

コミュニケーションから読み解く  
意思決定の見える構造

筑波大学審査学位論文（博士）

2008

前 野 義 晴

筑波大学大学院  
ビジネス科学研究科 企業科学専攻

## 概要

本研究は、観測データに含まれない、見えない構造を探る問題に取り組む。以下の3点を本研究の課題として問題提起する。

1. 観測にかからない、見えない重要な構造の特徴を考察し、ネットワーク上の空白地帯にノードが隠れている可能性を定量的に示すアルゴリズムを開発すること。

2. アルゴリズムをもとに、人間の解釈・事前理解と相互作用しながら、見えない重要な構造に対する仮説を進化させる成長型のプロセスを開発すること。

3. 個別の問題に対して、見えない重要な構造を理解し、仮説を探索的に導出できるか検証し、個別の問題に適した応用プロセスを開発すること。

見えない構造についての仮説を創出する手法として、ヒューマンインタラクティブアニーリングを提案する。ヒューマンインタラクティブアニーリングでは、人間の事前理解と観測データへの認知とを活用し、それらのずれから気づきを獲得し、仮説を創出する。コミュニケーション、意思決定などの観点から、人間の思考や行動にかかわる、見えない構造を探るのに適した手順を明らかにしていく。

はじめに、見えない構造への足がかりとして、ヒューマンインタラクティブアニーリングのプロセスの提案とシミュレーションによる基本的な特性の研究について述べる。ヒューマンインタラクティブアニーリングは、クラスタ間の隙間についての仮説を創出するプロセスを体系化したものである。隙間とは、関連性の高い事象群（クラスタ）の間の空白地帯でありながら、クラスタとの関係性が観測された領域である。参加者は、空白地帯を含むシナリオマップを描きながら、シナリオを創出する。自己の事前理解（誤解）を確認しながら、シナリオマップと仮説とを同時に作成するプロセスに特色がある。隙間を探すためのアルゴリズムを開発した。ネットワーク上のコミュニケーションから、見えないハブノードを探すシミュレーションでは、事前知識やネットワーク構造がゆらいでも、良好な性能が得られることを示した。見えないノードの確からしさを評価する様々なランキング関数の性能、コミュニケーション強度に対する依存性についても検証した。

次に、組織的な意思決定を取り巻く、見えない構造を探る問題に取り組んだ。20名の参加者で、意思決定の隠れた黒幕を探る実験を行った。組

織の意思決定を目的とした電子メールによるコミュニケーションを観測データとした。黒幕は、電子メールを使用せず、観測データに残らない方法で参加者に指示を与え、意思決定に影響を与える。この見えない黒幕を探る実験において、有効な仮説を創出できた。提案するプロセスに従って、事前知識によると独立性の高そうな参加者クラスタ間に、クラスタを結びつけるダミーノードが出現した。このヒントを解釈し、事前理解と観測データへの認知とのずれから気づきを獲得し、事前知識には欠けていた仮説を創出した。ヒューマンインタラクティブアニメーションが、組織のコミュニケーションと意思決定を誘導する、見えない黒幕を探るための有効な情報を与えることを示した。

9/11 テロ組織についての文献の情報をもとに、実行犯を支援した共謀者についてのネットワークを分析した。トポロジから見たグローバル (peripherarity, partiality)・ローカル (degree, clustering coefficient) な重要性指標を導入した。重要性比率によって、共謀者が2種類に分類できることを見出した。タイプ1の共謀者は、文字通り、クラスタ間の bridge の役割を行う。タイプ2の共謀者は、外部世界との bridge の役割を担う。このネットワークに基づくシミュレーションを行った。タイプ1の共謀者について、標準偏差を用いたアルゴリズムが効果的に機能することを示した。一方、タイプ2の共謀者を探ることは難しく、関連性の高い上位2つのクラスタに注目すると性能が向上するものの、不完全であることがわかった。

応用として、グループ討議で好みについての行動・心理への理解を深める問題に取り組んだ。アートの好みに対するアンケートを題材とし、二部グラフ構造を持ったアンケートの回答結果を取り扱えるよう、アルゴリズムを改良した。ヒューマンネットワーク形式のシナリオマップからコミュニケーションの活性化シナリオを創出した。シナリオの品質を提案可能性、気づきにくさ、成長可能性で評価すると、クラスタ全体をつなぐシナリオよりも、個々のクラスタに着目した細かい粒度のシナリオの方が、具体的でよいことがわかった。さらに、作品群クラスタ形式のシナリオマップから潜在的な好みを表出するシナリオを創出した。ダミー事象から、新しい要因を追加したシナリオを作成できた。また、グループ討議によって個人の理解を深める手法として、視覚化・言語化によって、自己の認知を再確認し、解釈の枠組みを再構築することを支援するプロセスに改良した。気づかなかった好みの発見、心理的側面への言及、好みの具体化といった、被験者ごとに異なる発見が得られた。



# 目次

1	問題提起	1
2	関連研究	7
2.1	まれな事象の解明	7
2.2	確率的推論	9
2.2.1	推論の諸手法	9
2.2.2	多変量解析と関連手法	11
2.3	ネットワーク分析	13
2.3.1	複雑ネットワーク	14
2.3.2	社会ネットワーク分析	15
2.4	応用問題	16
2.5	まとめ	17
3	ヒューマンインタラクティブアニメーリング	20
3.1	事前理解と観測のはざまに潜む発見の契機	21
3.1.1	シナリオマップ	21
3.1.2	参加者間の相互作用	24
3.2	理解の粒度に応じた成長型の繰り返しプロセス	26
3.3	見えない事象を表出化するアルゴリズム	29
3.3.1	クラスタリング	29
3.3.2	ランキング関数	32
3.4	アルゴリズムの精度と安定性	34
3.5	まとめ	39
4	ネットワーク上の観測できないノードの発見	45
4.1	不均一な組織構造	45
4.1.1	ネットワークモデル	45
4.1.2	コミュニケーション強度を特徴付ける指標の導入	48
4.1.3	精度のランキング関数に対する依存性	48
4.1.4	精度のノードに対する依存性	50
4.1.5	精度のコミュニケーション強度に対する依存性	50
4.2	均一な組織構造	55
4.2.1	ネットワークモデル	55
4.2.2	精度のノードに対する依存性	55

4.3	まとめ	55
5	意思決定の隠れた黒幕を探る実験	59
5.1	電子メールによる組織の意思決定の方法	59
5.2	影響力のある見えない重要人物を探る実験	60
5.3	組織の意思決定での見えない黒幕を探る実験	62
5.3.1	黒幕からの指示内容の一致性	62
5.3.2	黒幕からの被指示者の一致性	68
5.4	まとめ	69
6	ケーススタディに基づく共謀者の探知	70
6.1	9/11 テロ組織のネットワーク構造の分析	70
6.2	人物のグローバル・ローカルな重要性指標の導入	72
6.3	重要性指標による共謀者の分類	73
6.4	隠れた共謀者を探るシミュレーション	74
6.4.1	ランキング関数に対する依存性	79
6.4.2	ネットワーク構造の事前知識に対する依存性	79
6.4.3	コミュニケーションの伝播距離に対する依存性	79
6.4.4	共謀者の分類に対する依存性	83
6.5	仮説の創出に向けたシナリオマップ	83
6.6	まとめ	86
7	グループ討議から読み解く個人の好み	88
7.1	二部グラフ構造へのアルゴリズムの改良	88
7.2	コミュニケーションの活性化シナリオの創出	90
7.3	潜在的な好みを表出するシナリオの創出	96
7.4	グループ討議によって深める好みの理解	100
7.4.1	グループ討議のプロセス	100
7.4.2	実験条件	102
7.4.3	プロセスの評価	107
7.4.4	気づきに役立った論点	107
7.4.5	気づきの特徴	110
7.5	まとめ	112
8	結論	113

<b>A</b>	<b>最尤法による解法</b>	<b>129</b>
A.1	観測モデル . . . . .	129
A.2	尤度とランキング . . . . .	130
A.3	単純な例と応用展開 . . . . .	132
<b>B</b>	<b>テキスト解析・発想支援への応用</b>	<b>135</b>
B.1	特許からの発想支援 . . . . .	135
B.2	特許テキストの解析 . . . . .	137
<b>C</b>	<b>公表済みの研究成果</b>	<b>141</b>
C.1	論文投稿 . . . . .	141
C.2	出版 . . . . .	141
C.3	学会投稿 . . . . .	142

# 1 問題提起

はじめに、本研究が取り組む問題の位置づけを述べる。本章では、心理、意思決定、組織行動の见えない構造をめぐる問題と、コミュニケーションから见えない構造を読み解くためのアプローチとを提起する。ビジネスにおいては、人、モノ、金、情報が4つの経営資源と言われている。近年では、知識創出型のビジネスモデルへの移り変わりと共に、人と情報の重要性が増している。個人や組織の意思決定に重大な影響を与えている、観測しにくい潜在的な構造を探ることが、新たな問題として認識されるようになった。これらを見えない構造と総称する。

いくつか、具体的な問題の例を述べる。

人が見えない構造となる場合には、組織に重要な影響を与えている隠れたメンバの特定、グループの行動を決定する、見えないキープレイヤーの推定などの問題が該当する。例えば、テロリスト集団や犯罪組織には、背後に組織を動かす、見えない重要人物が潜んでいる。本人が意識しないまま、見えない重要人物となる場合もある。営業部門の社員が、日ごろ暖めているアイデアを知人や同期の開発部門の社員たちに吹き込むとしよう。これがヒントとなり、開発プロジェクトが立ち上がることもあるだろう。開発部門をサポートする企画部門にとっては、営業部門の社員は見えない重要人物である。また、9/11事件（2001年9月のアル・カイダによる同時多発テロ）以降、リスク管理の対象としての優先度が急速に高まったテロ活動においては、一般の人間にまぎれた共謀者がテロ活動の根幹を支えていると言われている。共謀者は、テロの実行犯に対して、資金や武器を提供するだけでなく、仲間を紹介したり計画立案を支援したりする。共謀者は、コミュニケーションやミーティングの監視にかからないことが多く、その存在を突き止める方法が求められている。このように、観測された組織の活動には、観測にかからない、見えない重要人物の意向が働いている場合がある。組織の見えない重要人物は、潜在的な機会と脅威を読み解き、有効な意思決定を導く重要な鍵になる。見えない重要人物を探る新たな方法の開発が必要である。

情報が見えない構造となる場合には、ライバル企業の水面下での動向の推測、業界外の未知の企業の重要性を見出す仮説の導出などの問題が該当する。特に、企業の技術開発では、サブマリン特許のようなライバル企業が隠している技術資産の特定、潜在的な可能性を秘めたニッチなテクノロジー要素の特定といった問題が、リスク管理や研究開発戦略の策定の上で重要である。技術開発の面だけでも、企業や技術者に有用な情



報は特許文献だけに限らず，実に多様である．メーリングリストの履歴，インターネット上でアクセス可能な情報ソース，論文などのデータベース，学会・展示会・標準化会議の記録といった実世界のさまざまな情報が利用できる．このような容易に利用可能な情報から，欠落した情報を復元し，水面下での動向に対する解釈・理解を進化させ，有効な仮説の創出と仮説を現実にするための効果的な意思決定を導く新たな方法の開発が必要である．

本研究の狙いは，「意思決定の背後で作用している，直接的にはデータとして観測できない空白地帯に潜む，コミュニケーションや人間関係の見えない重要な構造を探る新しい手法を開発し，他者の意思決定，自己の意思決定に係わる構造についての理解を深めること」にある．

ここで，この問題の数理的な位置づけを明らかにしておく．組織をネットワークでモデル化すると，見えない重要人物を見つける問題は，ネットワークの中でのノードを見つける問題となる．同様に，情報や概念の体系をネットワークでモデル化すると，欠落した情報や新しい概念を見つける問題は，やはり，ネットワークの中でのノードを見つける問題となる．このようなモデルを用いると，本研究で取り組む問題は，ネットワークのノード発見問題と位置付けられる．コミュニケーションの履歴から見えない重要人物を探るケースを例に，図1にノード発見問題の要点を図示する．

はじめに，観測された見える構造  $S_{\text{vis}}$ ，観測できない見えない構造  $S_{\text{inv}}$ ，コミュニケーションのモデル  $M$  についての事前仮説  $H$  を設定する．仮説は，式(1)で示される．

$$H = S_{\text{vis}} \cup S_{\text{inv}} \cup M. \quad (1)$$

多くの場合に，見えない構造についての事前仮説は，明確に形式化されたものではない．観測の過程で捕えることができない何らかの構造がある，といった直感のようなものである．

見える構造  $S_{\text{vis}}$ ，見えない構造  $S_{\text{inv}}$  とは，式(2)におけるグラフ  $G$  についての知識を指す．

$$S_{\text{vis}} = G_{\text{vis}}(p_i, r_{ij}). \quad (2)$$

ノードは，コミュニケーションの主体となる人物  $p_i$  を表す．識別子  $i$  によって，個々の人物を識別する．リンクは，人物  $p_i$  と  $p_j$  との関係性を表す．リンクがあると ( $r_{ij} = 1$ )，コミュニケーションがある，リンクが

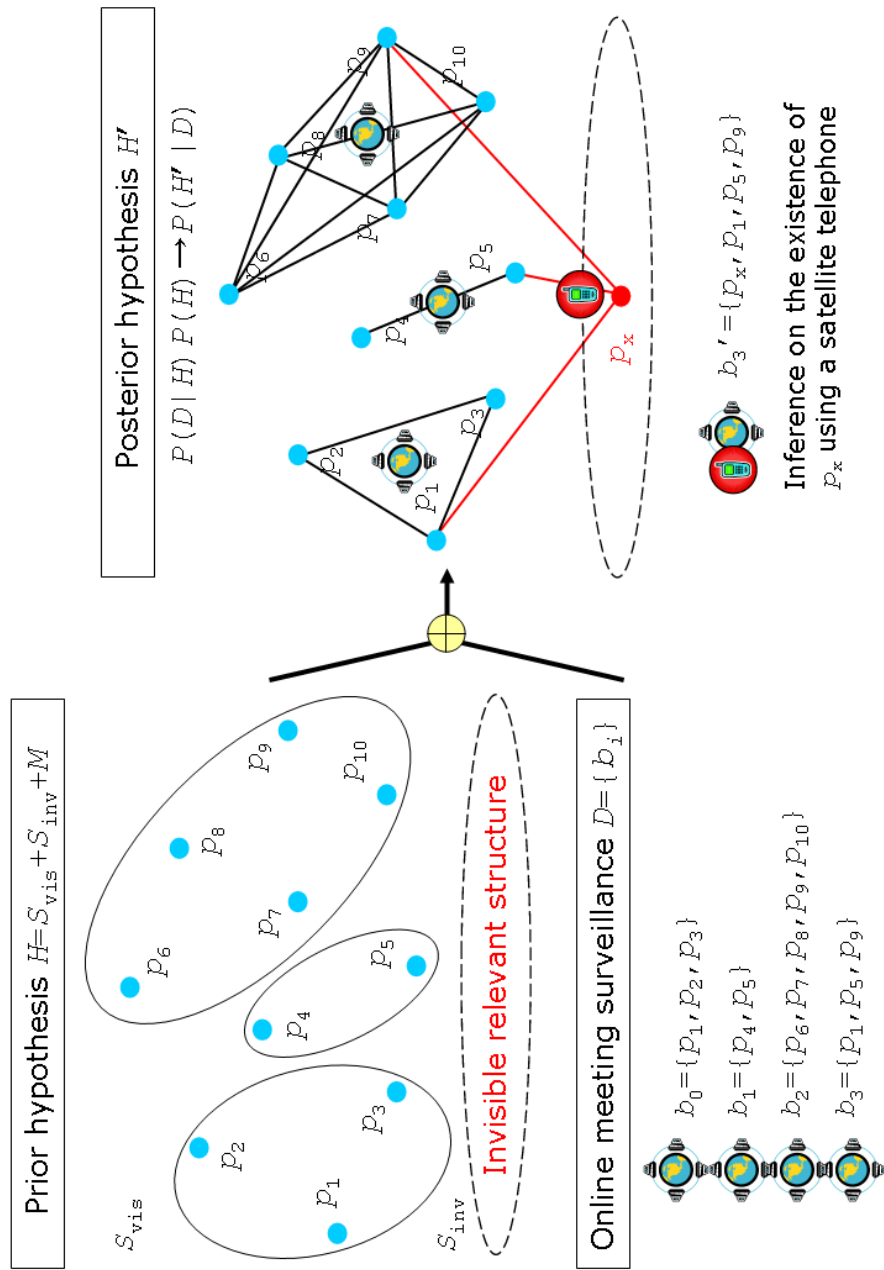


図 1: [左側上 prior hypothesis] 観測された見える構造  $S_{\text{vis}}$  , 観測できない見えない構造  $S_{\text{inv}}$  , コミュニケーションのモデル  $M$  についての事前仮説  $H$  と , [左側下 online meeting surveillance] コミュニケーションの監視から得た観測データ  $D(S_{\text{vis}})$  とから推測する , [右側 posterior hypothesis] 事後仮説  $H'$  .

