめに

人々の知識の相違やオーバーラップを考慮して効果的にコラボレーションを支援するためには、従来のような場所、時間、プレゼンス等のコンテキストだけでなく、ソーシャルネットワークを含むコミュニティ構造に関連した「深い」コンテキストを考慮する必要がある。日常生活にコンピューティングおよびセンシング技術が浸透することにより、動的で複雑な実世界の現象を捉えた大量のデータにアクセスすることが容易になりつつある。このようなデータを用いた個人の行動推定については盛んに研究が行われているが、グループやコミュニティの構造を体系的に考慮したソーシャルなコンテキストウェアサービスについては比較的未開拓である。

本稿では、学術的なコラボレーションの支援に着目し、文献データベースから抽出したアフィリエーションネットワーク構造を用いて研究者コミュニティを支援するコミュニティマッピングツールについて述べ、コミュニティ構造を体系的に考慮したコンテキストウェアサービスの可能性を議論する。

研究者コミュニティは、研究者の日々の仕事や学習、協調、知識創造等の活動に様々な影響を与えている。コミュニティの構造は、構成員のつながりによる内部的な構造と、コミュニティ同士のつながりによる外部的な構造の2種類に分けて考えることができるが、特に後者はコミュニティの境界を超えた学際的なコラボレーションを行う際に問題になると思われる。これはCampbell[2]の鱗モデル(図1)が示唆するところでもあり、参加者の専門分野が少しずつ重なり合いながら面的に広がる「集合的包括性」や「適切な混ざり合い」[6]といった状態を可視化して、学際コラボレーション環境のデザインや改善に役立てることができると考えられる。



(a) 専門性のクラスタ

(b) 鱗モデル

図1 学術分野の重なりパターン

研究者の実際のソーシャルネットワークは、オフィシャルな研究分野の定義(例:「データベース」「ロボティクス」)に必ずしも一致するとは限らない。近年、研究業

† 東京大学空間情報科学研究センター  
Center for Spatial Information Science, the University of Tokyo

務の遂行や成果の公開等において、様々なデジタルツール・メディアが利用されている（例：文献データ、電子的な論文やプレゼン資料、ウェブサイト、ダウンロード可能なツールやデータセット、ブログ、ウェブサービス、科学的ワークフロー14。）更に、これらのツールやメディアの利用履歴データも用いて、「実際の」研究コミュニティ構造をより正確に反映したソーシャルネットワークを自動抽出することができる可能性がある。

本稿では、研究者コミュニティの構造に基づき初期段階の学際コラボレーションを支援するシステムを提案する。具体的には、文献データベース5)に基づいてコンピュータサイエンス分野の数百に及ぶコミュニティを抽出・可視化するデータセントリックなコミュニティマッピングツールを導入する。このツールは、成果発表の場(論文誌や国際会議などの publication venue)をイベント、研究者をアクターとするアフィリエイトネットワーク15)を生成し、論文誌や国際会議の重なり合いに基づいてコミュニティを抽出することができる。更に、「コミュニティの地図」の可視化を行うとともに、学際研究チームのデザインを支援する機能を提供する。

## 2. 関連研究

既知の研究者とのコラボレーションを支援するデジタルツール（電子メールや遠隔会議システム等）は数多く存在するが、共同研究相手を発掘しプロジェクトを立ち上げる過程を支援するシステムはそれほど多くない。研究者同士の出会いの場でもある学術会議においてコミュニケーションを支援するシステムとしては、AutoSpeakerID や Ticket2Talk9)がある。DeaiExplorer10)は共著者のネットワークを大画面に提示して学会参加者のコミュニケーションを支援するシステムであるが、コミュニティ構造に基づくコンテキストウェアサービスは提供していない。

科学者の分散コラボレーションを支援することができるシステムとしては、コラボラトリーと呼ばれる環境を構築する取り組みがある3)。特に、Tanら14)は科学者によるワークフローの利用パターンを、ソーシャルネットワーク分析の技法を用いて吟味している。分散型の学際コラボレーションについてはCSCW分野で特に盛んに議論が行われており4)11)、学際コラボレーションにおいて既存の社会的つながりや仲介者の存在が重要な役割を果たすことを示唆する結果が得られている。Birnholtzら1)は、コラボレーションの立ち上げに関連して、「誰が何を知っているか」に関する知識が重要であることを議論している。

