

無線センサノードを用いた近接関係によるコミュニティ推定手法

江口 洋平[†] 岩井 将行[‡] 瀬崎 薫^{*}[†] 東京大学情報理工学系研究科[‡] 東京大学生産技術研究所^{*} 東京大学空間情報科学研究センター^{*} 科学技術振興機構戦略的創造推進事業

1 はじめに

ソーシャルネットワークの利用は、主にコミュニケーション用としてオンライン上で大きく広がっている。しかし、これらのソーシャルネットワークは主にオンライン上での 1-click 関係により形成されているため、実世界での人と人との関係を全て網羅することはできない。

この問題を解消するために、位置情報や他人との近接関係など、実世界から得られる情報を用いてソーシャルネットワークを生成する手法が報告されている。本稿では、まず近距離無線を用いた人と人との近接関係推定によるソーシャルネットワークの生成手法を整理する。その後、ソーシャルネットワーク内でのコミュニティ推定手法の理論的な検討をする。

この手法により、組織の実態調査や組織内の非公式な関係の抽出等が可能になる。また、小型無線ノードを用いて近接関係を推定することが可能であれば、無線ノードを猿などの動物に装着することにより、それらの動物における社会行動の調査への応用の可能性がある。

本稿では、まず 2 節で用語の定義をした後、3 節で近接関係を用いたソーシャルネットワークの構成と、グラフからのコミュニティ抽出に関する既存の研究を述べる。4 節で近接関係によるソーシャルネットワークの特徴とその定式化を述べる。そして、5 節でそのソーシャルネットワークを用いてコミュニティを推定する手法を検討し、6 節で今後の課題について議論し、まとめる。

2 用語の定義

無線ノードとは、無線通信機能を持った端末であり、一定の半径内の他の端末に対して自端末の ID を広告するものである。また、他端末からの広告を受けた場合、その端末の ID と広告を受けた時刻を端末内の記憶領域に保存する。行為者とは、無線ノードを持っている人物（若しくは動物）のことである。近接関係が存在するとは、ある時刻において 2 人の行為者 i と j

の持つ無線ノードが互いを検出できる位置に存在していることである。この近接関係は、行為者 i と j について対称でなければならない。つまり、行為者 i から行為者 j の近接関係が存在している場合、行為者 j から行為者 i への近接関係も存在していなければならない。なぜならば、行為者 i の無線ノードの ID を行為者 j の無線ノードが検出可能な場合には、行為者 j の無線ノードの ID を行為者 i の無線ノードからも検出可能だからである。したがって、ある 2 人の行為者間の近接関係は無向である。本研究におけるソーシャルネットワークは、行為者とそれらの間の近接関係が与えられた場合に決定される、行為者間の関連の総体のことである。無線ノードにより断続的に近接関係が測定されるので、このソーシャルネットワークは時々刻々と変化していく。本研究では、行為者の性別やその他の特徴といった行為者の属性は扱っていない。コミュニティとはあるソーシャルネットワークの中で、他の行為者の集合と比べて、より強い連結度を持つ行為者の集合とする。本研究の目的は、あるソーシャルネットワークから、コミュニティを推定することが目的である。

3 関連研究

近年、無線端末を用いて近接関係を取得することが可能になっている。近接関係を用いたソーシャルネットワークを生成する試みが存在する。その中で代表的なものについて、それぞれの特徴を整理する。

[1] では、肩に装着するタイプの無線ノードを開発した。以下、この無線ノードを Sociometer と呼ぶ。Sociometer には IR（赤外線）の送受信機とマイクが搭載されている。Sociometer が自分の ID を IR で送信すると同時に、他の端末からの IR を受信することで近接関係を取得している。Sociometer では、近接関係以外に会話の頻度や持続時間などの行為者間の関係の取得も目指したため、会話の情報を取得するためにマイクが搭載されている。[1] では、行為者の会話の頻度や持続時間による解析が主であり、コミュニティの推定は行われていない。