ないと ( $r_{ij} = 0$ ) , コミュニケーションがない , ことを表現できる . 通常 , コミュニケーションは双方向に発生するため , リンクには向きがない無向グラフであると仮定する . 図 1 左側の prior hypothesis は ,  $p_1$  から  $p_{10}$  までの 10 人の人物があり ,  $p_1, p_1, p_1$  の 3 人のグループ内にはコミュニケーションがあってリンクが存在するだろう , といった事前仮説を示している . 見えない構造は , 未知の潜在的な人物  $p_x$  とその人物を含む関係性  $r_{xj}$  を表す . 関係性は , 未知の人物同士の関係性である場合もあるが , 未知の人物を探る上では , 既知の人物との関係性が重要なヒントになるだろう . コミュニケーションのモデル  $M$  は , グラフ  $G$  からコミュニケーションについての観測  $D$  への写像を表す . 形式的には , 写像 Fnc を用いて , 式 (3) で表現できる .

$$M = \text{Fnc}(S_{\text{vis}} \cup S_{\text{inv}} \rightarrow D(S_{\text{vis}})). \quad (3)$$

次に , 観測手段によって , 見える構造についての観測データ  $D(S_{\text{vis}})$  を収集する . 例えば , 警察の捜査官が , インターネット上のオンラインチャットミーティングを監視し , 参加者の記録を作成する . この場合の観測した記録は , 1 回のチャットミーティングへの参加者  $p_j$  の集合  $b_i$  (式 (4)) である . マーケティングにおいて買い物かごに入れられた商品群にちなんで , この集合をバスケット形式のデータと呼ぶ . 1 回目のチャットミーティングに対するバスケット  $b_0$  は , 人物  $p_1, p_2, p_3$  が参加したことを示している . 集合であるため , バスケット  $b_i$  内に現れる  $p_j$  の順序には , 意味はない .

$$b_i = \{p_j\}. \quad (4)$$

複数のチャットミーティングに渡って , バスケット  $b_i$  を蓄積しリストとしたものが , 観測データ (式 (5)) である . 図 1 左側の on-line meeting surveillance は , 4 回のチャットミーティングに対する参加者を記録した観測データ ( $b_0$  から  $b_3$  までの 4 つ) である .

$$D(S_{\text{vis}}) = \{b_i\}. \quad (5)$$

最後に , 事前仮説と観測データに対して , 式 (6) のベイズの定理を用いれば , 観測データを用いて , 仮説を更新することができる [Hastie 2001] .

$$P(H'|D) \propto P(D|H) \times P(H). \quad (6)$$

$P(x)$  は、変数  $x$  についての確率分布である。更新された仮説（事後仮説  $H'$ ）では、見えない構造  $S_{inv}$  についてのより確からしい知識が得られるかもしれない。例えば、捜査官は、衛星携帯電話を用いていて、インターネット上のオンラインチャットミーティングの監視では捕らえられない人物  $p_x$  を存在を推理できるかもしれない。図 1 右側の posterior hypothesis は、 $p_x$  はバスケット  $b_3$  に関与しており、3本のリンク  $r_{x1}, r_{x5}, r_{x9}$  が存在する可能性を示している。このように、コミュニケーションから見えない構造を読み解くことは、数理的には、ネットワークのノード発見問題と捉えると理解しやすい。

しかし、現実の組織構造、人間の行動、概念の体系といったものが、単純なネットワークのモデルでどの程度記述できるのか明らかではない。人間が持つ環境認識や状況証拠、見えない構造についての漠然とした直感といったものを、パラメータを用いた定量的な事前仮説の形に書き下すことも難しい。また、1,000人規模の組織構造を分析するには、計算量の面において、式(6)の解を求めることは容易ではない。より実用的な観点で、コミュニケーションから見えない構造を読み解くには、ネットワークのノード発見問題という数理的な面での理解をふまえ、数理的には表現しにくい人間の理解や認知を有効に活用したヒューリスティックなアプローチが必要になる。

本研究では、観測データに含まれない、見えない構造を探る問題に取り組むため、以下の3点を課題として問題提起する。見えない構造についての仮説を創出する手法として、新しいアルゴリズムを用いたヒューマンインタラクティブアニメーリングを提案する。ヒューマンインタラクティブアニメーリングでは、人間の事前理解と観測データへの認知とを活用し、それらのずれを可視化して気づきを獲得し、仮説を創出する。組織、意思決定、コミュニケーションなどの観点から、人間活動にかかわる、見えない構造を探るのに適した手順を明らかにしていく。

- 観測にかからない、見えない重要な構造の特徴を考察し、ネットワーク上の空白地帯にノードが隠れている可能性を定量的に示すヒューリスティックなアルゴリズムを開発すること。
- 上記のアルゴリズムをもとに、人間の解釈・事前理解と相互作用しながら、見えない重要な構造に対する仮説を進化させる成長型のプロセスを開発すること。
- 個別の問題に対して、見えない重要な構造を理解し、仮説を探索的

に導出できるか検証し，個別の問題に適した応用プロセスを開発すること．

本研究は，2章で関連研究と課題をまとめた後で，ヒューマンインタラクティブアニメーションの手法の提案，手法の基本的な特徴の分析，組織的な意思決定についての実証実験，いくつかの具体的な問題への応用展開について，詳細を述べる．見えない構造への足がかりとして，ヒューマンインタラクティブアニメーションの手法とアルゴリズムの提案(3章)と，シミュレーションによるアルゴリズムの基本的な特徴の分析(4章)とについて述べる．心理，意思決定，組織的行動を取り巻く，見えない構造を探る問題に取り組み，20名の参加者で行った意思決定の隠れた黒幕を探る実験(5章)と，9/11を実行したテロ組織における共謀者を探るケーススタディ(6章)とについて述べる．心理，意思決定，組織行動応用として，グループ討議で好みについての行動・心理への理解を深める問題に取り組み，コミュニケーションの活性化(7.2節)，潜在的な選好構造の理解(7.3節)についての成果について述べる．

付録Aでは，最尤法によるネットワークのノード発見問題の解法について述べる．付録Bでは，テキスト解析・発想支援への応用展開として，特許の可視化について述べる．

## 2 関連研究

関連研究の特徴と課題について、まれな事象の解明、確率的推論、ネットワーク分析、それらの応用問題に分けて述べる。まれな事象の解明と確率的推論とは、データとして観測できない空白地帯の解明に関連した研究である。ネットワーク分析は、意思決定にかかわるコミュニケーションや人間関係のモデルとして重要である。

### 2.1 まれな事象の解明

まれな事象の解明にかかわる代表的な研究として、チャンス発見の手法 [Ohsawa 2006] がある。チャンスは、まれだが意思決定にとって重要な事象である。観測データの可視化、新たな気づきの獲得、新たな観測データの収集、の繰り返しからシナリオを創出する。チャンスは、人間の意思決定に重要なインパクトを与える事象と定義される。チャンスは、機会である場合もあれば、脅威である場合もある。チャンスを発見することは、チャンス的重要性を意識し理解することを意味する。チャンス頻度がまれな場合には、特に重要となる。機会からの望ましい結果が促進され、脅威からの望ましくない結果が回避されるよう、意思決定し行動をとるべきである。多くの社会的、ビジネス上の領域において、チャンス発見を用いた、潜在的な機会と脅威の理解が必要である。例えば、人間は、マーケティング、技術開発・設計、災害対策、製薬・医療などの分野でチャンスを探している。マーケティング、災害対策、技術開発・設計、製薬・医療などでの応用事例がある。チャンス発見は、そのような領域において、将来を予測するのではなく、将来を創るための手段を提供している。

チャンス発見の中核をなす手法は、人間のプロセスとコンピュータのプロセスとが融合した二重らせんプロセスである。Leeらの研究 [Le05] に見られる、多くのヒューマン・コンピュータ・インタラクティブ・システムが要求の衝突やデッドロックの解消を目指すのに対して、チャンス発見は未来に向けてのシナリオを創出する協調的な手法である。チャンス発見のプロセスは、観測された事象が示唆する兆候に気づき、新しい有意のシナリオを具体化するために設計された手法である。人間のプロセスは、特定の興味を持ち、兆候のチャンスとしての意味を理解し、選択したチャンスからシナリオを具体化し、シミュレーションやシナリオに基づく行動をとり、新しい興味を持つことで成り立っている。他方、コ

コンピュータのプロセスは、人間の興味に基づく環境データを分析し、人間のプロセスに基づく人間の思考を記録した主観データを解析することから成り立っている。

コンピュータは、KeyGraph[Ohsawa 2006]のようなツールを用いて、解析結果をシナリオマップ上のグラフとして可視化する。[Sugiyama 2002], [Fruchterman 1991] は、バネモデルによるグラフ描画法を含む、グラフに作成に有用な手法を示している。グラフの辺は、共起性の強さに応じたばね定数のばねに置き換えられる。ばねによって構成された力学的な系となる。Jaccard 係数が一定以上の事象間には、引力が働くようになり、お互いに近づこうとする。力学的な系が平衡に達するまで、頂点は移動する。平衡後の頂点間の距離に厳密な意味はないが、頂点間の近さは、事象の関連性に関するおおその知見を表現している。

シナリオマップに表示されるグラフの頂点は事象を表す。事象は、入力データから取り出される。入力データ形式は、通常、要素の集合のリストであるバスケットデータである。バスケットは、何らかの性質を共有する事象の集合である。テキストの場合には、センテンスごと、パラグラフごとに形態素解析によって単語の集合を生成し、これをバスケットとする。グラフの辺は、事象間の類似性や共起性を表す。共起性は、Jaccard 係数によって評価できる。Jaccard 係数は、ウェブ構造マイニングの URL 分類や文書の内容の比較などで、よく使用される指標である。[Matsumura 2001], [Murata 2003] に、その活用方法を見ることができる。KeyGraph では、チャンスは、頻度が高く共起性の高い事象群クラスタ間をつなぐ橋のような構造として表現される。チャンスは、クラスタからクラスタへ乗り移る際のトリガのような役割を担っている。潜在的な大きな変化の兆候と考えられる。

チャンス発見を越えようという新しい試みとして、[Ohsawa 2005] のデータ結晶化の方法がある。これは、データに新規の事象を追加すると、その事象が、新規のノードとしてデータを表現するグラフ上のどの位置に出現するのかを可視化する方法といえる。まれな事象の解明にかかわる研究には、他にもある。[Suzuki 2005] は、例外ルールの発見方法である。[Weiss 1998] は、まれな事象に注目した最も初期の研究である。乏しい観測データから未知の事象を発見する手法 [砂山 1999] も研究されている。

チャンス発見は、まれだが意思決定に重要な事象を見つける手法として確立されてきた。しかし、問題も残されている。まれな事象は、必ずしも観測データの含まれるとは限らない。まれであるとは、観測データ

の中で相対的に低頻度であるという意味で，絶対的に希少かどうかは問われない．また，低頻度の基準も明示されていない．データ結晶化の方法では，見えない重要な構造とは何か定義されておらず，問題解決のアルゴリズムというよりも，データ内の事象の位置をグラフ上のノードの位置に変換する描画手段である．KeyGraph に依存するため，描画可能な範囲での探索に留まる．見方を変えると，大規模な観測データのすべての事象に対して，見えない重要な構造が存在する可能性を定量的に評価した上で，見えない重要な構造に関与する事象をグラフ上のノードとして示すことができない問題である．したがって，まれな事象からさらに一歩進んで，観測できない，または，観測が困難な事象を取り扱う必要がでてくる．見えない事象，見えない構造は，データとして観測されない，観測できないが，人間が理解可能な潜在性の高い事象を指す．そのような見えない事象，見えない構造への理解を深める手法の開発が課題と言える．

## 2.2 確率的推論

広い意味で本研究に関連する基礎的な数理的な手法としては，確率的推論，ベイズ統計，多変量解析が重要である．

### 2.2.1 推論の諸手法

まず，見えない事象に近い問題として，潜在変数の学習がある．潜在変数を含むベイジアンネットワークでは，線形潜在変数グラフ (linear latent variable graph) の学習 [Silva 2006] が研究されている．因果関係に関するモデルの妥当性の検証と因果関係の強さの推定・検定とを行う．観測できない時系列的な状態遷移を扱うには，隠れマルコフモデルやカルマンフィルタモデルに代表されるダイナミックベイジアンネットワークの学習 [Friedman 1998] を用いる．因果推定の手法や問題点は，[Glymour 1999] に網羅されている．機械学習や統計的学習については，[Hastie 2001] に詳しい．[Duda 2000] は，パターン識別を網羅的に記載し，クラスタリングにも詳しい．

[Heckerman 1996] は，ベイジアンネットワークにおける因果の独立性 (条件付独立性の集合) によって，確率の検証と推論を容易にする手法を示した．条件付独立性によって上流側のノードの数を削減することと，