研究者のコラボレーションパターンを理解するために、論文データベースから共著関係に基づくソーシャルネットワークを抽出し利用することができる8)。例えば、研究者のコミュニティ構造を理解するために、抽出したネットワークに含まれるクリーク

等を求めることができる。あるいは、クラスタリング手法7)を用いることも可能である。Pallaら12)は、互いに重なり合ったコミュニティ構造を抽出する手法を提案している。しかし、抽出された複雑な重なりのパターンを効果的に可視化することはそれほど容易でない。コンピュータサイエンス分野のコミュニティ構造を研究者ネットワークから抽出するためには、ノード数が百万前後のネットワークデータを処理する必要があると考えられ、非常に計算コストが高くなる。本稿では、アフィリエイトネットワークの概念に基づき、研究者のネットワークと双対的な関係にある成果発表の場(venue)のネットワーク構造を処理する。コンピュータサイエンス分野の文献データベースであるDBLP5)から発表の場のネットワークを抽出した場合ノード数は6,000程度であり、研究者のネットワークよりも非常に高速に処理を行うことができる。

## 3. その場の限りのインタラクション支援を超えて

本稿で提案するコミュニティマッピングツールは、DBLP文献データベースに基づき研究者のソーシャルネットワークを表示するシステムDeaiExplorerと統合されている。このデータベースには、コンピュータサイエンスとその関連分野において100万人弱の著者が出版した150万件程度の論文データが登録されている。DeaiExplorerシステムは学術会議参加者の名札に取り付けたRFIDタグに反応して、論文の共著関係や引用関係に基づき、その場にいる人たちのソーシャルネットワークを表示する。

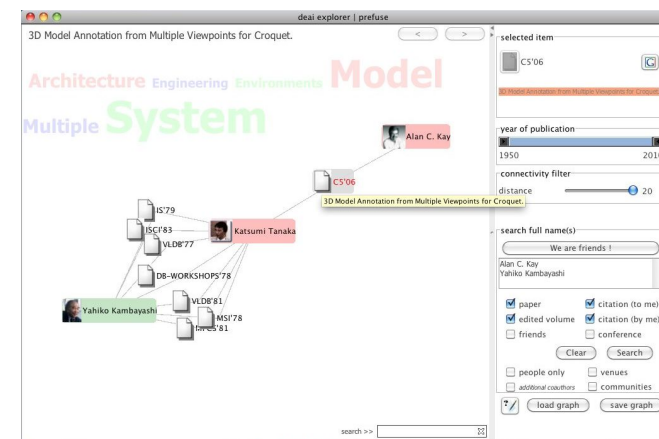


図2 DeaiExplorerの表示例

図2はDeaiExplorer上に研究者の間接的なつながりが表示された例である。このシステムはJavaと可視化ツールキットprefuse13を用いて実装されており、4つの国際会議で利用された。会議参加者のソーシャルインタラクションのきっかけ作りに役立てることができた。もちろん、会議参加者がちょっとした会話を行って打ち解けることができたとしても、必ずしもそれが有意義な研究コラボレーションへと発展する訳ではない。そのような発展を促すためには、会議参加者同士が互いの専門分野や知識・能力の近接性と境界を理解できるようにすることが重要であると考えられる。コミュニティマッピングツールは、ソーシャルネットワーク分析の技法に基づいてそのような理解を助けることを目指している。

#### 4. 発表の場のネットワーク

共著関係のデータからコミュニティ構造を効率良く導出するために、システムはまず発表の場のネットワークを生成する。発表の場のネットワークのノードは会議や論文誌等の発表の場を、エッジは発表の場間の関連を示している。発表の場間の関連は、発表の場と研究者の関連に基づいて導かれる。基本的には、2つの発表の場 $v_1$ と $v_2$ はそれら両方で発表した研究者の存在によって連結されることになる。これは、発表の場をイベント、研究者をアクターとするアフィリエーションネットワーク15(p.291)からイベントのみのネットワークを取り出すことに相当する。以下の式を用いて発表の場の類似性を計算し、得られた値を対応するエッジに割り当てる。

$$\text{Similarity}(v_1, v_2) = A_1 \cap A_2 / \min(|A_1|, |A_2|) \quad (1)$$

ここで、 $A_1$ と $A_2$ は、発表の場 $v_1$ と $v_2$ の全ての著者をそれぞれ表している。この類似度(シンプソン係数)は、発表の場が小さい場合に望ましくないバイアスが生じるため、著者数が200以下である発表の場は除外する。以下では、このネットワークを用いて、類似性の高い発表の場をまとめていき、同じ研究コミュニティに属す発表の場の集合を求める。そこで、前準備として閾値(0.15)よりも類似度の低いエッジを除去しておく。

