

知的生産集団のコンフィギュレーションの分析 －半導体分野と遺伝子治療薬分野のネットワークの例－

井上寛康

Analyses of Configuration in Groups that Produce Intellectual Property:
Examples of Semiconductor Sector and Gene Remedy Medicine Sector

Hiroyasu Inoue

Abstract

It is generally recognized that technological innovation drives economical advance, and this process is promoted by scientific findings. However, we know little about how scientific findings assist technological innovation. This paper focuses on collaborations that are conducted by scientists and engineers, and investigated mutual affects between them. Concretely, I created two networks for each semiconductor sector and gene remedy medicine sector by using papers and patents data. Then, I analyzed the networks by random graph models for multiple relations. The analysis is logistic regression analysis and uses network configurations as explanatory variables, and emergence probability of networks as explained variables.

キーワード：論文，特許，ネットワーク，科学的知見，発明

Keywords：Paper, Patent, Network, Scientific Knowledge, Invention

目次

1. はじめに
2. 多重ネットワークデータ
3. 多重ネットワーク依存モデル
4. 分析結果
5. 議論
6. 結論

1. はじめに

イノベーションは、新しいアイデアに起因して社会全体が変容することと定義されている [1]。我が国ではイノベーションを技術革新に限定して議論することが多いが、本論文においても同様にイノベーションの技術的な側面について議論したい。

技術的イノベーションが経済発展を支えていることは広く認知されており、それは科学的知見に依拠しているといわれている [2]。しかしながら、そのプロセス、すなわち科学的知見がいかに技術的イノベーションに発展しているかの過程はいまだによくわかっていない。

現代において、技術的イノベーションのプロセス自体に変化が見られていることは、広く認識されているところである。1つにはプロセスのサイクルそのものが短くなっていることがある。ほかに、従来クロズドであったプロセスがオープンに行われるようになった [3] ということがある。クロズドであるとは、研究開発投資、新技術発見、新製品販売、売上げ・利益増加、というサイクルを1つの企業（あるいは企業グループ）内で行うということである。これまでの伝統的なイノベーションのプロセスであり、我が国が得意とする形であった。これに対してオープンであるとは、上記のクロズドプロセスにおける新技術発見の部分に、ある企業と関係のないところで行われていた研究プロジェクトの成果を自在に取り込むということである。

ここで大切であるのは、今世紀に入ってからこのようなオープンイノベーションを行う企業が非常に強い競争力を示していることである。そのような企業に IBM や P&G があげられることが多い。すべての企業にとって、オープンイノベーションである必要があるかどうかは議論の余地があるとして、いったいそのようなプロセスを実現するにはどのような組織を構築すべきかどうかは議論すべき問題である。

本論文では、技術的イノベーションにおいて重要なプレーヤーである、科学者、技術者の間でどのような協業が行われているのかに焦点をあて、その相互の影響の解明に接近する。これにより、上述のような現代的なイノベーションプロセスの理解に寄与できると考える。本論文では、データとして論文と特許を用いる。そして、科学的知見の発見者たる論文著者のネットワーク、およびその具現者である発明者のネットワークを構築する。その上で、それらネットワークで共通するノードを元に、それぞれのネットワークが互いにどのように影響するのかについて分析を行う。その際に、おのおののネットワークの部分的な構造（コンフィギュレーション）が、どのように発生確率に影響しているのかについて、ロジスティック回帰分析する。

2. 多重ネットワークデータ

ネットワークはノードと呼ばれる点とそれをつなぐリンクからなる。本論文が対象とした多重ネットワークデータは2つのネットワークからなり、それぞれ論文著者ネットワー

クと発明者ネットワークである。重要な点は、これらネットワークのノードは、ある個人に対応しており、ネットワーク間で共通していることである。

イノベーションのプロセスを考えると、論文により科学的知見が発見され、特許によりそれが具現化されるというのが順として理解しやすいが、ここではデータの作成の都合から逆の順、すなわち発明者のネットワーク、論文著者のネットワークの順に説明する。なお本論文とは直接関係ないが、特許は公知の事実は申請できないため、論文が純粋に科学的知見でない限りは、特許申請を先に行うのが一般的である。したがって、説明の順という意味ではこちらの方が正しいといえる。