複数の条件付独立性によって構造を再構築することで，計算が簡単になり，高速化が図れる．[Silva 2006], [Silva 2003] は，潜在変数を含むベイジアンネットワークで，線形潜在変数グラフ (linear latent variable graph) における潜在変数と因果関係についての情報を導くアルゴリズムである．[Friedman 1998] は，ダイナミックベイジアンネットワークの追加・削除，向きの逆転，ノードの追加などの構造の変更を伴う，学習 (EM によって，観測データの尤度を最大化するトポロジと確率構造の発見) の手法を示した．

具体的な問題の特性に特化した手法の研究も多い．[Allanach 2004] は，隠れマルコフモデルによる状態遷移を推論しながら，テロリストの活動を予測・分析する手法を示した．さらに，[Haiying 2006], [Singh 2004] は，隠れマルコフモデルとベイジアンネットとの組み合わせで，意思決定を支援する ASAM system を開発し，テロリストの活動に対処する意思決定支援に応用した．[Memon 2006] もテロリストネットワークを分析している．

[Adamic 2003] は，ウェブ上の情報 (個人のホームページ間のリンク関係，個人のホームページに記載されたテキスト，個人が登録しているメールアドレス) から友人関係を推定する方法を示した．[Liben-Nowell 2004] は，さまざまな指標によるリンクの存在確率の推定方法を比較した．科学者の協力関係を学習し，予測する問題に適用している．これは，リンク発見問題の代表的な成果である．リンク発見問題は，既知のリンクの存在やノードの属性から未知のリンクの存在を推定する問題である．ランキングや分類と並ぶリンクマイニング [Getoor 2005] の中心的な問題である．リンクの間の依存関係を無向グラフィカルモデルで学習する Markov ランダムネットワーク [Taskar 2003] は，代表的な手法である．Markov ランダムネットワークは，dependence graph [Besag 1974], [Frank 1986], [Anderson 1999] の一種であり，社会ネットワーク分析で広く用いられる．拡張として，階層モデル [Lazega 1999], multiple network [Pattison 1999], valued network [Robins 1999], 高次の依存性 [Pattison 2002], 2-block chain graph [Robins 2001] などが研究されている．

命題論理にも，類似の学習の概念がある．帰納論理プログラミングの構成的帰納学習 [Shafti 2005] は，既存の属性を組み合わせ，普遍性の高い有効な属性を導出する手法である．

確率的推論では，事象の属性として直接観測できない潜在変数に係わる確率的構造の推定や仮説の検証が重要視される．見えない重要事象の

存在そのものについての仮説を創出することが課題である。また、潜在変数に対する確率的推論は、複雑性の高い現実のコミュニケーションや人間行動へ適用することが難しい。複雑性は、双方向でループを形成する影響の伝播、観測変数から潜在変数への影響の伝播、多数の変数などに起因する。これらは考慮されていないため、現実の問題に適用できる実用的な手法の開発が課題である。

ベイズ統計とは、ベイズの定理を中核とした、統計的な推論、検定などの技術の総称である。ベイズの定理とは、ある観測データが得られた時、事前確率から、その観測結果を反映した事後確率を再帰的に求めるための定理である。事後確率は、事前確率と尤度関数の積に比例する。つまり、観測データを得た下で、事前確率を既知の尤度を使って事後確率へと更新する関係を示した定理である。

ベイズの定理は、事前確率と尤度を仮定した下で事後確率を与える、というあくまで相対的なメカニズムを表わした定理にすぎない。したがって事後確率の計算結果の信憑性や有用性は、事前分布と尤度の設定にかかっており、慎重を期すことが必要である。これはベイズの定理を含むベイズ統計学が、不確実性を含む問題を人によって異なる確率を用いて定式化することを許容する主観確率という立場をとっていることによる。この立場はまだ解析対象となっていない新たな問題へのアプローチを可能にするという利点がある一方で、確率の決め方について客観性に欠けるといった批判もある。ベイズ統計を巡る問題については、[Gill 2002] に詳しく議論されている。

ベイズ統計には、事前知識として人間の知見を活用する面があるものの、人間の解釈・理解を更新しながら、繰り返していくプロセスとはなっていない。事前知識以外には、気付きといった、人間のプロセスへの参画が弱い問題がある。また、事前知識と観測データを融合して収束する方向に作用するだけで、事前知識と観測データとのずれから新たな発見へ収束する方法論にはないにくい面がある。刻々と更新される人間の解釈・理解と観測データとのずれから、新たな発見を促進する手法を取り込むことが課題といえる。

### 2.2.2 多変量解析と関連手法

最後に、多変量解析において、共分散構造分析と呼ばれる一群の手法がある [宮川 1997]。共分散構造分析は、従来の因子分析が発展したもので、因子分析と回帰分析とを統合的、検証的に行うものだと考えること

ができる．基本的な定式化としては，母集団の相関係数行列と標本の相関係数行列との逸脱度を最小化する．最小化によって，未知の量を推定する．[稲田 2002] は，潜在変数を含む潜在構造モデルを扱っている．特に，構造方程式モデリング [Bollen 1989] は，因果関係を調べる統計的手法である．直接観測される変数（観測変数）から，直接観測できない潜在変数を導き出し，その潜在変数と観測変数との因果関係について仮説（数理モデル）を設定して，さまざまな現象を理解するアプローチである．

構造方程式モデリングは，検証的な使い方が多く，仮説の導出方法について明確なプロセスを提供するものではない．

一方，グラフィカルモデリングでは，2変数間の関係性を表わす尺度として，偏相関係数（偏相関行列）を利用する．擬似相関と呼ばれる見かけ上の相関関係を排除して，たくさんの変数がある中で，2変数間の純粋な関係の強さ（条件付き独立性）を分析する手法である．グラフィカルモデリングには，潜在変数を扱えないという問題がある．共通の問題として，アンケートへの回答などの離散変数や，ウェブ情報，ビジネス文書，技術文献などのテキストの解析には不向きである．

多様な情報源を活用しながら，人間の事前知識とを融合しながら，潜在的な構造を探索的に検索し，仮説を導出する明確なプロセスを確立していくことが課題といえる．

関連する数理的手法のなかで，クラスタリング手法とデータの低次元表現は，データ分析の基本手法として重要である．クラスタリング手法とデータの低次元表現については，本研究の中で有効に活用していく．

クラスタリングとは，データ解析手法の1つである．教師なしデータ分類手法，つまり与えられたデータを外的基準なしに，データ間の類似性や相違性（距離）に基づいて，自動的に分類する手法である．クラスタリングアルゴリズムを指す場合も多い．教師なし学習の基本は，最尤推定とベイズ推定である．これらを元に，さまざまな手法が提案されている．[Dempster 1977] は，EM アルゴリズムの最初の文献で，未知構造を推定しながら，尤度を最大化するパラメータを決定する手法である．[Bishop 1998] は，EM アルゴリズムでパラメータ推定する潜在変数モデルを積み上げて，階層的な可視化を行う手法を示した．トップレベルの全データから，ボトムレベルの細かいクラスタまでスムーズに可視化するアルゴリズムを提案した．

[Kohonen 1990] は，自己組織化マップの最初の文献である．大きく分けるとデータの分類が階層的になされる階層型手法と，特定のクラスタ数に

分類する非階層的な手法とがある。階層型手法では、系統図と呼ばれる関係木を生成する。最小のスパニングツリーを生成する凝集的(ボトムアップ)手法、分割的手法(トップダウン)手法などのさまざまな方式が提案されている。非階層的な手法の代表的な手法として、K平均法(K-means)、K-medoid法、自己組織化マップなどがある。参考文献 [Hastie 2001]、[Duda 2000] に、詳しく述べられている。

クラスタ数が未知の場合のクラスタ数に関する仮説検定・評価尺度設定、ファジークラスタリング、オンラインクラスタリング、グラフ理論を活用したクラスタリング、次元数の圧縮、主成分分析・独立成分分析など、さまざまな関連研究へと広がっている。

データの低次元表現は、次元数の圧縮、主成分分析・独立成分分析の応用とも考えられる。[Carreira-Perpinan 1998]、[Carreira-Perpinan 1999] は、観測された高次元プロセスを低次元で表現する、潜在変数モデルを用いた多変数回帰、次元圧縮の手法を示した。[Keim 2002]、[Keim 1996]、[Hinneburg 1999]、[Nabney 2005]、[Zoeter 2003] は、高次元データのさまざまな可視化手法を述べている。[Shen 2006] は、情報を集約した可視化によって、アクタの潜在的な関係を理解することを支援する手法である。[Aggarwal 2004] は、高次元データを部分空間でクラスタリングすると、ランキング関数の最適性とは裏腹に、人間に理解しにくいものになるため、人間の直感を取り入れながらクラスタリングを実行する方法を示した。

### 2.3 ネットワーク分析

近年、ネットワーク分析は、2つの流れが融合し、著しい発展を見せている。第一の流れは、物理学に関連した大規模な複雑ネットワークの研究である。90年代後半から、科学者の協力関係のネットワークや映画の出演者、ウェブのリンク構造、化学反応などの大規模な複雑ネットワークには、共通の興味深い性質があることがわかってきた。スケールフリー性 [Barabási 1999a] やスモールワールド性 [Watts 1998] によって、人間の活動に限らず、自然現象をも記述できることから、複雑ネットワークは普遍的なモデルとして広く受け入れられている。第二の流れは、社会学における社会ネットワーク分析の研究である。社会ネットワーク分析は、人間関係や活動の関連性をグラフ構造で表現し、中心性の計算やクリーク構造の抽出を通して、組織全体の構造的な特性を分析する手法である。いずれも、問題領域として、本研究と関連が深い領域である。

### 2.3.1 複雑ネットワーク

[Albert 1999] は、WWW のページごとのリンクの次数  $d$  の分布には、特定のスケールが存在せず、スケールフリー ( $P(d) \propto d^{-\alpha}$ ) になることを観測した。さらに、[Barabási 1999a] は、さまざまな実在のネットワークがスケールフリーになることを観測した。スケールフリーネットワークの理論的基礎が、[Barabási 1999b] に示され、preferential attachment によるネットワークの成長が重要であることがわかった。Center-and-periphery, winner-takes-all といった経済システムに見られる格差や不均質性は、スケールフリー性に起因すると考えられるようになっている。大規模ネットワークのスケールフリー性が、さまざまな社会、経済、自然界のモデルの出発点となっている。[Albert 2000] は相互作用するネットワークの時間発展の周期構造、[Albert 2000] はネットワークの攻撃と破壊に対する耐性を明らかにした。[Yook 2001] は、リンクの重みが増加するモデルである。[Bianconi 2001] は、複雑ネットワークに温度に相当する概念を取り入れ、動的な非平衡性が、winner-takes-all な特性を生み出すことを示した。[Menezes 2004], [Menezes 2004] は、システムのコンポーネントの同時刻の挙動を観測して、内部のダイナミクスと外部からの揺らぎとを分離する。ノイズや作為が多い場合にも機能するアルゴリズムは、重要なテーマであろう。[Fenner 2007] は、ノードが不活性になる時間を設けると、分布のパワーローのテールにエクスポネンシャルのカットオフが発生することを確率モデルで示した。[Adamic 2001] は、スケールフリーネットワーク上での検索が効率的に行えることを示した。検索の容易性や攻撃に対する耐久性は、組織構造を検討する上でも重要な論点である。

[Watts 1998] は、スモールワールドネットワークを論じた最初の報告である。[Newman 2001b] は、科学者のネットワークが、スモールワールド性を持つことを示した。コミュニティなどの社会ネットワークでは、凝集性が高くクラスタが形成されやすいにもかかわらず、ネットワーク全体での平均距離が短くなる傾向にある。スケールフリーネットワークがハブノードによってノード間の距離を短縮するのとは対照的に、スモールワールドではクラスタ間のショートカットリンクが距離を短縮することに大きく寄与している。[Newman 2003a], [Newman 2002] は、ヒューマンネットワークでは、ハブノード同士が隣接しやすい assortative mixing という特性を持つことを示した。[Kauffman 1996] は、ネットワークだけでなく、カオスの観点も論じている。静的なトポロジの分析から、動的な複雑な挙動の発現へと展開している。

複雑ネットワークは多様な展開を見せ、問題領域として極めて興味深いものの、関心の多くが分析的・記述的な観点でなされている。探索的・発見的な手法へ展開し、より能動的な人間の活動への支援へ役立てることが課題であろう。

### 2.3.2 社会ネットワーク分析

社会ネットワーク分析における、具体的、実証的な研究として、オンラインコミュニティのプロファイリング[松村 2003]、ウェブ情報を用いた人間関係の抽出[松尾 2005]、企業関係の抽出[金 2007]、不正会計で破綻した Enron の社内メールの因子分析[Keila 2006]、ベイジアンネットワークと隠れマルコフモデルを用いたテロリストの行動分析[Singh 2004] などの実証研究が活発である。社会ネットワークの分析手法は、[Carrington 2005] に網羅されている。[Freeman 1979] は、中心性 (degree, closeness, betweenness など) のさまざまな尺度を定義した。さらに、[Borgatti 1997] は、2-mode data (bipartite graph) における中心性 (degree, closeness, betweenness など) を定義した。

[Keila 2006] は、2001 年に会計不祥事で破綻したアメリカのエネルギー会社 Enron の社内 email を SVD (singular value decomposition) によって分析し、メッセージの長短、出現する単語のまれ性、単語の使い方と組織特性、犯罪の特徴とに相関があることを示した。[Popp 2006] は、犯罪組織、テロリストに関する研究をまとめたもので、コンピュータを用いたさまざまな分析手法が紹介されている。[Klerks 2002]、[Morselli 2007]、[Krebs 2002] は、犯罪組織、9/11 を中心としたテロリストネットワークの構造や特性を議論した。テロリストネットワークは、クラスタがゆるく結合した構造を持っている。首謀者と共謀者の連携形態、コミュニケーション効率と秘匿性のトレードオフ、目的達成までの時間の拘束などによって、構造が変わるという興味深い特徴を持っている。同様の考察が、[Sageman 2004] においてもなされている。

[Krackhardt 1987] は、観察ではなく、被験者が認知した関係性によるネットワークの分析手法である。アンケートの手法といった、かなり実用面に重点が置かれている。[Johnston 2000] は、企業において、社会ネットワークが、環境に優れたプロセスを採用するのにプラスの効果をもたらす実証研究である。[Batallas 2006] では、中心性とブローカレッジ係数を用いて、航空機エンジンの製品開発チームのネットワークを分析し、Information leaders team の重要性を示した。

社会ネットワーク分析も、複雑ネットワークと同様に、関心の多くが分析的・記述的な観点でなされている。ネットワーク構造の表現であり、観測データに現れない、見えない重要人物を探る問題意識とは異なる。探索的・発見的な手法へ展開し、より能動的な人間の活動への支援へ役立てることが課題であろう。