図3は、ある研究者(中央に位置するノード)に関連する発表の場のネットワークを示したものである。発表の場を表すノードは、二つの円に沿ってレイアウトされており、内側の小さな円に沿ってレイアウトされたノードは、この研究者が発表した場であり、外側の大きな円に沿ってレイアウトされたノードは、この研究者が発表した場と類似した発表の場である。研究者が容易に発表の場を発見できるように、研究コミュニティの集成的な履歴データに基づいたこのようなネットワークを利用することができる。

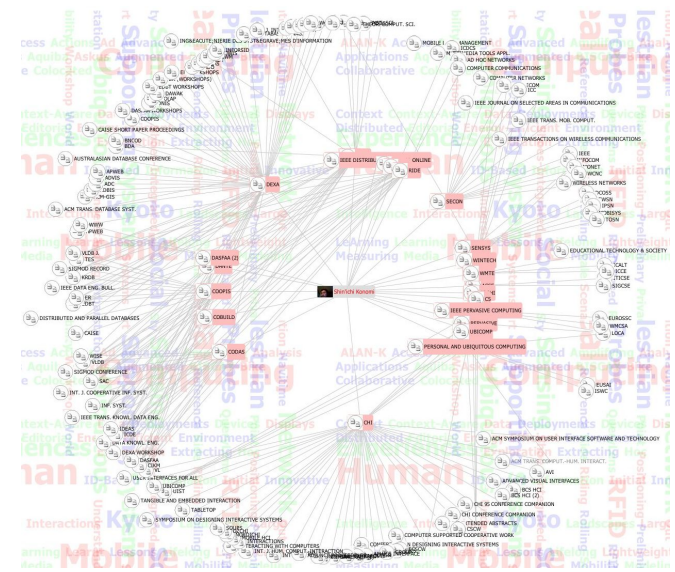


図3 研究者を中心とする発表の場のネットワーク

#### 5. コミュニティマッピング

コンピュータサイエンス分野における類似する発表の場のクラスタを高速に求めるために、DBLPデータベースの発表の場のネットワークに対してマルコフクラスタリングアルゴリズム7)を適用した。図4は結果として得られた249個のクラスタとそこに含まれる発表の場をバネモデルに基づいてレイアウトしたものである。各クラスタの名前は手作業で入力を行ったが、インタラクティブな入力ツールやクラウドソーシングを利用すれば、手作業による入力のコストを小さくすることができると考えている。

なお、開発したツールは、色を用いて指定したコミュニティを強調表示することができるが、図4では次の3つのコミュニティが強調表示されている: アルゴリズムと理論(ピンク)、データベース(緑)、HCIとユビキタス(青)。図5はデータベースコミュニティに含まれる発表の場を示したものである。

このように、コミュニティマッピングツールでは、研究コミュニティ群の広がりを



俯瞰しつつ、その中での研究者の位置を容易に把握することができるように可視化を行う。これを用いて、研究者は過去や現在の研究について省察したり、将来の研究の方向性を考えることができる。