Reality Mining では、Bluetooth による通信機能を組み込んだ携帯電話を用いて、それぞれの被験者の行動や、行為者間の関係性を推定している。近接関係の推測は、他の行為者の端末と Bluetooth による通信が確立したかどうかにより判断している。また、Reality mining では、携帯電話の利用履歴等の行為者の属性も考慮されている。その後、行為者間の近接関係と行為者の属性から、行為者の行動に関する考察がなされている。また、彼らによる後の研究 [3] では、収集された情報から、行為者間の友人関係を推測するなど、詳細な解析が行われている。

A Method for Inferring Communities from Proximity Data using Wireless Sensor Node

Yohei EGUCHI[†], Masayuki IWAI[‡], and Kaoru SEZAKI^{*}[†] Graduate School of Information Science and Technology, the University of Tokyo

153-8505, Tokyo, Japan

[‡] Institute of Industrial Science, the University of Tokyo

153-8505, Tokyo, Japan

^{*} Center for Spatial Information Science, the University of Tokyo

277-8568, Chiba, Japan

^{*} JST CREST

102-0075, Tokyo, Japan

{eguchi, masa, sezaki}@mcl.iis.u-tokyo.ac.jp

あるグラフから、その他の部分よりも連結度の高い頂点の部分集合を抜き出すことを目的とした研究として [4] 等がある。[4] では、頂点数と辺数が共に大きいグラフから、効率的にコミュニティを推定する手法が示されている。

4 ソーシャルネットワークの構成

一般的にソーシャルネットワークはグラフとして扱われる。本節で、近接関係を用いたソーシャルネットワークの特徴とそのグラフとしての定式化を試みる。

いま、 g 人の行為者の集合を頂点集合 $N = \{n_1, n_2, \dots, n_g\}$ で表す。無線ノードによる近接関係の収集を時刻 $0 \leq t \leq T$ の範囲で行ったとする。ある時刻 t を中心とする幅 Δt の時間内に、行為者間の全体に m_t 組の近接関係が存在したとする。 ($0 \leq m_t \leq \frac{g(g-1)}{2}$) 行為者 i と行為者 j の間の近接関係は頂点間の辺 $l_{tk} = (n_i, n_j)$, ($0 \leq k \leq m_t$) として表される。 l_{tk} の集合 $L_t = \{l_{t1}, l_{t2}, \dots, l_{tm}\}$ が、グラフの辺集合となる。 N, L_t を用いて、時刻 t におけるソーシャルネットワークは、グラフ $G_t = (N, L_t)$ として表される。この G_t は $\frac{T}{\Delta t}$ を超えない整数 τ 個だけ生成される。 $1 \leq a \leq \tau$ と、 $t_a = \frac{(2a-1)\Delta t}{2}$ を用いると、 τ 個のソーシャルネットワークは $\mathcal{G} = \{G_{t_1}, G_{t_2}, \dots, G_{t_\tau}\}$ と表される。

5 コミュニティ推定手法

ある G_{t_a} が与えられたとき、その中でのコミュニティの推定には既存の研究を利用することが可能である。既存研究 [4] 等を用いて、 G_{t_a} から α 個のコミュニティ $C_{t_a} = \{C_{t_a1}, C_{t_a2}, \dots, C_{t_a\alpha}\}$ が推定できる。コミュニティは行為者の集合であるから、 $1 \leq \beta \leq \alpha$ を用いて、 $C_{t_a\beta} \subseteq N$ である。例えば、 G_{t_a} が $N = \{n_1, \dots, n_7\}$ とその間の辺として表されたとする。その図を図 1 に示す。図中では、点で頂点を点と点を結ぶ線で辺を表している。この G_{t_a} 中でのコミュニティを N の部分集合で完全グラフとなかつ極大なものをコミュニティと定義する。この定義によると $C_{t_a1} = \{n_1, n_2, n_3\}$, $C_{t_a2} = \{n_4, n_5, n_6, n_7\}$ の二つのコミュニティを推定できる。コミュニティの推定結果はコミュニティの定義方法によって変わるが、全ての G_{t_a} について同様にコミュニティを推定できる。