## 2.4 応用問題

見えない構造への理解が重要となるビジネスへの応用領域として、マーケティングにおける消費者行動・心理分析がある。

マーケティングでは、消費者ごとに、思考や行動における潜在的な傾向を理解することが重要である。例えば、明確に意識していない好みは、新しい製品に興味を持ったり、未知の人と新しいコミュニケーションを始めたりするきっかけになりうる。そのような消費者ごとの傾向を理解するためのさまざまな手法が提案されている。[Gill 2002] は、マーケティング応用の観点からベイズ推論、特に、階層ベイズを論じている。[Arora 1998] は、マーケティングにおける、階層ベイズモデルによる要求の分析手法である。平均的な挙動よりも、個々人の挙動をより重視する点に特徴がある。[Green 1981] は、マーケティングにおける、コンジョイント分析による製品の要素の最適な組み合わせを見つける手法を示した。一種の組み合わせ最適化である。Stated choice method は、その一形態である。

これらの研究は、明示的・意識的な好みを扱っているが、より潜在的、無意識的な好みに焦点を当てた研究もある。無意識的知覚の研究として興味深いのは、[Greenwald 1998] の IAT (implicit association test) である。無意識的に、存在を知覚できなくても、関係性を知覚できる点に着目している。関係性が、潜在的な構造へアプローチする際のヒントになることを示している。潜在クラスモデルは、潜在変数のモデル化と統計的な検証を扱う [Cheung 2004], [Zhang 2004] 。

これらの研究は、個人の傾向についての仮説の検証を重視し、仮説を創出するための方法を提供するものではない。通常、潜在性の高い傾向そのものについての情報は、欠落している。しかし、潜在性の高い傾向と観測可能な明示的な傾向との関係性についての情報が得られる場合がある。このような関係性を理解し、潜在性の高い傾向そのものを解き明かすヒントとするような手法の開発が課題である。

最後に、広い意味で、コミュニケーション、意思決定、マーケティング

に関連する応用技術について述べる。

グループ討議の手法，Webからの知識抽出，技術文書の解析など多様な手法が存在する。[Kwok 2002]は，Webを用いた集団によるよりよい意思決定の支援の手法，集団で合意を形成するための手法を示した。[Rush 1997]は，多数のエキスパートからの知識抽出のための手法である。[Hori 2004]，[Hori 1994]は，基本的な重要要素の解体・再形成によるコンセプト形成の手法と知識マネジメントの手法を提案した。[Mukherjea 2005]は，バイオ分野の特許を活用した知識抽出方法である。これらの手法は，グループ討議に利用でき，ヒューマンインタラクティブアニーリングに採り入れることができる。

テキスト解析では，データマイニング，Webマイニング，パターン認識，クラスタリング，特徴抽出，学習などに関連する重要な研究分野がある。文書内の語の出現頻度，順番，位置などの統計情報から要約，構造，キーワードなどを抽出する研究がなされている。PAIは，代表的な手法であり，チャンス発見にも活用されている[松村 2002]。これらの単独の文書を扱う研究に対して，文書間の類似性や分類を扱う研究の例もある。Word vectorは，文書を分類し検索を容易にするための実用レベルの技術として，概念検索の技術などに応用されている[Schuetze 1998]。Webマイニングの研究では，Webページでの出現頻度から組合せの有効性や独自性を評価する手法[西原 2004]や，Webページの分類法[井山 2004]が研究されている。これらの手法は，ヒューマンインタラクティブアニーリングのアルゴリズムの前処理として，Webやテキストの処理に採り入れることができる。

## 2.5 まとめ

関連研究の問題点をまとめる。

- まれな事象を解明するための手法には，問題が残されている。まれな事象は，観測データの含まれるとは限らない。まれであるとは，観測データの中で相対的に低頻度であるという意味に過ぎない。低頻度の基準も明示されていない。データ結晶化の方法では，見えない重要な構造とは何か定義されておらず，問題解決のアルゴリズムというよりも，データ内の事象の位置をグラフ上のノードの位置に変換する描画手段である。KeyGraphに依存するため，描画可能な範囲での探索に留まる。見方を変えると，大規模な観測データのす



すべての事象に対して、見えない重要な構造が存在する可能性を定量的に評価した上で、見えない重要な構造に關与する事象をグラフ上のノードとして示すことができない問題がある。

- 確率的推論では、事象の属性として直接観測できない潜在変数に係わる確率的構造の推定や仮説の検証が重要視されてきた。見えない重要事象の存在そのものについての仮説を創出することには寄与しない。潜在変数に対する確率的推論は、複雑性の高い現実のコミュニケーションや人間行動へ適用することが難しい。複雑性は、双方向でループを形成する影響の伝播、観測変数から潜在変数への影響の伝播、多数の変数などに起因する。これらは考慮されていないため、現実の問題に適用できる実用性な水準にあるとは言えない。ベイズ統計には、事前知識として人間の知見を活用する面があるものの、人間の解釈・理解を更新しながら、繰り返していくプロセスとはなっていない。事前知識以外には、気付きといった、人間のプロセスへの参画が弱い問題がある。また、事前知識と観測データを融合して収束する方向に作用するだけで、事前知識と観測データとのずれから新たな発見へ収束する方法論にはないにくい面がある。
- 構造方程式モデリングは、検証的な使い方が多く、仮説の導出方法について明確なプロセスを提供するものではない。グラフィカルモデリングには、潜在変数を扱えないという問題がある。共通の問題として、アンケートへの回答などの離散変数や、ウェブ情報、ビジネス文書、技術文献などのテキストの解析には不向きである。多様な情報源を活用しながら、人間の事前知識とを融合しながら、潜在的な構造を探索的に検索し、仮説を導出するには不向きである。
- 複雑ネットワークは多様な展開を見せ、問題領域として極めて興味深いものの、関心の多くが分析的・記述的な観点でなされている。探索的・発見的な手法へ展開し、より能動的な人間の活動への支援へ役立てるには至っていない。社会ネットワーク分析も、複雑ネットワークと同様に、関心の多くが分析的・記述的な観点でなされている。ネットワーク構造の表現であり、観測データに現れない、見えない重要人物を探る問題意識とは異なる。
- マーケティングにおける消費者行動・心理分析の研究は、個人の傾向についての仮説の検証を重視し、仮説を創出するための方法を提

供するものではない。通常，潜在性の高い傾向そのものについての情報は，欠落している。しかし，潜在性の高い傾向と観測可能な明示的な傾向との関係性についての情報が得られる場合がある。このような関係性を理解し，潜在性の高い傾向そのものを解き明かすヒントとするような手法は知られていない。

関連研究が解決していない課題をふまえて，本研究は，意思決定の背後で作用している，直接的にはデータとして観測できない空白地帯に潜む，コミュニケーションや人間関係の見えない重要な構造を探る新しい手法を開発し，他者の意思決定，自己の意思決定に係わる構造についての理解を深めることを目指す。

### 3 ヒューマンインタラクティブアニメーリング

本章では、人間の解釈、理解と相互作用しながら、見えない構造に対する仮説を進化させる成長型のプロセスとして、ヒューマンインタラクティブアニメーリングを提案する。ヒューマンインタラクティブアニメーリングにおいて、観測データを処理し、見えない構造が隠れている可能性が高い観測データを指摘し、意思決定が重要であることを提示するヒューリスティックなアルゴリズムを提案する。

はじめに、アルゴリズムとプロセスの開発の方針を述べる。本章で扱う重要な構造とは、独立性の高いクラスタの間を結びつけるハブのような構造だと定義する。典型的には、多くのクラスタを結びつけるハブノードである。例えば、テロリスト組織を構成する世界各国の無数の実行部隊の背後に潜むリーダーである。6章では、重要な構造の定義をハブノードからより一般的なトポロジ上の特徴を持ったノードに拡張する。このような構造が観測にかからない場合、観測データのどの部分が見えない重要な構造に関連が深いのか、ランキング関数を開発する。ランキング関数は、観測データの部分ごとに、見えない重要な構造を含む確からしさを数値として算出し、観測データの部分ごとの確からしさのランキングを行えるようにする。

プロセスでは、アルゴリズムの出力を無向グラフ構造として描画し、グラフを解釈しながら人間の事前理解と比較できる形態とする。グラフは、事象と事象間の関係性のトポロジを表現したもので、シナリオマップと呼ぶ。シナリオマップには、見えない重要な構造を含む確からしさの高い観測データをもとに、見えない重要な構造が存在する可能性の高いグラフ上の空白地帯を特定し、目印としてダミーノードが配置されている。人間の事前理解とシナリオマップの粒度の違い、事前理解とは一致しない構造に着目しながら、ダミーノードが対応する、見えない重要な事象についての仮説を探索的に創出する。なお、アルゴリズムが扱うのはハブノードであるが、プロセスはハブノードを扱うという前提に立つものではない。他の特徴を持つノードにも適用できるものである。

## 3.1 事前理解と観測のはざまに潜む発見の契機

### 3.1.1 シナリオマップ

ヒューマンインタラクティブアニメーションでは、仮説の創出プロセスに参加する参加者がまとめた概念として理解している独立性の高い事象群の間にある弱い干渉を可視化し、参加者の事前理解に含まれない仮説の創出を促す。干渉とは、事象群の独立性を汚染する、事象群の共起構造に現れる重なりである。観測データから、干渉を計算するのが、アルゴリズムである。計算された干渉の源に対応する仮説的な構造を、見えない重要事象の候補として描画する。これがシナリオマップである。シナリオマップは、コンピュータのデータ処理と人間の理解とのインターフェースとなるだけでなく、プロセスの参加者が相互のコミュニケーションを円滑化するためのツールとも言える。

さて、シナリオとは、映画の筋書きや脚本を意味し、未来に向けての事象の系列、特に時間的な系列を指す。事象とは、人間の発言や行動などの観測可能なできごとを指す。時間的な系列という観点からは、状態の遷移の流れとも言える。時間的な系列を因果関係と捉えることができる場合もあるだろうが、特に、因果関係に限定するものではない。シナリオは、必然的に起こるであろう未来予測の結果でも、意思決定者に推奨する最適な行動計画を自動導出した結果でもない。それ自体は、観測データの上で未来に向けて矛盾なく繋ぎ合わせることが可能な事象の系列に過ぎない。したがって、シナリオが定義できることと、そのシナリオがある特定の環境下で意思決定に有用であることとは異なる。シナリオは、仮説とも言える。シナリオの有用性は、環境と問題に対する人間の理解が深まる中で、はじめて判断できるものである。環境の変化に対する予想と自身の行動計画の意思が混ざり合ったものである。あるシナリオが想定外の環境下で有用であると見出すこと、あるいは、想定外のシナリオがある環境下で有用であると見出すことは、典型的な発見の契機となるだろう。

シナリオマップは、多様なシナリオを読み取る元となる問題領域の見取り図である。事象をノードで、事象間の関係性の有無をリンクで表現する無向グラフである。双方向の関係性を前提としている。方向性のあるベイジアンネットワークの条件付き確率とは異なる。関係性については、事象の発生についての観測可能な共起性を基礎とする。共起性は、時間的な場合と空間的な場合とがある。さらに拡張された共起性として、あ

る観点から見た事象間の類似性を考えることもできる。本研究におけるシナリオマップの特長は、観測可能なできごとをノード、クラスタとして可視化するだけでなく、観測できない見えない重要事象の候補も合わせて図示することにある。

シナリオマップの例を図2に示す。John. F. Kennedy 暗殺事件に関する Wikipedia の記事から作成したヒューマンネットワークである。ノードは人物、リンクは人物の共起性を表現する。事象という意味では、ノードは、その人物が John. F. Kennedy 暗殺事件において何らかの役割を果たしたことを意味する。役割には、「政治的に対立した」、「計画を立てた」、「暗殺を実行した」、「証拠を隠滅した」といった犯人側のものあれば、「捜査を行った」、「証言を残した」といった警察側のものもあるだろう。「親族であった」というのもひとつの役割とみなせるだろう。ここでは、役割を細分化し分析を詳細化することが目的ではないので、さまざまな役割をまとめて、単純に扱う。具体的には、Wikipedia の記事のひとつの文に現れる人物名の集合から、人物の間の共起性を算出し、無向グラフに描画したものである。シナリオマップには、2つの構造が現れる。第一に、観測された事象が持つクラスタ構造である。関係性の強い人物のグループで、図2に黒いノードとリンクで示されている。第二に、クラスタ構造が捉えきれない潜在性のある弱い構造である。クラスタ間の空白地帯に位置し、観測できない見えない重要事象の候補として図2に赤いノードとリンクで示されている。

クラスタ構造が捉えきれない潜在性のある弱い構造は、観測にはかからないダミーノードとして、クラスタ間の空白地帯に描画される。John. F. Kennedy 暗殺事件であれば、暗殺者などの実行犯グループの背後で、政治的な陰謀に関連した政治家、軍の関係者、産業界の実力者、あるいは彼らのプレーンであった弁護士などに対応する。この空白地帯が、見えない重要な構造を示唆し、観測されていない重要な人物の役割に気づくきっかけを与えるものと期待される。クラスタ間を埋めるダミーノード  $DE_i$  は、空白地帯にあるクラスタ間の干渉の源に対応する仮説的な構造で、見えない重要事象の候補 (crystallized dark event) を表す。ダミーノードは、観測データに含まれない、仮説的な事象である [Ohsawa 2005]。これらが潜在構造 (annealed latent structure) である。人間の事前理解では独立に見えていたクラスタ間の潜在構造を含むシナリオマップから受ける印象、新たに想起するノードの連鎖の意味、未知のダミーノードへの関心などが、観測データへの認知である。通常、事前理解と観測デー

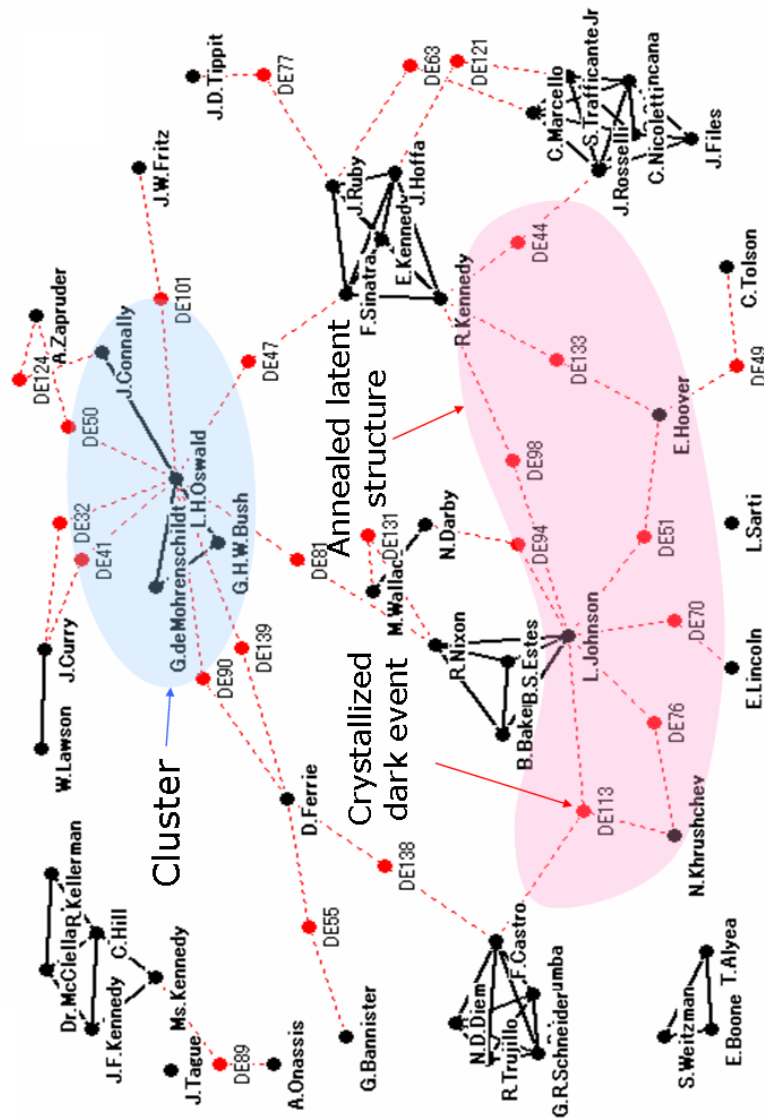


図 2: クラスタと観測できない見えない重要事象の候補としてのダミーノード (crystallized dark events) を無向グラフに可視化したシナリオマップの例。John. F. Kennedy 暗殺事件に関する Wikipedia の記事から作成したヒューマンネットワークである。ノードは、人物が暗殺事件で何らかの役割を果たしたことを意味する。リンクは、その役割の間に関係性があることを示す。複数のクラスタ間を埋める、見えない事象群によって、観測にかからない潜在構造 (annealed latent structure) が構成される。