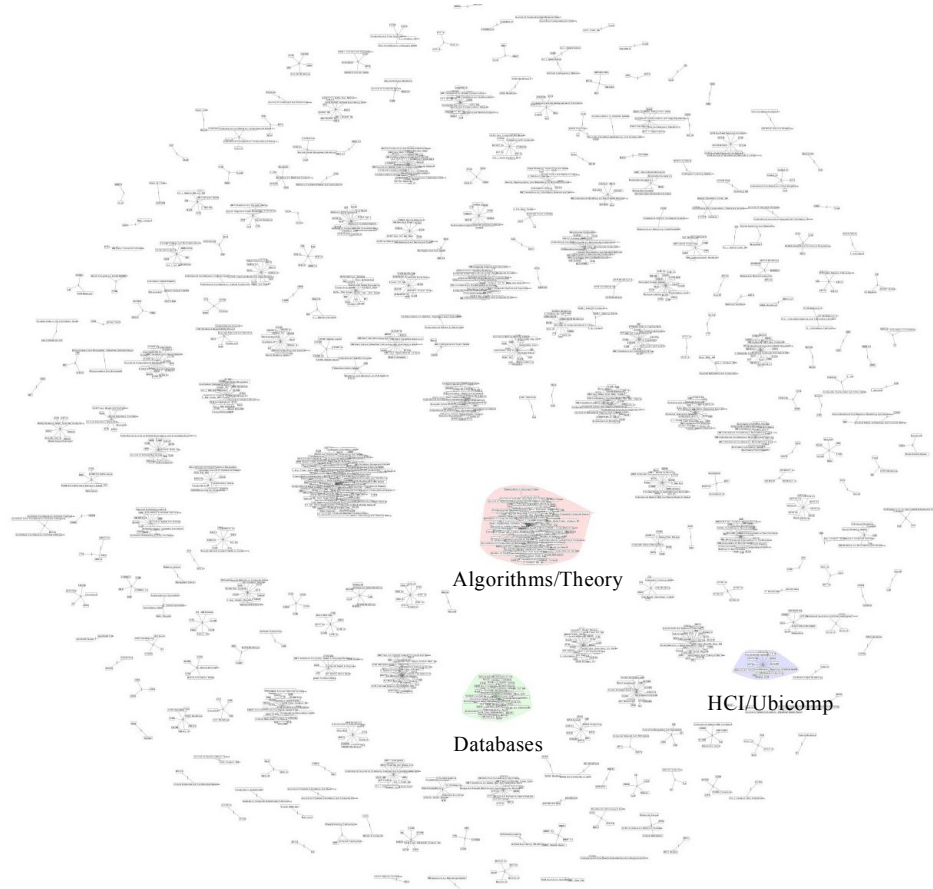


図 4 コンピュータサイエンス分野のコミュニティマッピング

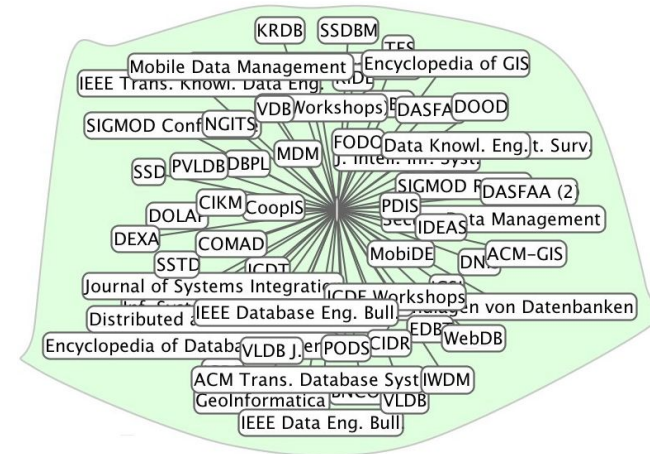


図 5 データベースコミュニティの詳細

## 6. 共同研究の立ち上げ支援

共同研究を立ち上げる際の重要なステップの一つは、研究チームを作ることである。チームメンバーが異なる研究コミュニティに分散しており、研究の実践法や使うツール、語彙、文化が異なる場合、研究チーム作りは容易でない。ここでは特に、過去と一緒に仕事をした経験 4) と重なり (図 1) に着目して、分野をまたがる研究チームの作成を支援する。

研究者が過去に誰と一緒に仕事をしたかについては、提案システムが扱う文献データベースの共著関係から多くの情報が得られる。研究チーム内のコミュニティの重なりについては、以下で定義する参加率  $P(a,c)$  に基づいて取り扱う。