発明者ネットワークの元となるデータは、日本の公開特許公報において1993年1月から2002年12月の10年間に記載された4,998,464件の特許データであり、TamadaDatabase [4] を利用する。発明者ネットワークにおいて、ノードは発明者である（ただし論文著者ネットワークと共通である。）。また、リンクは発明者の間で共同で1つでも特許が出願されていれば形成される。

図1は、発明者ネットワーク構築の様子である。黒丸は発明者であり、ノードである。それらをつなぐリンクは、それらノードの間で1つでも特許が申請されていた場合に引かれる。したがって、リンクについては多重度を考慮せず、リンクが存在するかしないかの2値である。これは後述のロジスティック分析の適用にも少なからず関係する。国内の発明者のネットワークはすでに構築済みであるが、論文著者ネットワークのデータについては、すべてを保有していないため、このネットワークの一部を用いる。これについては後述する。

次に、論文著者ネットワークの元となるデータは、国立情報学研究所が提供する CiNii [5] において、2009/02/15時点での論文データであり、12,039,089件の論文が対象である。論文著者ネットワークにおいて、ノードは論文著者である（ただし発明者ネットワークと共通である。）。また、リンクは論文著者の間で共同で1つでも論文が執筆されていれば形成される。

ここまで論文著者ネットワークおよび発明者ネットワークをそれぞれ述べたが、次に、どのようにして発明者ネットワークと論文著者ネットワークのノードを一致させるのかについて述べる。これにはまず、(1)発明者ネットワークからある部分のノードを取り出す。(2)それらの間のすべての2者の組み合わせを生成する。(3)2者の氏名が同一の論文の著者となっていれば、それらのノードの間に論文著者ネットワークのリンクがあるとす。ここで1つの問題は、同じ氏名の人物がいるという可能性である。確かに同じ氏名の人物は大量に存在するが、同じ氏名を持つ2組の2者が論文を作成する可能性は、きわめ

て低いことから無視することができる。より詳しく説明すると、同じ氏名の別人は国内に数名存在するが、2者が揃って別人であるような論文は、単純な計算をすれば（同一氏名の共著者）/（全論文著者）になるためである。

本論文では論文著者ネットワークおよび発明者ネットワークの全体を対象とはしない。本論文では、科学的知見を元に発明に至っている、数十人程度の部分ネットワークを2つ抽出した。

1つは同志社大学教授山口栄一氏を中心とした73名からなるネットワークである。同氏は固体物理学が専門であり、(株)パウデック、Algan(株)などを起業した論文著者・発明者である。同氏から3ステップで到達できるノードは発明者ネットワークにおいて72であり、これら73ノードの間の論文著者ネットワークを取得した。これは半導体分野を代表する1つのネットワークといえる。以後このデータを Yamaguchi-3 とする。

もう1つは大阪大学客員教授森下竜一氏を中心として127名からなるネットワークである。同氏は臨床遺伝子治療が専門であり、アンジェス MG(株)などを起業した論文著者・発明者である。同氏から2ステップで到達できるノードは発明者ネットワークにおいて126であり、これら127ノードの間の論文著者ネットワークを取得した。これは遺伝子治療薬分野を代表する1つのネットワークといえる。以後このデータを Morishita-2 とする。

3. 多重ネットワーク依存モデル

本論文で知りたいことは、論文著者と発明者という異なる関係性を示すネットワークの間でどのような関係性が形成されているかである。本論文では、Random Graph Models for Multiple Relation[6](多重ネットワーク依存モデル)を用いる。これは p^* に基づくネットワーク分析[7]を多重ネットワーク分析に拡張したものである。この p^* モデルとは、

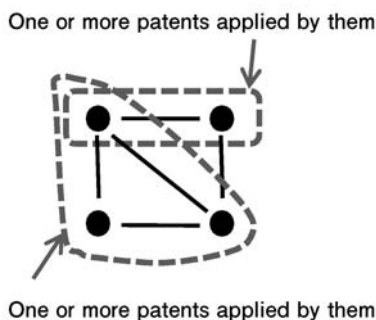
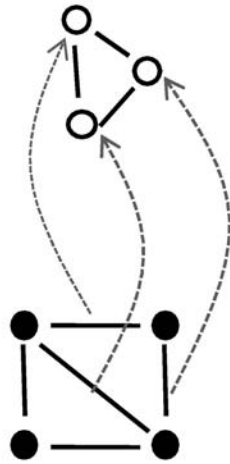