タへの認知との間には、ずれがあり、これを発見の契機と捉える。

なお、シナリオマップ上の絶対的な位置には意味はなく、リンクのトポロジの関係性にのみ意味がある。絶対的な距離の値にも意味はない。通常、グラフ描画法としてよく用いられるばねモデル [Fruchterman 1991], [Sugiyama 2002] を用いてノードを配置するため、人物の共起性の強さがある程度反映した配置となっている。

### 3.1.2 参加者間の相互作用

ヒューマンインタラクティブアニメーションのプロセスの参加者が、コミュニケーションを円滑化するために、シナリオマップを用いる。参加者は、観測データを収集し、観測データをシナリオマップに可視化してシナリオを創出し、そのシナリオの有用性を評価して意思決定に役立てることになる。シナリオマップに可視化してシナリオを創出する段階とシナリオの有用性を評価する段階は、グループディスカッションを用いることとする。上記の段階を踏むには、参加者として、プロセスの流れを調整し意思決定につなげる調停者、観測データの元となる被験者、コンピュータを操作しながらシナリオマップを作成する作成者、シナリオマップを読み解きながらシナリオを創出する創出者、シナリオの有用性を評価する評価者の役割がありうる。

上記の参加者を図3にまとめる。ヒューマンインタラクティブアニメーションの通常の運用では、被験者、シナリオマップ作成者、シナリオ創出者は同一人物群である。被験者が、自ら、グループディスカッションでシナリオマップを作成しながらシナリオを創出することで、被験者自身への理解を深め、シナリオの創出を促す。自分自身の事前理解を確認しながら、思い込みを生みやすい解釈の枠組みを再構築しつつ、理解を深め仮説の創出を進める作業が織り込まれている。シナリオマップを指差しながら、解釈を表明したり、解釈をさらにメタ認知・集団メタ認知することで、自己と他者への理解を深めることが期待できる。このような意味で、コミュニケーションのツールとしてのシナリオマップの意義は大きいと考えられる。調停者、被験者、シナリオマップ作成者、シナリオ創出者、評価者の5者が、すべて同一の場合もある。必ずしも、客観的な評価者がシナリオを評価することがいいとは限らない。自分自身への理解を深める問題においては、理解の良し悪しを第三者が評価できるものではないからである。

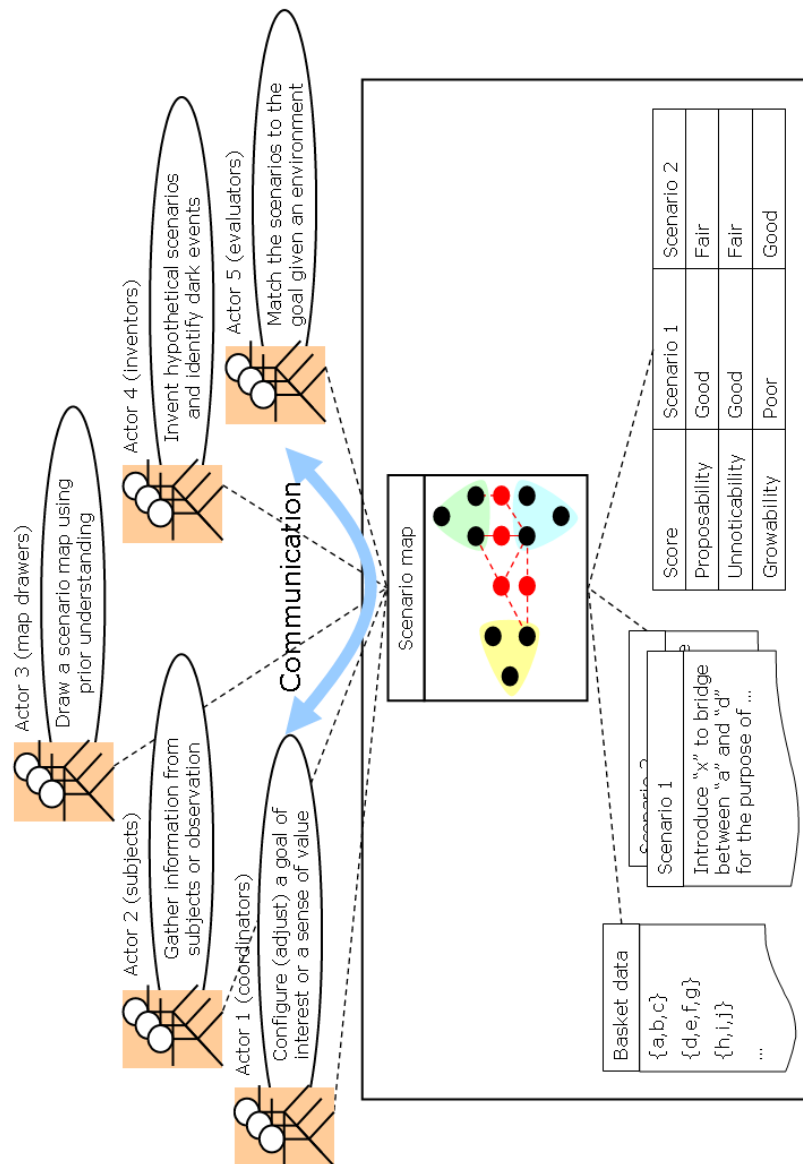


図 3: ヒューマンインタラクティブアニメーションにおけるプロセスへの参加者。調停者, 被験者, シナリオマップ作成者, シナリオ創出者, 評価者の 5 者がある。利用環境に依存して, これら 5 者は, 異なる場合も同じ場合もある。ヒューマンインタラクティブアニメーションの通常の運用では, 被験者, シナリオマップ作成者, シナリオ創出者は同一である。被験者が, シナリオマップを作成しながらシナリオを創出することで, 被験者自身への理解を深め, 仮説創出を促す。



最後に、プロセスの中で、シナリオマップを作成しシナリオを創出する上で重要となる、粒度への適応の概念を説明する。図2のシナリオマップに現れるクラスタの大きさ（クラスタの数、クラスタ内の事象の平均的な数）を粒度と呼ぶ。小さいクラスタ（事象の数が少ないクラスタ）ばかりのシナリオマップは、粒度が細かく、大きいクラスタ（事象の数が多クラスタ）ばかりのシナリオマップは粒度が粗いとする。「粒度をどう設定するか？」は重要な問題である。3.2節で、理解の粒度への適応について述べる。

### 3.2 理解の粒度に応じた成長型の繰り返しプロセス

図3の中で、グループディスカッションを通して、シナリオマップに可視化してシナリオを創出し、その有用性を評価する手順において、参加者が自らシナリオマップを修正し更新することになる。このシナリオマップの成長のための繰り返しプロセスが、ヒューマンインタラクティブアニメーションである。ヒューマンインタラクティブアニメーションでは、シナリオマップの粒度を参加者の理解のレベルに適応される点に特長がある。仮説を創出する参加者が、描画されたクラスタ群を直感的に理解しやすいと感じれば、クラスタ群と参加者の獲得済みの概念体系の粒度とが一致したと言えよう。理解しやすいクラスタ数は、参加者の事前理解の粒度に対する尺度とみなせるだろう。クラスタの大きさや数は、観測されたデータから自動的に決まるといっても、仮説を創出する参加者の事前理解の程度や粒度に応じて適切に設定されるものである。このような適応性がシナリオマップを作成する際に重要である。

参加者は、シナリオマップを描きながら、シナリオを創出する。自己の事前理解（誤解）を確認しながら、シナリオマップと仮説とを同時に作成するプロセスに特色がある。参加者がシナリオマップを成長させる繰り返しプロセスを図4に示す。縦軸は、参加者が自分の理解の程度に合わせて自ら調整するパラメータで、温度  $|T|$  である。図4の縦軸の上方は、高い温度で、クラスタの粒度が小さく、数が多い。図4の縦軸の下方は、低い温度で、クラスタの粒度が大きく、数が少ない。ただし、温度  $T$  は比喩的な意味であり、シミュレーテッドアニメーションにおける状態の遷移確率 ( $P \propto \exp(-\epsilon/T)$ ) を決定する際のような厳密な意味は持たない。3.3節で述べる、観測データからシナリオマップを作成する際の見えない事象を表出化するアルゴリズムにおいては、温度はクラスタ数  $|c|$  に置き

かえられる。クラスタの粒度が細かい場合を温度が高いと呼ぶ。横軸は、プロセスにおける繰り返しの回数である。参加者は、温度を更新しながら、シナリオマップを成長させる。つまり、クラスタの粒度を調節することになる。参加者が問題の構造への理解を深め、仮説の創出に至ると、繰り返しプロセスを終了する。

繰り返しは、図4の縦軸の上方へ向かう加熱過程と、下方へ向かう冷却過程との組合せである。加熱過程では、最高温度  $T_{\max}$  におけるクラスタの粒度を参加者の事前理解の粒度に一致させることが目標である。冷却過程では、見えない事象を表出化するアルゴリズムでクラスタ群を埋め合わせるダミーノードが参加者の事前理解とのずれに相当し、このずれの原因を想起することが目標である。加熱過程と冷却過程の内容を以下にまとめる。

- 加熱過程： 温度を上げると、クラスタが断片化し、クラスタ数が増える。断片化したクラスタは、細かい粒度の概念に対応する。参加者が1回前の加熱・冷却過程での理解の度合いを認知しながら、シナリオマップの概念の粒度を理解の度合いに近づくよう、最高温度を更新する。プロセスを開始する前の理解の度合いは、参加者の事前理解を指す。参加者が仮説の創出らず、理解がまだ浅いと認知すれば、最高温度をあまり上げずに、粗い粒度の概念で全体を理解する。参加者が、仮説の創出には至らなくても、理解が深まっていると認知すれば、最高温度を上げ、細かい粒度の概念で詳細を理解する。
- 冷却過程： 見えない事象を表出化するアルゴリズムによって、加熱過程の最高温度での断片化したクラスタ群をダミーノードで埋め合わせ、ひとつの全体構造を生成する。ダミーノードは、加熱過程での参加者の理解とのずれに相当する概念を表す。獲得済みの理解に、ゆらぎを与える疑問のような役割を果たす。ダミーノードが表す、見えない重要事象の鍵となる、概念間の欠けた要素を想起しながら、全体構造を解釈し、仮説を探る。理解が深まると同時に、先入観を解くきっかけが得られるかもしれない。理解が深まると、多数の細かい粒度の概念にまたがる、複雑で想定外の仮説を創出できるかもしれない。仮説の創出に至ると、プロセスを終了する。

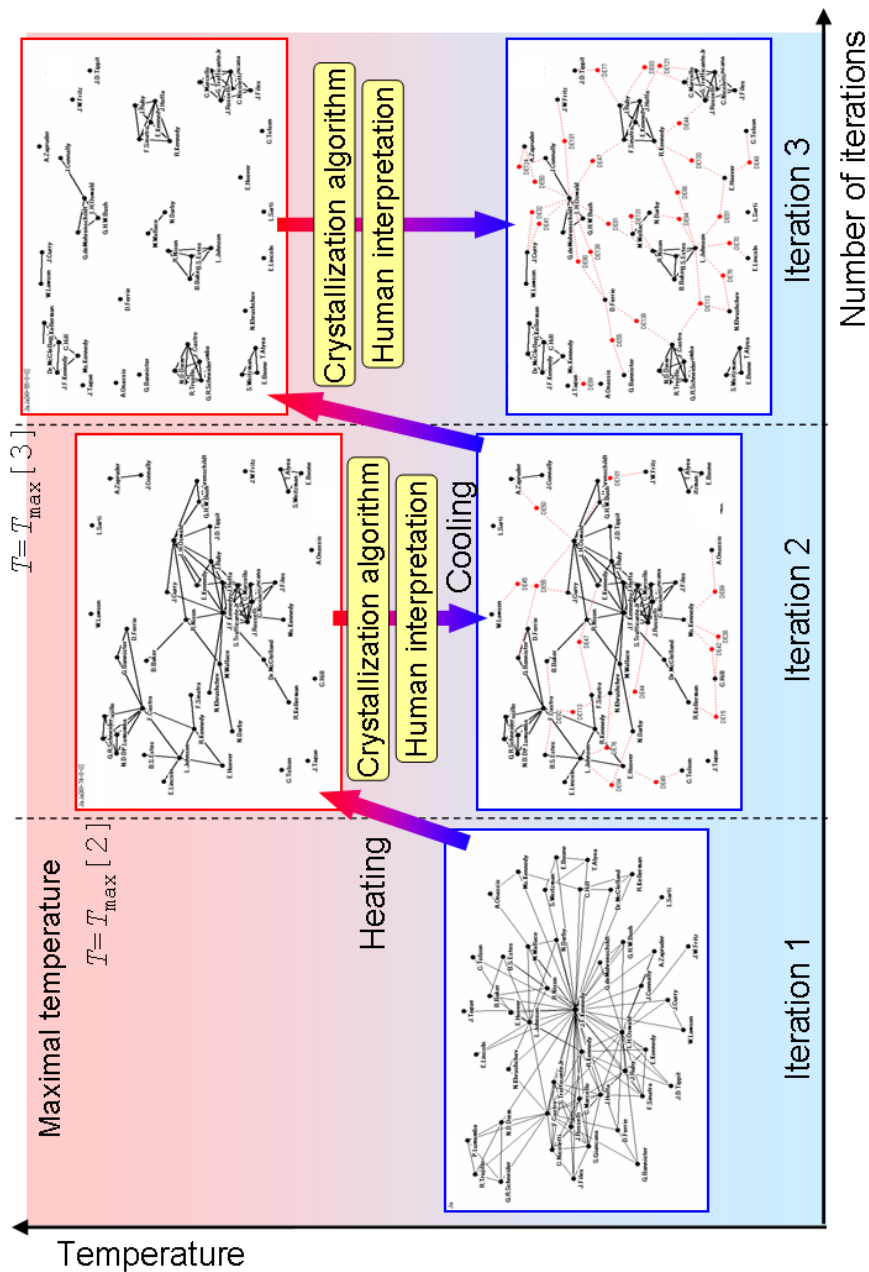


図 4: 参加者がシナリオマップを成長させる, ヒューマンインタラクティブアニーリングの繰り返しプロセス. 縦軸は, 温度に相当し, クラスタ数に対応する. 上方は, 高い温度で, クラスタの粒度が小さく, 数が多い. 下方は, 低い温度で, クラスタの粒度が大きく, 数が少ない. 横軸は, 繰り返し回数である.