著者  $a$  の全論文数と発表の場  $v_i$  における論文数をそれぞれ  $N(a)$  ,  $N(a,v_i)$  とする。著者  $a$  のコミュニティ  $c$  に対する参加率  $P(a,c)$  は、コミュニティ  $c$  に含まれる全ての発表の場  $v_i$  についての論文数  $N(a,v_i)$  の和を全論文数  $N(a)$  で割ったものである。

$$P(a,c) = \sum N(a,v_i) / N(a) \quad (2)$$

参加率  $P(a,c)$  が閾値  $th$  以上である場合、システムは著者  $a$  とコミュニティ  $c$  を連結する。

コミュニティマッピングツールは、過去の経験と重なるのパターンを両方考慮して、研究者のつながりを可視化することができる。コミュニティの重なりに基づき3人の研究者のつながりを表示した例を図6に示す。文献データベースの共著関係によれば、これらの研究者同士が合同で論文を発表した記録はなく、「過去に共同で仕事をしたことがない」ことになる。しかしながら、実際はこの中の少なくとも二人の研究者は同じ研究プロジェクトと一緒に仕事をした経験があり、社会的なつながりが既に存在していた。つまり、文献のデータベースから得られる共著関係のデータは有用であるものの、研究者の過去の活動を断片的に捉えたものでしかないことに注意しなければならない。コミュニティマッピングが捉えることのできない情報をユーザから取得するインタラクティブなツール等を用いれば、より効果的にチーム作成を支援できる可能性がある。

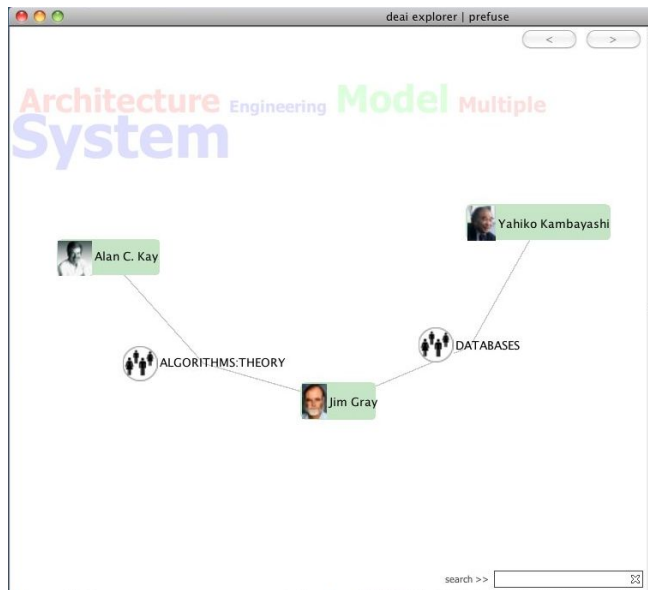


図 6 コミュニティの重なりによる研究者のつながり

新たに学際的な共同研究プロジェクトを立ち上げた場合、その成果をどこで発表すればよいか自明でない場合がある。そのような場合に適切な発表の場を探しやすくするために、発表の場  $v$  に関係するコミュニティの集合  $C(v)$  を以下のように定義する。

$$C(v) = \cup \text{community}(a_i) \quad (3)$$

$C(v)$  は、発表の場  $v$  に属すすべての著者  $a_i$  のコミュニティの和集合である。

図7はDBLP文献データベースにおける発表の場  $v$  とコミュニティの集合  $C(v)$  の関係を示したものであり、横軸はコミュニティの数  $|C(v)|$ 、縦軸は発表の場の数を示している。例えば、グラフのピークの部分を見ると、関連するコミュニティを20個程度持つ発表の場が100以上存在することが分かる。また、少数ではあるが、関連するコミュニティを200個以上持つ「門戸の広い」発表の場が存在することも分かる。このようないくつかの初歩的な分析に基づいて、関連する発表の場を研究チームに推薦する機能をシステムに統合することができると考えている。