図1：発明者ネットワーク構築の様子

Dependence network (model)



Actual (authors' or inventors') network

図2：依存ネットワーク

ネットワーク x (あるノード集合上のリンクの組み合わせ) は, 他のリンクに依存して決まるような確率分布に従って出現したとすると, 次のように表すことができる。

$$P(x) = \kappa^{-1} \exp\left(\sum_{A \subseteq N_D} \lambda_A z_A(x)\right) \quad (1)$$

ここで $\kappa = \sum_x \exp(\sum_{A \subseteq N_D} \lambda_A z_A(x))$ は正規化定数, A はノードの部分集合, N_D は依存ネットワーク (後述) のノード, $z_A(x) = \prod_{x_{ijm} \in A} x_{ijm}$ であり, x_{ijm} はネットワーク m におけるノード i から j へのリンクである。ネットワーク m というのは, ネットワークの種類を表しており, この場合は論文著者ネットワークと発明者ネットワークを表している。

依存ネットワークとはネットワークのリンクをノードとするグラフである。図2は依存ネットワークの例である。下の黒丸のネットワークが実在のネットワークであり, 上の白丸のネットワークが依存ネットワークである。依存ネットワークは実在のネットワークのリンクをノードとしている。依存ネットワークのリンクは, それらの間で相関があることを意味する。これはモデルであり, この依存ネットワークを多重ネットワークの間で求めることが, この論文の目的といえる。

前述したように, 式 (1) における λ を求めることが本論文の目的になるが, このままでは λ の数が多すぎ, 計算ができない。そこで, 同位性を用いてその数を減らす。す

なわち、依存ネットワークにおける同じコンフィギュレーションはすべて同じ効果とみなすことになる。コンフィギュレーションとは前述したように、ネットワークにおける部分的な構造である。本論文で扱うコンフィギュレーションは Choice, Multiplexity, Role interlocking, Transitivity とする。これを表したものが図3である。これらを選んだのは、先行研究 [6] における基本的な有向グラフのモデルに合わせたためであり、本論文は無向グラフを扱うため、修正を施してある。これ以上に複雑なモデルはいくらでも考えられるが、本論文ではまずこの簡単な構造について分析を行う。

Choice は、2者の間でのリンクの発生はそれぞれのネットワーク（ここでは論文著者ネットワークと発明者ネットワーク）で関係がないとするモデルである。前述したように同位性を考慮するので、 λ はそれぞれのネットワークに1つずつであり、これらがすべてである。このモデルを調べることで、個々のネットワークを別々のモデルとすることに意味があるかがわかる。Multiplexity は、Choice とは異なり、2者の間でのリンクの発生は2つのネットワークで同時に起きるとするモデルである。この場合の λ は1つだけである。このモデルを調べることで、2つのネットワークにおいて、リンクが重複する傾向にあるかわかる。Role interlocking は、3者（A, B, Cとする）が関係しており、AとBが論文を共著、BとCが特許を共願するようなモデルである。この場合の λ は1つだけである。このモデルを調べることで、ある人を中心として論文と特許のパートナーがそれぞれ存在する傾向であるかがわかる。Transitivity は、3者（A, B, Cとする）が関係しており、AとB, AとCが論文を共著、BとCが特許を共願するようなモデルである。この場合の λ は2つある。このモデルを調べることで、ある人を中心に論文を書く3人のグループがあるが、その周りの2人で特許を出願するような傾向にあることがわかる。これの論文と特許を交換した場合も同様である。

一般的に式（1）を同位性を考慮した形に直すと、

$$P(x) = \kappa^{-1} \exp\left(\sum_{[A]} \lambda_{[A]} z_{[A]}(x)\right) \quad (2)$$

となる。この式を用いて実在の多重ネットワークが尤度最大になるように λ を求めればよい。この尤度を求める方法には、最尤法 [8,9], マルコフ連鎖モンテカルロ法 [7] がある。前者は依存ネットワークの構造が複雑なため、この場合計算できない。後者は比較的最近提案された推測値を出す方法であるが、まだ十分に確立しているとはいえない。一方で、これまでに多数の分析に用いられ、その有効性が示されている擬似尤度法 (Pseudolikelihood function) [10, 11, 12, 13] があり、本論文ではこれを用いる。