### 3.3 見えない事象を表出化するアルゴリズム

図4の冷却過程で，シナリオマップを作成する際に用いるアルゴリズムについて述べる．このアルゴリズムは，大量の観測データを処理し，観測できない見えない事象に関する定量的な情報を出力できる．シナリオマップは，出力を描画したものに過ぎない．描画を重視したアルゴリズムではない．この点は，KeyGraphをベースに可視化を重視した，2.1節で述べた結晶化アルゴリズムとは異なるアプローチを採っている．

観測データのどの部分が，観測できない見えない事象に関連が深いのか？この関連の程度を算出するランキング関数と，ランキング関数から見えない事象へのヒントを取り出すアルゴリズムを考案した．1章で述べたノード発見問題を解くのではなく，ヒューリスティックなアプローチを用いる．ここでは，その詳細を説明する．ランキング関数の設計が，アルゴリズムの中核的な部分である．アルゴリズムの入力は，観測データである．観測データの形式は，1章の式(4)の事象の集合のリストとする．観測データに含まれる事象を観測された事象と呼ぶ．その他に，観測されない事象が存在すると想定している．観測されないのは，観測方法の限界や観測対象として認知されていないことに起因する．観測アルゴリズムの出力は，事象の集合に対して，ランキング関数から算出された観測されない事象への関連の深さで順位を付けた，確からしさのランキングである．このランキングの過程で，観測されない事象が強く関連している事象の集合，強く関連している事象のクラスタ，強く関連している観測された事象が特定される．シナリオマップは，アルゴリズムが出力したこれらの関連している事象とともに，観測されない事象をダミー事象として無向グラフ構造に可視化したものである．観測されない事象は，クラスタ群の隙間の空白地帯に補間された構造とも言える．補間された構造を想起しながら，ダミー事象とクラスタ群のつながりである全体構造を捉えることが，シナリオが作成されることにつながる．

#### 3.3.1 クラスタリング

観測データについて述べる．1回の観測結果は，式(7)のバスケット $b_j$ で表される．これは，時間，空間，ある特定のトピックスなどに基づいて分類された事象 $e_i$ の集合である．同時刻に記録された人物の行動の集合，同じ場所で目撃された人物の集合などが該当する．コミュニケーションであれば，メールサーバから収集した，電子メールの受信者の集合や会

話のキーワードの集合も同じ形式で表現できる．ひとつのバスケットに含まれる事象の数は，任意の値を取りうる．電子メールの受信者は，ひとりの技術者の場合もあれば，技術部門の全員の場合もある．

$$b_j = \{e_i\} \quad (0 \leq j \leq |b| - 1). \quad (7)$$

また，テキストデータを入力として処理することもできる．文や段落を形態素解析し，単語に分けたものをひとつのバスケットとする！「コミュニケーションから読み解く意思決定の見えない構造」であれば， $b_0 = \{\text{コミュニケーション, から, 読む, 解く, 意思決定, の, 見る, ない, 構造}\}$ となる．通常は，助詞や助動詞は削除し，主に，名詞，動詞，形容詞だけを扱い， $b_0 = \{\text{コミュニケーション, 読む, 解く, 意思決定, 見る, ない, 構造}\}$ とする．特許や論文などのテキストデータを用いたアイデア創出といった研究も進めているが，本論文では取り扱わない．

バスケットの数を  $|b|$  で表す．観測の規模や観測を行う期間によって，1以上の任意の値を取りうる．連続的に，オンラインで観測をし続ける場合もあるだろう．事象の種類を  $|e|$  で表す．ひとつの事象は，複数のバスケットの出現することがある．事象の出現回数の合計は， $|e|$  よりも大きい．まず，事象をクラスタリングする．そのために，事象間の距離（近さ）を定義する．事象の発生頻度，共起性をもとに，事象間の距離を計算する．事象  $e_i$  の発生頻度  $F(e_i)$  は，ブーリアン関数  $B(s)$  を用いて，式 (4) の入力データから式 (8) で計算することができる．ブーリアン関数  $B(s)$  は，式 (9) で定義される．論理式  $s$  が真なら 1 を，偽なら 0 を返す．

$$F(e_i) \equiv \sum_{j=0}^{|b|-1} B(e_i \in b_j). \quad (8)$$

$$B(s) = \begin{cases} 1 & \text{if } s \text{ is TRUE} \\ 0 & \text{otherwise} \end{cases}. \quad (9)$$

式 (8) は，事象  $e_i$  が現れるバスケットの個数である．事象  $e_i$  がひとつのバスケットに複数回現れても，発生頻度には 1 が加算される．電子メールのサブジェクトをひとつの観測単位とする場合に，同じサブジェクトで何回も返信し続けると，その返信者は何回も出現する．事象  $e_i$  と事象  $e_j$  との共起性の尺度について，式 (10) で定義される Jaccard 係数を用いる．Jaccard 係数は，リンク発見問題，ウェブマイニング，テキスト処理などで広く用いられている [Liben-Nowell 2004] ．

$$J(e_i, e_j) \equiv \frac{F(e_i \cap e_j)}{F(e_i \cup e_j)}. \quad (10)$$

式 (10) は，式 (11) を用いて，式 (8) に変換できる．

$$J(e_i, e_j) = \frac{\sum_{k=0}^{|b|-1} B((e_i \in b_k) \wedge (e_j \in b_k))}{\sum_{k=0}^{|b|-1} B((e_i \in b_k) \vee (e_j \in b_k))}. \quad (11)$$

Jaccard 係数に代わり，単純な共起頻度（式 (12)）や dependence 係数（式 (13)）を共起性の尺度として用いることもできる [Ohsawa 2006]．

$$C(e_i, e_j) \equiv F(e_i \cap e_j). \quad (12)$$

$$D(e_i, e_j) \equiv \frac{F(e_i \cap e_j)}{F(e_i) \times F(e_j)}. \quad (13)$$

クラスタリングには，k-medoids 法 [Hastie 2001] を使用する．Medoid  $e_{\text{med}(j)}$  はクラスタ  $c_j$  の最も中心に位置する事象を表す．数値データに対する k-means 法におけるクラスタの重心に相当する．クラスタ内の事象群と medoid との間の Jaccard 係数の合計値を最大化するよう，medoid と事象群の分類とを選ぶ．これは EM 法 (expectation-maximization) の一種であり [Dempster 1977]，計算量が少なく，1,000 以上の事象でも実時間で計算できる．まず，クラスタ数と同じ  $|c|$  個数の medoid をランダムに選ぶ．他の  $|e| - |c|$  個の事象は，最も近い medoid を含むクラスタに分類する．次に，クラスタごとに分類された事象群から新しい medoid を生成し，更新する．式 (14) で定義されるクラスタ内の Jaccard 係数の合計値が最大となる事象を新しい medoid とする．

$$M(c_j) \equiv \sum_{(e_i \in c_j) \wedge (e_i \neq e_{\text{med}(j)})} J(e_{\text{med}(j)}, e_i). \quad (14)$$

つまり，式 (15) を実行する．

$$\text{Select } e_{\text{med}(j)} \text{ to maximize } M(c_j). \quad (15)$$

$|c|$  個の新しい medoid をもとに，他の  $|e| - |c|$  個の事象は，最も近い medoid を含むクラスタに分類する．この手順を medoid が収束して変化しなくなるまで，繰り返す．あるいは，階層型クラスタリング [Duda 2000] や自己組織化マッピング (self-organizing mapping) [Kohonen 1990] など，他のクラスタリングアルゴリズムを使用してもよい．

### 3.3.2 ランキング関数

観測されない，見えない事象が関連していた確からしさについてのランキング関数を定義する．式 (16) の  $I_{av}(b_i)$  を用いる．この値が大きなバスケットは，観測されない事象が関連していた可能性が高いとして，取り出す候補とする．どのクラスタとも距離が遠くない（近さが小さくない）バスケットに，クラスタ群を引き付ける観測されない事象が潜んでいるとの考えに基づいている．

$$I_{av}(b_i) \equiv \frac{1}{|c|} \sum_{j=0}^{|c|-1} \max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)}. \quad (16)$$

$e_k \in b_i$  であれば，式 (16) の分子は 1 なので，式 (16) は，単純な式 (17)) に変換できる．

$$I_{av}(b_i) = \frac{1}{|c|} \sum_{j=0}^{|c|-1} \min_{(e_k \in c_j) \wedge (e_k \in b_i)} F(e_k). \quad (17)$$

同時に，ランキング関数の値が大きなバスケットにおいて，クラスタごとに関連の強い（距離が小さい，近さが大きい）事象を選ぶ．バスケット  $b_i$ ，クラスタ  $c_j$  に対して，式 (18) を用いる．このような事象をゲートウェイ事象と呼ぶ．ゲートウェイ事象は観測された事象であるため，これらを注意深く観察することで，観測されない事象へのヒントが得られるだろう．

$$e_{gtw(i,j)} = \arg_k \max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)}. \quad (18)$$

他に，式 (19) で定義される  $I_{nu}(i)$  もランキング関数になりうる． $I_{nu}(i)$  は，ランキングするバスケットと相関のあるクラスタの個数を表す．つまり，バスケットが多くのクラスタにまたがる場合に，大きな値をとるランキング関数である． $I_{nu}(i)$  の大きな値を高く評価する．

$$I_{nu}(i) \equiv \sum_{j=0}^{|c|-1} u\left(\max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)}\right). \quad (19)$$

$$u(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (20)$$

式 (21) で定義される全クラスタに対する標準偏差  $I_{sd}(i)$  もランキング関数になりうる。ランキングするバスケットが、どのクラスタとも相関があると、小さな値を示す。  $I_{sd}(i)$  の小さな値を高く評価する。

$$I_{sd}(i)^2 \equiv \frac{1}{|c|} \sum_{j=0}^{|c|-1} \left( \max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)} - I_{av}(i) \right)^2. \quad (21)$$

式 (22) のクラスタとの相関の上位 2 つ (最も強く結びついたクラスタ対) の合計  $I_{tp}(i)$  もランキング関数になりうる。これは、特定の 2 つのクラスタを結びつけるバスケットに注目することになる。なお、式 (22) において、  $T(x_k, n)$  は、  $x_k$  の値を降順にソートしたときの  $n$  番目の値である。  $T(x_k, 1) + T(x_k, 2)$  は、  $x_k$  のなかの最大値と 2 番目に大きな値との合計値を表す。  $I_{tp}(i)$  の大きな値を高く評価する。

$$I_{tp}(i) \equiv \frac{1}{2} \left( T\left(\max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)}, 1\right) + T\left(\max_{e_k \in c_j} \frac{B(e_k \in b_i)}{\sum_l B(e_k \in b_l)}, 2\right) \right). \quad (22)$$

ランキング関数の値が大きなバスケットから順番に、バスケットをランキングし、アルゴリズムの出力として取り出す。上位何位まで取り出すか、指定する。どのランキング関数が、クラスタを結びつけるハブノードを探るのに適しているのか、7.4.3 節で評価を行う。4 章では、コミュニケーションや組織構造のさまざまな側面から詳細な評価を行う。

これまでの式 (16) や式 (19) は、1 次のランキング関数である。式 (23) のように、2 次以上のランキング関数を定義することもできる。通常、観測されない事象は、複数のバスケットと関連しているため、1 次のランキング関数で逐次的に取り出すよりも、高次のランキング関数を用いる方が精度が高いと予想される。ただし、計算量の増大が問題となるだろう。

$$I_{av}(b_i, b_j) \equiv \frac{1}{|c|} \sum_{k=0}^{|c|-1} \max_{e_l \in c_k} \frac{B(e_l \in b_i) + B(e_l \in b_j)}{\sum_m B(e_l \in b_m)}. \quad (23)$$

バスケット  $b_i$  と  $b_j$ 、クラスタ  $c_k$  に対して、ゲートウェイ事象には、式 (24) を用いる。

$$e_{gtw(i,j,k)} = \arg_l \max_{e_l \in c_k} \frac{B(e_l \in b_i) + B(e_l \in b_j)}{\sum_m B(e_l \in b_m)}. \quad (24)$$

可視化の際には、取り出したバスケット  $b_i$  に対応したダミーノード  $DE_i$  とゲートウェイノードをリンクで接続して描画することで、無向グラフ



とする．2次のランキング関数の場合には，取り出したバスケット  $b_i, b_j$  に対応したダミーノード  $DE_{ij}$  とゲートウェイノードをリンクで接続する．描画するダミーノードの数は，任意である．クラスタ数と同じオーダーであれば，視認性の点で良いと考えられる．ダミーノードをクラスタ間の空白地帯にノードとして配置し，クラスタ間の空白地帯と関連するクラスタとの関係性が可視化される．3.1で述べた通り，シナリオマップ上の事象の位置や事象間の距離の値には意味はなく，リンク  $r_{ij}$  のトポロジ上の関係性にのみ意味がある．しかし，通常は，グラフ描画法としてよく用いられるばねモデル [Fruchterman 1991], [Sugiyama 2002] を用いてノードを配置するため，近くにある事象群，遠くに離れた事象群は，共起性の強さをある程度反映した配置となっている．

### 3.4 アルゴリズムの精度と安定性

ここでは，3.3節で述べたアルゴリズムの基本的な特性を分析する．

スケールフリーネットワークから生成したテストデータを用いて，アルゴリズムの基本的な特性を評価する．評価は，プレジジョンを尺度に，アルゴリズムが観測データから見えない構造について，どの程度の情報を復元できるか調べる．プレジジョンは，情報検索などで，取り出した情報のうちの正解の比率である．スケールフリーネットワークの生成方法として，ネットワークの growth とノードの preferential attachment とを特徴とする Barabasi-Albert モデル [Barabási 1999a] を使用する．スケールフリーネットワークは，人間のコミュニケーションや社会的な関係性・依存性を記述する際に，用いることができる．スケールフリーネットワークは，いくつかの中心的なハブノードも持つ傾向がある．一般に，ハブは，大きな次数（隣接ノードへのリンクの本数）を持つノードである．ハブノードは，ネットワーク全体の機能に大きく影響する．リンクをたどりながら，ノード間のコミュニケーションやノードの検索を行う際に，ハブノードを経由すると効率が高くなる．一方，個々のノードをランダムに削除しても，ネットワークの接続性や効率にはほとんど影響がない．ノードの次数  $d$  の分布は， $P(d) \propto d^{-2.5}$  といったべき乗となる．