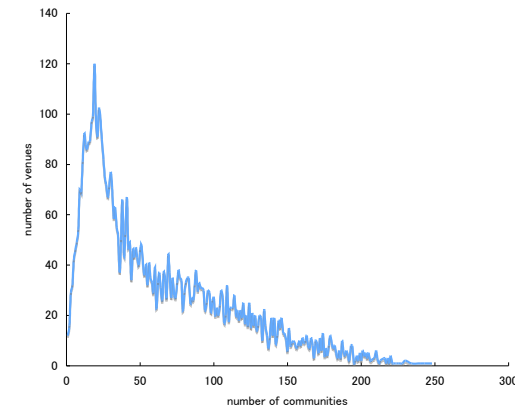


図 7 発表の場のコミュニティ多様性

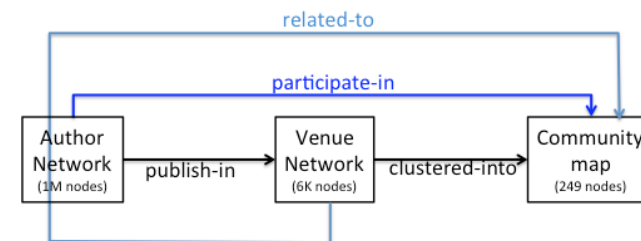


図 8 複数のネットワークとそれらの関連

## 7. むすび

複数の（ソーシャル）ネットワーク構造を体系的に用いてコラボレーションを支援するソーシャルなコンテキストウェアサービスのモデルを確立するために、学際的な共同研究の立ち上げを支援するためのコミュニティマッピングツールを開発した。

コミュニティマッピングにおいては、図8に示すように共著者のネットワークだけでなく、発表の場のネットワーク、コミュニティ、そしてそれらの様々な関係性を考慮しており、共同研究活動に関する省察や将来のプランニングのための可視化や、蓄積された集合的なパターンに基づく発表の場の推薦などが可能である。従って、従来のシステムのような単なる会話のきっかけ作りにとどまらない様々なサービスの提供につながるかと考えている。また、同様の手法を他のユビキタスコンピューティングアプリケーションにも適用できる可能性がある。

## 参考文献

- 1) J. Birnholtz, Y. C. Yuan, and G. Gay, "Bridging social and awareness networks in distributed research collaboration," Proc. ACM CSCW 2010 Workshop on the Changing Dynamics of Scientific Collaborations, Feb. 2010.
- 2) D.T. Campbell, "Ethnocentrism of disciplines and the fish-scale model of omniscience," in Interdisciplinary Relationships in the Social Sciences, M. Sherif and C. W. Sherif, Eds. Chicago: Aldine Publishing Company, 1969, pp. 328-348.
- 3) G. Jr. Chin, and C. S. Lansing, "Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 2004), ACM Press, Nov. 2004, pp. 409-418, doi: 10.1145/1031607.1031677.
- 4) J. N. Cummings, and S. Kiesler, "Who collaborates successfully? prior experience reduces collaboration barriers in distributed interdisciplinary research," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 2008), ACM Press, Nov. 2008, pp. 437-446, doi: 10.1145/1460563.1460633.
- 5) DBLP Computer Science Bibliography. <http://www.informatik.uni-trier.de/~ley/db/>. (accessed: October 19, 2010)
- 6) G. Fischer, "From reflective practitioners to reflective communities," Proc. HCI International Conference (HCHI 2005), Las Vegas, July 2005.
- 7) S. M. van Dongen, Graph clustering by flow simulation, PhD thesis, University of Utrecht, May 2000.
- 8) D. B. Horn, T. A. Finholt, J. P. Birnholtz, D. Motwani, and S. Jayaraman, "Six degrees of Jonathan Grudin: a social network analysis of the evolution and impact of CSCW research," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 2004), ACM Press, Nov. 2004, pp. 582-591, doi: 10.1145/1031607.1031707.
- 9) J. F. McCarthy, D. W. McDonald, S. Soroczak, D. H. Nguyen, and A. M. Rashid, "Augmenting the social space of an academic conference," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 2004), ACM Press, Nov. 2004, pp. 39-48, doi: 10.1145/1031607.1031615.

- 10) S. Konomi, S. Inoue, T. Kobayashi, M. Tsuchida, M. and M. Kitsuregawa, "Supporting colocated interactions using RFID and social network displays," IEEE Pervasive Computing vol. 5, no. 3, pp. 48-56, 2006.
- 11) S. Nomura, J. Birnholtz, O. Rieger, G. Leshed, D. Trumbull, and G. Gay, "Cutting into Collaboration: Understanding Coordination in Distributed and Interdisciplinary Medical Research," Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW 2008)', 2008, pp. 427-436.
- 12) G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," Nature, vol. 435, June 2005, pp. 814-818, doi:10.1038/nature03607.
- 13) Prefuse. <http://prefuse.org/> (accessed: October 20, 2010)
- 14) W. Tan, J. Zhang, and I. Foster, "Network Analysis of Scientific Workflows: A Gateway to Reuse," IEEE Computer (43:9), 2010, pp. 54-61.
- 15) S. Wasserman and K. Faust, Social Network Analysis, Cambridge University Press, 1994.