準備として、下記の3つの記法を定義する。

$$\begin{aligned} X_{ijm}^C &= \{X_{klh} : X_{klh} \in N_D \text{ and } (k, l, h) \neq (i, j, m)\} \\ x_{ijm}^+ &= \{x_{klh}^+ : x_{klh}^+ = x_{klh}^+ \text{ for all } (k, l, h) \neq (i, j, m) \text{ and } x_{ijm}^+ = 1\} \\ x_{ijm}^- &= \{x_{klh}^- : x_{klh}^- = x_{klh}^- \text{ for all } (k, l, h) \neq (i, j, m) \text{ and } x_{ijm}^- = 0\} \end{aligned}$$

その上で、あるリンクが存在するようなネットワーク X_{ijm} が出現する条件付確率は、

$$\begin{aligned} P(X_{ijm} = 1 \mid X_{ijm}^C) &= \frac{P(X = x_{ijm}^+)}{P(X = x_{ijm}^+) + P(X = x_{ijm}^-)} \\ &= \frac{\exp \sum_{[A]} \lambda_{[A]} z_{[A]}(x_{ijm}^+)}{\exp \sum_{[A]} \lambda_{[A]} z_{[A]}(x_{ijm}^+) + \exp \sum_{[A]} \lambda_{[A]} z_{[A]}(x_{ijm}^-)} \end{aligned}$$

これによってオッズ比を表すと、

$$\begin{aligned} \omega_{ijm} &= \log \frac{P(X_{ijm} = 1 \mid X_{ijm}^C)}{P(X_{ijm} = 0 \mid X_{ijm}^C)} \\ &= \sum_{[A]} \lambda_{[A]} [z_{[A]}(x_{ijm}^+) - z_{[A]}(x_{ijm}^-)] \\ &= \sum_{[A]} \lambda_{[A]} (d_{[A]})_{ij}^m \end{aligned}$$

擬似尤度法は擬似的に尤度を、

$$PL(\lambda) = \prod_{i \neq j} \prod_{m=1}^r P(x_{ijm} = 1 \mid x_{x_{ijm}}^C)^{x_{ijm}} P(x_{ijm} = 0 \mid x_{x_{ijm}}^C)^{1-x_{ijm}}$$

により求めるとする方法である。この擬似尤度を最大化する係数 λ はロジスティック回帰分析により求まることが証明されている [12]。

次節にて、この λ を図3のようなChoice, Multiplexity, Role interlocking, Transitivity について設け、論文著者ネットワークと発明者ネットワークにおいて、それらコンフィギュレーションが有意に起きているのかについて調べる。

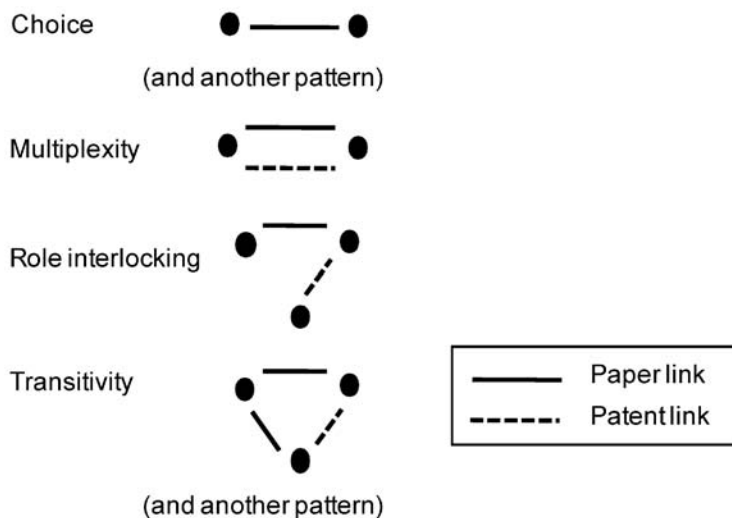


図3：ネットワーク依存パタンの種類

4. 分析結果

前節で述べた擬似尤度法をロジスティック回帰分析により行った。この分析には SPSS 17.0を用いた。表1は分析結果を表している。Yamaguchi-3および Morishita-2のデータに対して、多重ネットワーク依存モデルの各コンフィギュレーションに対する（擬似）-2対数尤度 (G_{PL}^2) およびそのときの説明変数の数を示している。一般的に回帰モデルは説明変数が増えるほど G_{PL}^2 が小さくなるので、その有意な差を決める必要がある。そのような値として用いられるのは $-2n(n-1)r \log(1-\delta)$ である。ここで n はノードの数、 r はネットワークの数、 δ は定数であり、一般的に0.001や0.005が用いられる。この値はそれぞれのデータについて、105.4、320.8である。