この評価では，ハブノードを見えない構造に設定する．上記に述べたように，ハブノードは，多くのノードを結びつける重要な役割を果たすためである．6章では，ハブノード以外の特徴を持つノードについても扱う．図5と図6に，400ノードのスケールフリーネットワークを示す．こ

のネットワークは、ひとつのプライマリハブと 20 個のセカンダリハブとを含む。プライマリハブは、ネットワークの中で次数が最大のノードである。プライマリハブは 0-00、セカンダリハブは 1-00, 2-00, ..., 20-00 とラベルが付けられている。セカンダリハブは、ネットワーク上のグループの中のハブである。つまり、ネットワーク全体は、20 個のグループが相互に接続された構造をとっている。個々のグループ内のノードは、1-xx, 2-xx, ..., 20-xx といったラベルが付けられている。グループは 20 個前後のノードを含む。リンクが異なるグループを接続する確率  $p_g$  は、調整可能なパラメータとした。図 5 では、確率は  $p_g = 0.057$  である。この図では、グループの境界は不明瞭であるが、目視によって 20 個のグループを識別することができる。一方、図 5 では、確率は 0.19 で、目視でグループを識別することは難しい。全体として、ひとつの大きな構造になっているように見えるネットワークの例である。

テストデータは、ネットワーク上を伝播するコミュニケーションのモデル化を想定して、以下の 2 つの手順で生成する。

1. 手順 1: あるノードを起点として、周辺ノードへ伝播するコミュニケーションに含まれるノード群をひとつパターン(コミュニケーションバスケット)とするデータを生成する。ある起点ノードから一定の距離以下のノード群をコミュニケーションへの参加ノード群としてバスケット化する。起点ノードに相当する人物が、ある話題を提供し、その周辺ノードが会話に参加する状況がひとつのバスケットに対応する。この話題の伝播範囲は、リンクのホップ数で数えた距離で表現できる。評価では、2 ホップとした。起点ノードは、400 ノードのなかの任意のノードであり、バスケットごとにランダムに選ばれる。評価では、400 バスケットを生成した。例えば、プライマリハブ (0-00) を起点ノードとするバスケットは、 $\beta_0 = \{0-00, 1-00, 1-01, \dots, 2-00, 2-01, \dots, 20-00, \dots\}$  となる。プライマリハブ (0-00) の発言に、周辺のノードが応え、それらの会話への参加者をひとつのコミュニケーション履歴としてまとめたものに相当する。
2. 手順 2: 手順 1 で生成したコミュニケーションのパターンのリスト  $\beta_i$  から特定のノードを消去することで、潜在構造(見えないノード)を織り込んだテストデータとする。バスケットから、プライマリハブ、セカンダリハブ、または、グループ内のその他のノードを消去する。消去されたノードは、観測にかからない見えない構造に対応

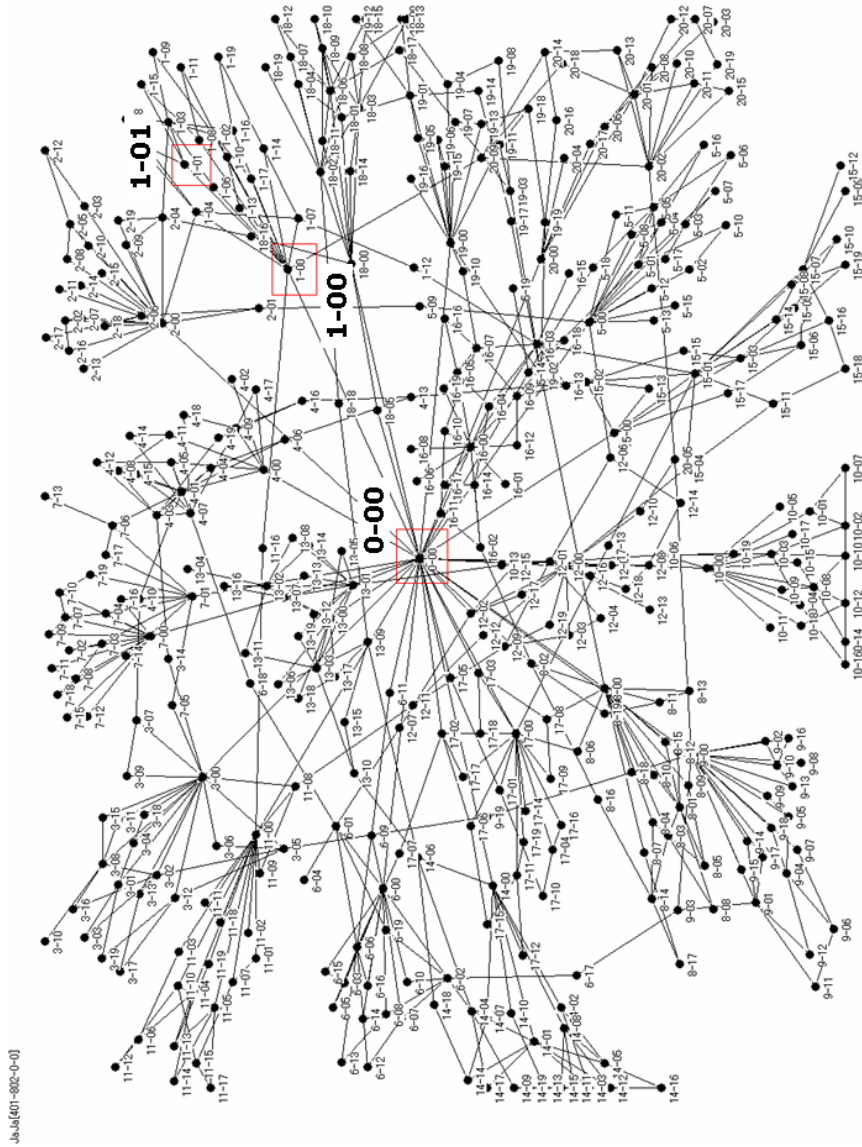


図 5: 400 ノードのスケールフリーネットワーク (グループ間接続確率  $p_g = 0.057$ ) . プライマリハブノード 0-00 , セカンダリハブノードの例 1-00 , 一般ノードの例 1-01 を図に四角の枠で示した .

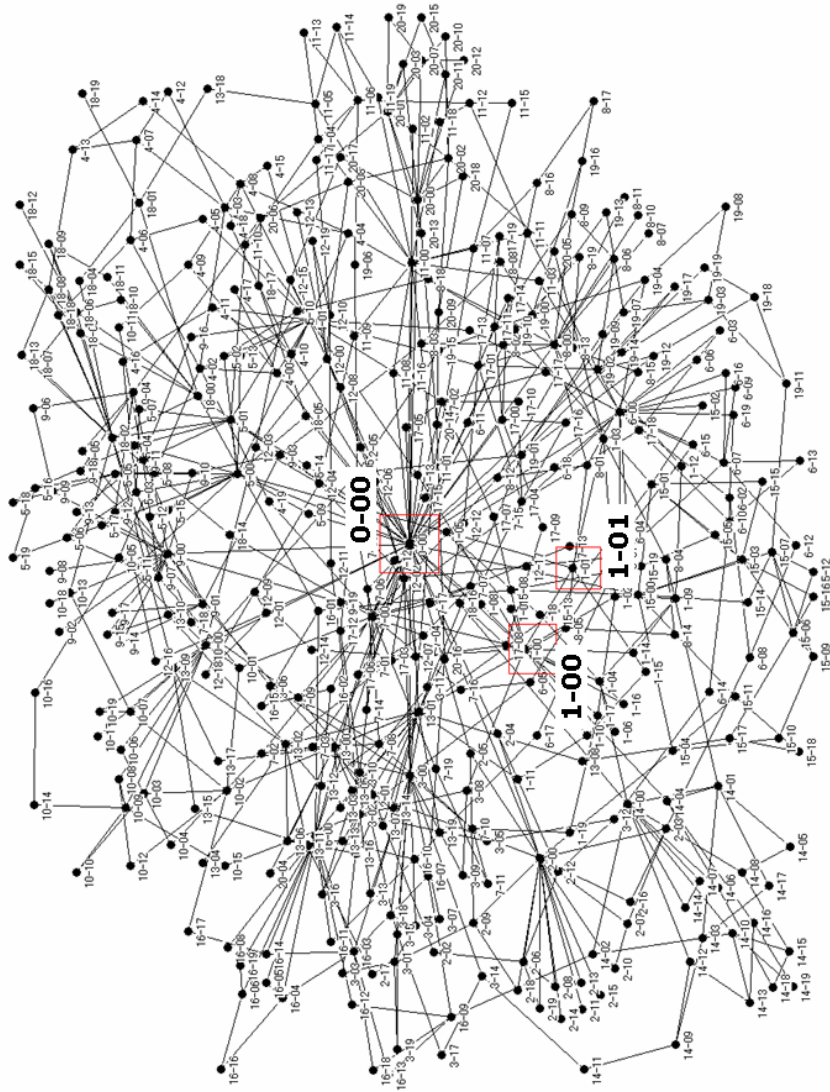


図 6: 400 ノードのスケールフリーネットワーク (グループ間接続確率  $p_g = 0.19$ ) . プライマリハブノード 0-00, セカンダリハブノードの例 1-00, その他のノードの例 1-01 を図に四角の枠で示した .

する．例えば，プライマリハブ (0-00) を見えない構造とする評価では，手順 1 のバスケットの例は， $b_0 = \{1-00, 1-01, \dots, 2-00, 2-01, \dots, 20-00, \dots\}$  となる．見えない構造が織り込まれたバスケットでは， $\beta_i \neq b_i$  となる．ノードが複数のバスケットに出現していれば，複数のバスケットに見えない構造が織り込まれる．ここで，ネットワークにおいてではなく，データにおいて，ノードを消去したことに注意する．消去によらず，ネットワークは同一で，手順 1 のコミュニケーションのパターン  $\beta_i$  は変化しない．同じネットワークで，特定のノードに関する情報のみが観測にかからなくなることに対応する．見えない重要人物が電話で指示を与える場合に，他の人物のメールの監視から得られるコミュニケーションの履歴が該当する．上記のバスケットのリスト  $b_i$  がアルゴリズムの入力となる．アルゴリズムは，すべてのバスケットをランキングし，見えない構造に関係している， $\beta_i \neq b_i$  となるバスケットを取り出すよう動作する．複数のバスケットからノードが消去されていれば，アルゴリズムは，これらすべてを取り出すことを目標とする．

このように生成されたテストデータ  $b_i$  をアルゴリズムの入力とする．アルゴリズムの出力である取り出したバスケット群（ランキング関数のランキングの上位から順番に取り出されたバスケットの系列）から，潜在構造（見えないノード）として設定したノードが消去されていたかどうかで，プレシジョン  $p$  を計測する．プレシジョンは，式 (25) から計算できる．取り出したバスケットの数に占めるノードが消去されたバスケットの数の割合である．

$$p = \frac{\sum_{i=0}^{m^{\text{ret}}-1} B(b_i^{\text{ret}} \neq \beta_i)}{m^{\text{ret}}}. \quad (25)$$

取り出したバスケットの数を  $m^{\text{ret}}$  で表す．この値は，1 から  $|b|$  までである．取り出されたバスケットを  $b_i^{\text{ret}}$  ( $0 \leq i \leq m^{\text{ret}} - 1$ ) で表す．アルゴリズムが適切に動作すれば，取り出すバスケットの数  $m^{\text{ret}}$  が小さいうちは，プレシジョン  $p = 1.0$  となることが期待される．なお，ここでの評価では，ランキング関数には式 (16) を用いた．

まず，事前知識としてのクラスタ数  $|c|$  に対する依存性を評価する．図 7 は，クラスタ数  $|c| = 20$  の場合で，正しい事前知識を持っていると仮定した場合に対応する．横軸は取り出したバスケットの数  $b^{\text{ret}}$  である．ネットワークは，図 5 を用いた．

図7の評価では、3回の実験を試行した。1回目では、バスケットからプライマリハブ(0-00)を消去してその位置を特定できるか試みた。2回目ではひとつのセカンダリハブ(1-00)を消去し、2回目ではひとつのハブではないノード(1-01)を消去した。プライマリハブ、セカンダリハブでは、バスケットの数が20程度まで、高いプレジジョンが得られる。アルゴリズムが良好に動作していることがわかる。セカンダリハブが手順1のバスケットに現れていた回数は、50回程度であるため、バスケットの数が増えるとプレジジョンは劣化する。一方、グループ内のその他のノードについては、プレジジョンの劣化が速い。アルゴリズムは、見えないハブノードが関連しているバスケットを取り出す目標に対して、有効に動作することがわかる。

図8は、クラスタ数=11の場合で、事前知識がずれていると仮定した場合に対応する。図8は、クラスタ数=29の場合で、やはり、事前知識がずれていると仮定した場合に対応する。いずれも、ネットワークは、図5を用いた。その他のノードについてはプレジジョンが劣化するものの、プライマリハブ、セカンダリハブについては、劣化は小さい。事前知識があまり正確でなくても、アルゴリズムは一定の正しい出力を提供できることがわかる。事前知識があいまいであったり、変わりやすかったりしても、おおよその正しい結果が得られ、データの追加的な収集や詳細な調査につながる知見が得られるだろう。

次に、ネットワーク構造に対する依存性を評価する。図10は、クラスタ数=20の場合で、グループ間接続確率を4段階に変更してネットワーク構造を変えた場合である。グループ間接続確率 $p_g$ は、0.019, 0.057 (図5), 0.095, and 0.19 (図6)である。消去したノードは、セカンダリハブ(1-00)である。プレジジョン $p$ は、あまり変化しない。グループ間接続確率 $p_g = 0.19$ で、クラスタを識別しにくい場合でも、アルゴリズムは動作していることが分かる。

アルゴリズムの特性として、事前知識としてのクラスタ数、ネットワーク構造への依存性が少ないことがわかる。この結果から、アルゴリズムは、安定でロバストな良い特性を持っていると考えられる。