$\mathcal{G}$ の中での特徴的なコミュニティを推定するためには、 $\mathcal{G}$ の要素それぞれからコミュニティを推定し、それらを全て考慮に入れなければならない。ある $N_s \subseteq N$ について、 $N_s = C_{t_a\beta} = \dots = C_{t_b\gamma}$ だったとする。 $0 \leq a \leq \tau$ の中で、 N_s と等しいコミュニティが出現した回数 σ_s をコミュニティの重複度と呼ぶ。重複度に対し、あるしきい値 σ_{thresh} を設定すると、 $\sigma_{thresh} < \sigma_s$ を満たす N_s は頻出するコミュニティであり、活発度の高いコミュニティを推定することが可能になる。また、ある $N_r \subseteq N$ と定数 δ について、 $N_r = C_{t_a\beta} = C_{(t_a+\delta)\gamma} = \dots = C_{(t_a+d\delta)\epsilon}$ だったとする。このコミュニティ N_r は、ある周期 δ 毎に定期的に現れるものであり、規則正しいコミュニティである。 $\mathcal{G}$ 全体を考慮に入れることで N_s, N_r のように、実世界である特徴を持つコミュニティを推定することが可能になる。

次に、 $\mathcal{G}$ でのコミュニティの時間による変遷を追う。 C_{t_a} と $C_{t_{a+1}}$ の中で、その要素間の関連を調べる。 $C_{t_{a1}}, C_{t_{a+1j}} \in N$ に対してそれらの関連度を、 $R_{ij} =$

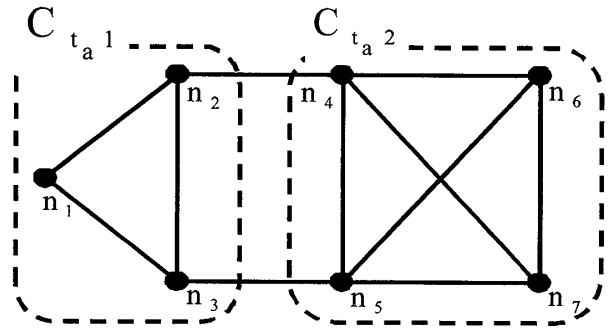


図 1: G_{t_a} と、 C_{t_a1}, C_{t_a2} への分割の例

$\frac{|C_{t_{a1}} \cap C_{t_{a+1j}}|}{|C_{t_{a1}} \cup C_{t_{a+1j}}|}$ と定義する。ただし、 $|\cdot|$ は集合の要素数を示す記号である。あるしきい値 R_{thresh} を用いて、 $R_{thresh} \leq R_{ij}$ であったとき、 $C_{t_{a1}}$ が時間 Δt のうちに $C_{t_{a+1j}}$ に変化したと考える。このように考えると、時間を経たコミュニティの変化を追うことができる。これによって、ある組織内での人間関係の移り変わりの把握等が可能になる。

6 まとめと今後の課題

本稿では、近接関係を用いてソーシャルネットワークを形成することが可能であることを示し、そのソーシャルネットワークにおいてコミュニティを推定する手法について検討した。

本稿では、ある時間幅内における特定の行為者間の近接関係の存在と、ソーシャルネットワーク内の頂点間の辺の存在を対応させていた。しかし、この手法ではその時間幅内にその行為者間の近接関係がただ 1 度検出された場合と多数検出された場合の区別がつかない。従って、 Δt 内での特定の近接関係の検出された頻度から計算される値を辺の重み w_k として採用し、ソーシャルネットワークを重み付きグラフとして扱うことで、より詳細な解析が可能になる。また、 Δt の大きさを 1 日、1 週間、1 ヶ月と変化させることで、推定されたコミュニティの意味合いも変わる。したがって、 Δt に関するより詳細な検討が必要である。

参考文献

- [1] T. Choudhury, et al.: "Sensing and Modeling Human Networks using the Sociometer", Proc. Seventh IEEE International Symposium on Wearable Computers, pp. 216-222, 2003.
- [2] N. Eagle, et al.: "Reality Mining: Sensing Complex Social Systems", Personal and Ubiquitous Computing, Vol. 10, NO. 4, pp. 255-268, 2006.
- [3] N. Eagle, et al.: "Inferring Social Network Structure using Mobile Phone Data", Proc. of the National Academy of Sciences, Vol. 106, No. 36, pp. 15274-15278, 2009.
- [4] A. Clauset et al.: "Finding community structure in very large networks", Physical Review, Vol. 70, Issue. 6, pp. 1-6, 2004