5. 議論

表1を見ると、どちらのデータの場合も Patent choice, paper choiceの方が Patent choice=paper choiceより優れている。すなわち特許と論文のリンクについて、発生する確率は別々と捉えるほうが、モデルの精度がよい。（-2対数尤度は正の数であり、小さいほうがよいモデルである。）これはそれぞれのリンクの発生確率が異なることに単に起因する。重要であるのは、Model 1と2の-2対数尤度の差の大きさである。Yamaguchi-3については前述した目安の閾値を超えているが、Morishita-2については、超えていない。

表 1 : 多重ネットワークモデル回帰結果

<u>Yamaguchi-3</u>		
Model	G_{PL}^2	# of variables
1. Patent choice, paper choice	3006.7	2
2. Patent choice = paper choice	3353.7	1
3. 1 + Multiplexity	2932.5	3
4. 3 + Role interlocking	2826.0	7
5. 4 + Transitivity	2328.6	11

<u>Morishita-2</u>		
Model	G_{PL}^2	# of variables
1. Patent choice, paper choice	9984.7	2
2. Patent choice = paper choice	10048.9	1
3. 1 + Multiplexity	8629.1	3
4. 3 + Role interlocking	7268.7	7
5. 4 + Transitivity	5856.1	11

(有意な差とはいえない。) これは、1つの解釈としては、半導体分野と遺伝子治療薬分野を比較すると、論文と特許の生産を両方行うグループの存在は後者の方がより可能性が高いということになる。

このモデル1に対して、Multiplexityを足したのがモデル3であるが、ここでも結果は異なる。Yamaguchi-3の方はモデル1と3が有意な差でないのに対して、Morishita-2の方は有意な差になっている。これは、論文と特許の生産を両方行うグループの存在が高いという前述の内容を補強している。このモデル3が有意であるということは、1つのネットワークでのリンクの存在が、他方のネットワークでの同じノード間のリンクの存在を刺激することを示している。モデル1, 2, 3のこの結果は、技術分野によって協業の傾向が異なることを明確に表している。

つづくモデル4のRole interlock およびモデル5のTransitivity はどちらのデータにおいても有意という結果である。モデル4を解釈すれば、ある人物がいて、論文か特許を出すパートナーがいる場合、それと異なる生産(論文なら特許, 特許なら論文)を行う別のパートナーがいる確率が高いということになる。すなわち、周囲に2つの集団(集団か1人かはわからない)が異なる役割を果たしており、適宜参加する集団を替えているともいえる。

さらに、モデル5を解釈すれば、3人が集団を構築する際、ある人を中心に論文あるいは発明が行われるが、それをともに行った別の2人が、それと異なる生産を行う可能性が高いということになる。モデル5はモデル4との対比で考えるとわかりやすい。モデル4では論文と特許の両方を生産する（リーダー的な）人がいる形であるが、モデル5では論文か特許かのどちらかを専門的に生産する（エキスパート的な）人がいて全体として両方が生産される形である。

最後に本論文の結果の意義について述べる。複数のネットワークの関係性を扱ったこれまでの研究は、主に Multiplexity の分析のみであった。この Multiplexity は、ネットワークの重複度合いの観察と言い換えることができる。本論文の結果からわかるように、Multiplexity は分野によってその様子が異なる。そのため、ある集団の結果に対して、この重複度合いだけを観察するような単純な分析をした場合、科学的発見と発明を行う集団は協業する／しないという誤った結論を導くことになる。これに加えて、本論文での重要な示唆は、集団における2人の関係を単純に観察しただけでは、協業の実際を見ることはできないということである。本論文で見たように、3人が関係するような関係（コンフィギュレーション）が科学的発見と発明での協業において有意に現れていることが、本論文において初めて明確にされた。

また上記の知見に加えて、経営学的な側面から見た本論文の知見は、以下の2つに集約できる。（1）近年ネットワーク分析は急速に発展しており、その経営学への応用が積極的に展開されている。多重ネットワーク分析はネットワーク分析の中でもカッティングエッジにあり、その分析手法が経営学的内容に適用可能であることを示した。（2）前項でも述べたとおり、経営学の分野でのネットワーク分析は盛んに行われているが、たとえば本論文の多重ネットワークを用いれば、企業ネットワークにおける次のような関係、株式保有・役員派遣・取引・（特許共願、論文共著などの）共同研究・コンソーシアムの参加などがどのように関係しあっているのかを議論できる。このことは、特定の企業が外部の能力を十分に活用しているかの判断に用いられる等、企業の価値を測る上での新たな指標になりうることを示しているといえる。

6. 結論

技術的イノベーションが経済発展を支えていることは広く認知されており、それは科学的知見に依拠しているといわれている。しかしながら、そのプロセス、すなわち科学的知見がいかに技術的イノベーションに発展しているかの過程はいまだによくわかっていな

い。本論文では、科学者、技術者の間でどのような協業が行われているのかに焦点をあて、その相互の影響の解明に接近した。具体的には、科学的知見の発見者たる論文執筆者とその具現者である発明者において、その協業ネットワークを抽出し、これらがどのように影響しあうかについて多重ネットワーク分析を行った。具体的には、これらネットワークの間で、コンフィギュレーションを説明変数としたロジスティック回帰分析を行い、どのようにネットワークが相互に関係するか分析した。

分析では、半導体分野と遺伝子治療薬分野を比較した。論文と特許の生産を両方行うグループの存在は、後者の方がより可能性が高かった。同様に Multiplexity は遺伝子治療薬分野が高かった。このように技術分野によって協業の傾向が異なることを明確に示した。

Role interlock および Transitivity によって、論文と特許の両方を生産する（リーダー的な）人がいる形、および論文か特許かのどちらかを専門的に生産する（エキスパート的な）人がいて全体として両方が生産される形が有意に現れることがわかった。

これまでの研究は Multiplexity の分析のみであった。本論文の結果からわかるように、Multiplexity は分野によってその様子が異なるため、そのような分析は十分ではないとわかった。また、3人が関係するようなコンフィギュレーションが科学的発見と発明での協業において有意に現れていることが、本論文において初めて明確にされた。

謝辞

本研究を進めるにあたり、国際電気基礎技術研究所の相馬亘氏、関西学院大学大学院の玉田俊平太氏に議論いただいた。ここに謝意を記す。本研究は科研費（20730268）の助成を受けたものである。

参考文献

- [1] J.A. Schumpeter (塩野谷祐一ほか訳). 経済発展の理論: 企業者利潤・資本・信用・利子および景気の回転に関する一研究. 岩波書店, 1977.
- [2] E. Mansfield. Academic research and industrial innovation. *Research Policy*, 20 (1), 1991.
- [3] H.W. Chesbrough. *Open innovation*. Harvard Business School, 2003.
- [4] S. Tamada, Y. Naitou, F. Kodama, K. Gemba, and J. Suzuki. Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68 (2):289-302, 2006.

- [5] Cinii ホームページ . <http://ci.nii.ac.jp>.
- [6] L.M. Koehly and P. Pattison. Random graph models for social networks: Multiple relations or multiple raters. In P.J. Carrington, John Scott, and Stanley Wasserman, editors, *Models and Methods in Social Network Analysis*, chapter 9. Cambridge University Press, New York, 2005.
- [7] P.J. Carrington, John Scott, and Stanley Wasserman, editors. *Models and Methods in Social Network Analysis*. Cambridge university press, 2005.
- [8] S. Wasserman. Conformity of two sociometric relations. *Psychometrika*, 52:3-18, 1987.
- [9] O. Frank and K. Nowicki. Exploratory statistical analysis of networks. In J. Gimbel, J.W. Kennedy, and L.V. Quintas, editors, *Quo Vadis Graph Theory? A Source Book for Challenges and Directions*. Amsterdam, 1993.
- [10] J.E. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179-195, 1975.
- [11] J.E. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Association*, 47:77-92, 1977.
- [12] D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85:204-212, 1990.
- [13] P. Pattison and S. Wasserman. Logit models and logistic regressions for social networks: Ii. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169-193, 1999.