### 3.5 まとめ

観測データに含まれない、見えない構造を探る問題に取り組むための新しいアプローチとして、ヒューマンインタラクティブアニーリングを

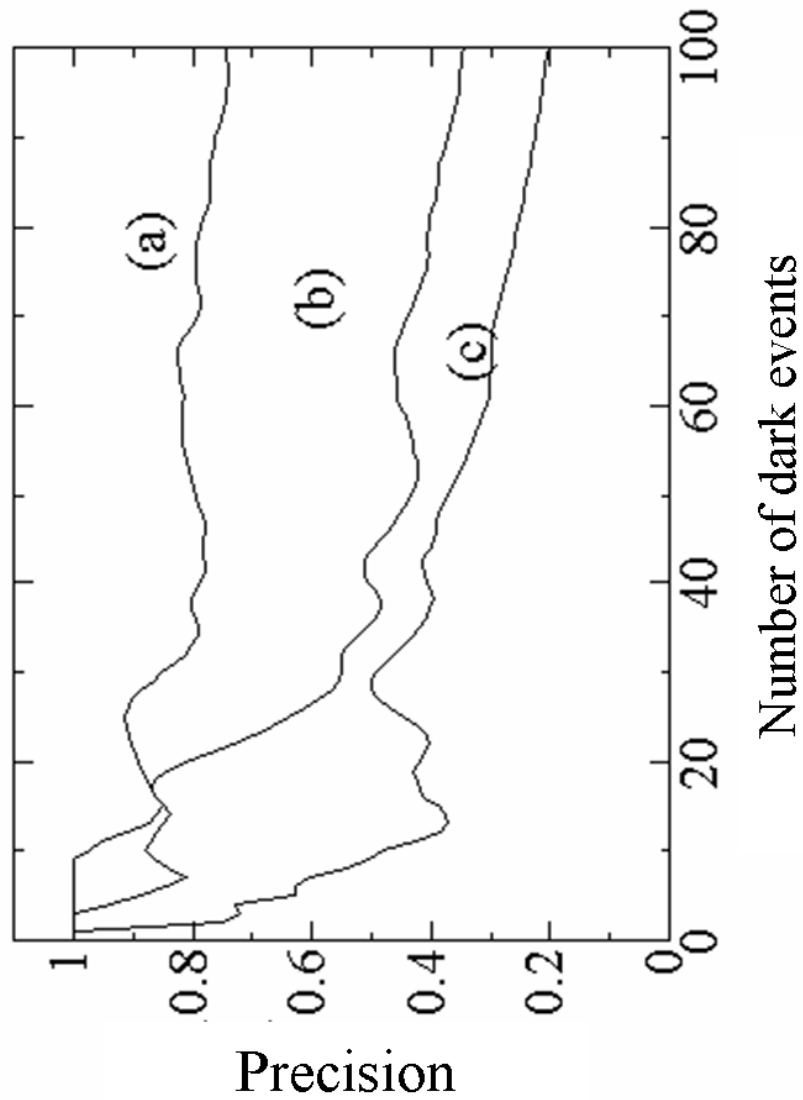


図 7: 取り出したバスケットに消去したノードが含まれていた比率 (プレジジョン  $p$ ) . 横軸は取り出したバスケットの数  $b^{\text{ret}}$  である . 事前知識としてのクラスタ数  $|c| = 20$  としてアルゴリズムを実行した場合である . (a) プライマリハブ (0-00) を消去, (b) セカンダリハブの例 (1-00) を消去, (c) その他のノードの例 (1-01) を消去して , アルゴリズムが消去したバスケットを取り出せるか測定した .

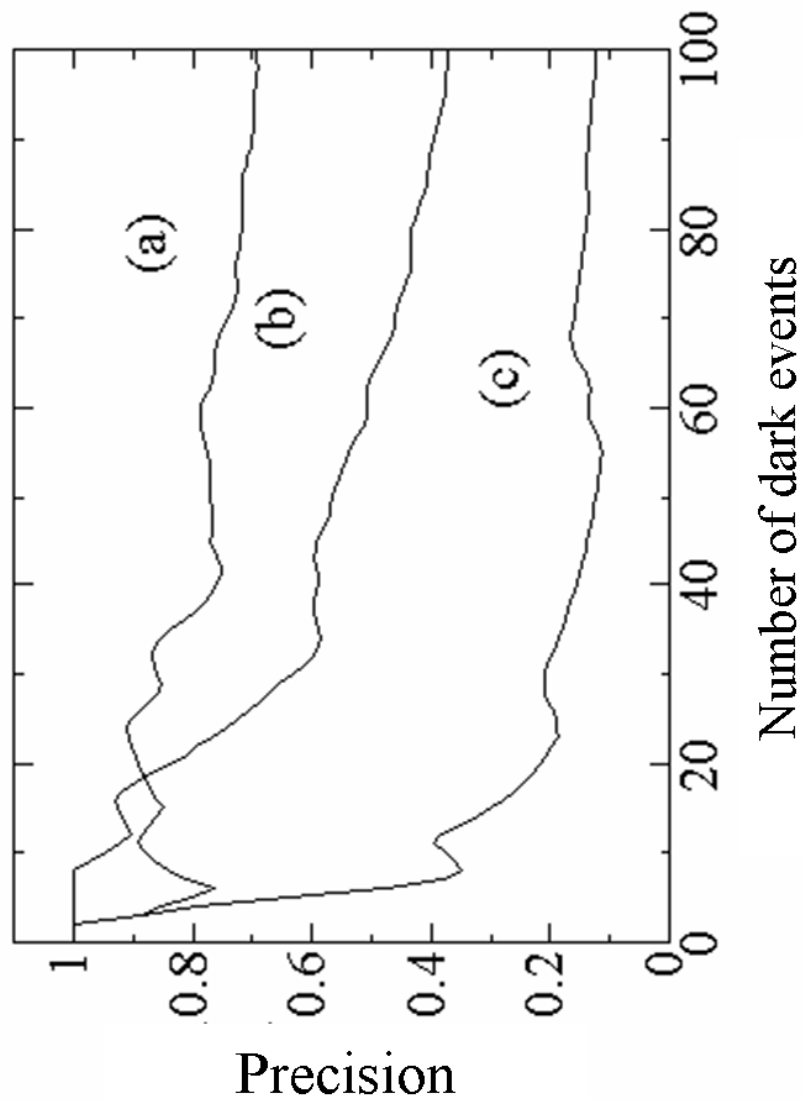


図 8: 取り出したバスケットに消去したノードが含まれていた比率 (プレジジョン  $p$ ) . 横軸は取り出したバスケットの数  $b^{\text{ret}}$  である . 事前知識としてのクラスタ数  $|c| = 11$  としてアルゴリズムを実行した場合である . (a) プライマリハブ (0-00) を消去, (b) セカンダリハブの例 (1-00) を消去, (c) その他のノードの例 (1-01) を消去して , アルゴリズムが消去したバスケットを取り出せるか測定した .



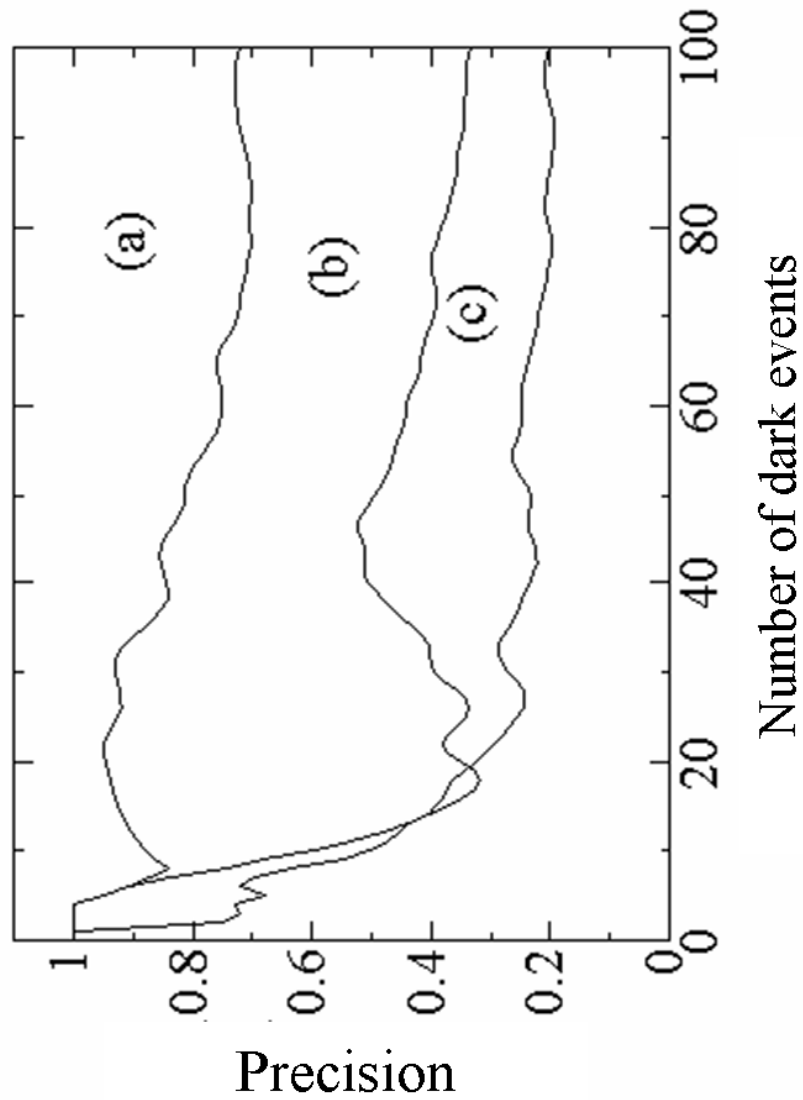


図 9: 取り出したバスケットに消去したノードが含まれていた比率 (プレジジョン  $p$ ) . 横軸は取り出したバスケットの数  $b^{\text{ret}}$  である . 事前知識としてのクラスタ数  $|c| = 29$  としてアルゴリズムを実行した場合である . (a) プライマリハブ (0-00) を消去, (b) セカンダリハブの例 (1-00) を消去, (c) その他のノードの例 (1-01) を消去して , アルゴリズムが消去したバスケットを取り出せるか測定した .

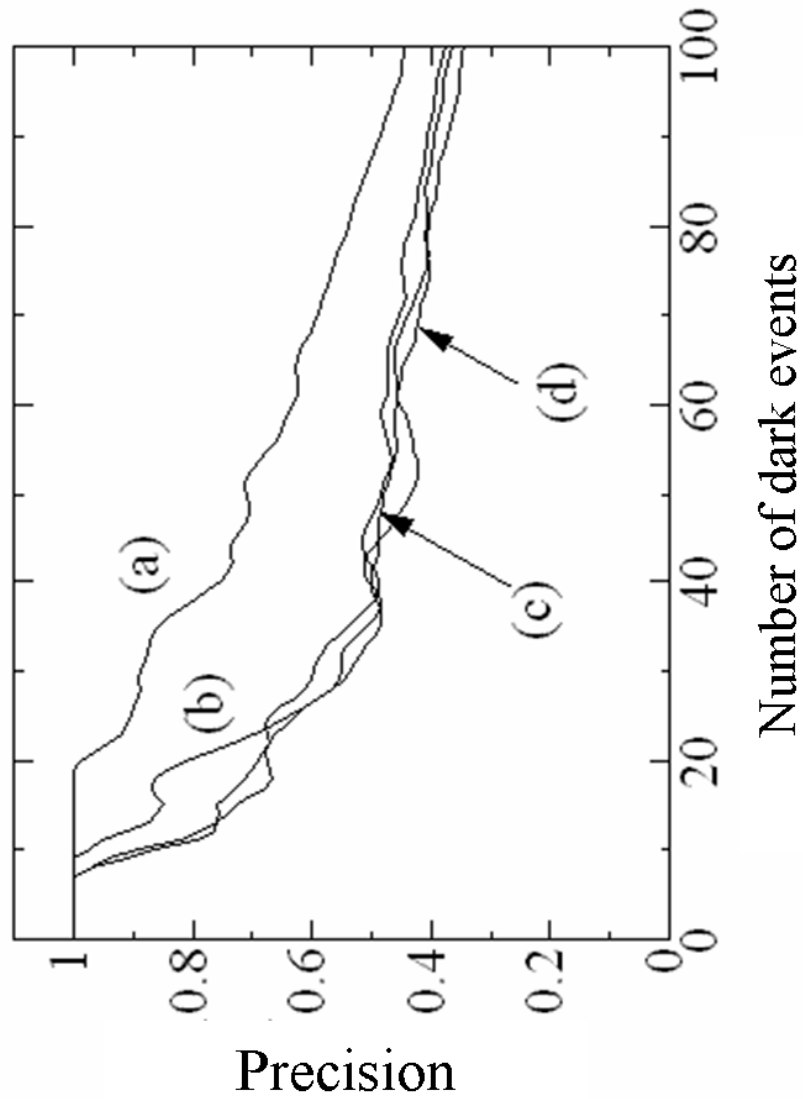


図 10: 取り出したバスケットにセカンダリハブ (1-00) が含まれていた比率 (プレジジョン  $p$ ) . 横軸は取り出したバスケットの数  $b^{\text{ret}}$  である . 事前知識としてのクラスタ数  $|c| = 20$  としてアルゴリズムを実行した場合である . グループ間接続確率  $p_g =$  (a) 0.019, (b) 0.057 (図 5 に対応する場合), (c) 0.095, (d) 0.19 (図 6 に対応する場合) である .

提案した．

ヒューマンインタラクティブアニメーションは，クラスタ間の隙間についての仮説を創出するプロセスを体系化したものである．隙間とは，関連性の高い事象群（クラスタ）の間の空白地帯でありながら，クラスタとの関係性が観測された領域である．ヒューマンインタラクティブアニメーションにおけるデータ処理手法として，見えない事象を表出化するアルゴリズムを提案し，精度と安定性について，基本的な特性を確認した．

## 4 ネットワーク上の観測できないノードの発見

本章では，3.3章で述べたランキング関数やネットワークのコミュニケーションの強度が，提案手法にどのように影響するのか，詳細に検討する．7.4.3節と同じ評価の手順を用いる．前半では不均一な組織構造のネットワークモデル，後半では均一な組織構造のネットワークモデルを扱う．均一性は，次数のようなノードのローカルな特徴のばらつき具合に対応している．不均一な組織構造は，インターネット上のコミュニティの自発的に形成される人間関係によく見られる．リーダーや人気のある人物に，人間関係が集中する傾向がある．スケールフリーネットワークが典型的な例である．一方，均一な組織構造は，特定の個人の露出を嫌う犯罪組織やテロ組織の観測事例でよく見られる．捜査機関が重要人物を特定できると，組織の存在に重大な影響が出るためである．

### 4.1 不均一な組織構造

#### 4.1.1 ネットワークモデル

7.4.3節と同じシミュレーションによって，組織のネットワークモデルからテストデータを生成し，アルゴリズムの性能を評価する．観測できない見えない重要ノードを探るヒントとなる情報を精度良く提供できるか調べる．代表的な不均一な組織のモデルは，スケールフリーネットワークである．シミュレーションに使用する1,000ノードのネットワークを図11に示す．ネットワークの成長過程で，新しいノードが古いノードにリンクを張る確率が古いノードの次数に比例する，Barabási-Albertモデル [Barabási 1999b] を用いた．次数の平均  $\bar{d} = 3.5$  で，分布はベキ乗則に従う ( $\text{Freq}(d) \propto d^{-2.4}$ )．ヒューマンネットワークに特有な，次数の高いノードが隣接する assortative mixing 特性 [Newman 2003a] を持っている．次数の分布を図12に示す．次数15以上のノードが14個，次数10以上のノードが42個，次数5以上のノードが170個存在する．次数13のノードID=1.00は，平均次数の4倍を持つハブノードの例である．次数21のノードID=42.00は，平均次数の6倍を持つ，大規模なハブノードの例である．

テストデータは，ネットワーク上を伝播するコミュニケーションのモデル化を想定して，以下の2つの手順で生成する．7.4.3節で述べた手順を再掲する．

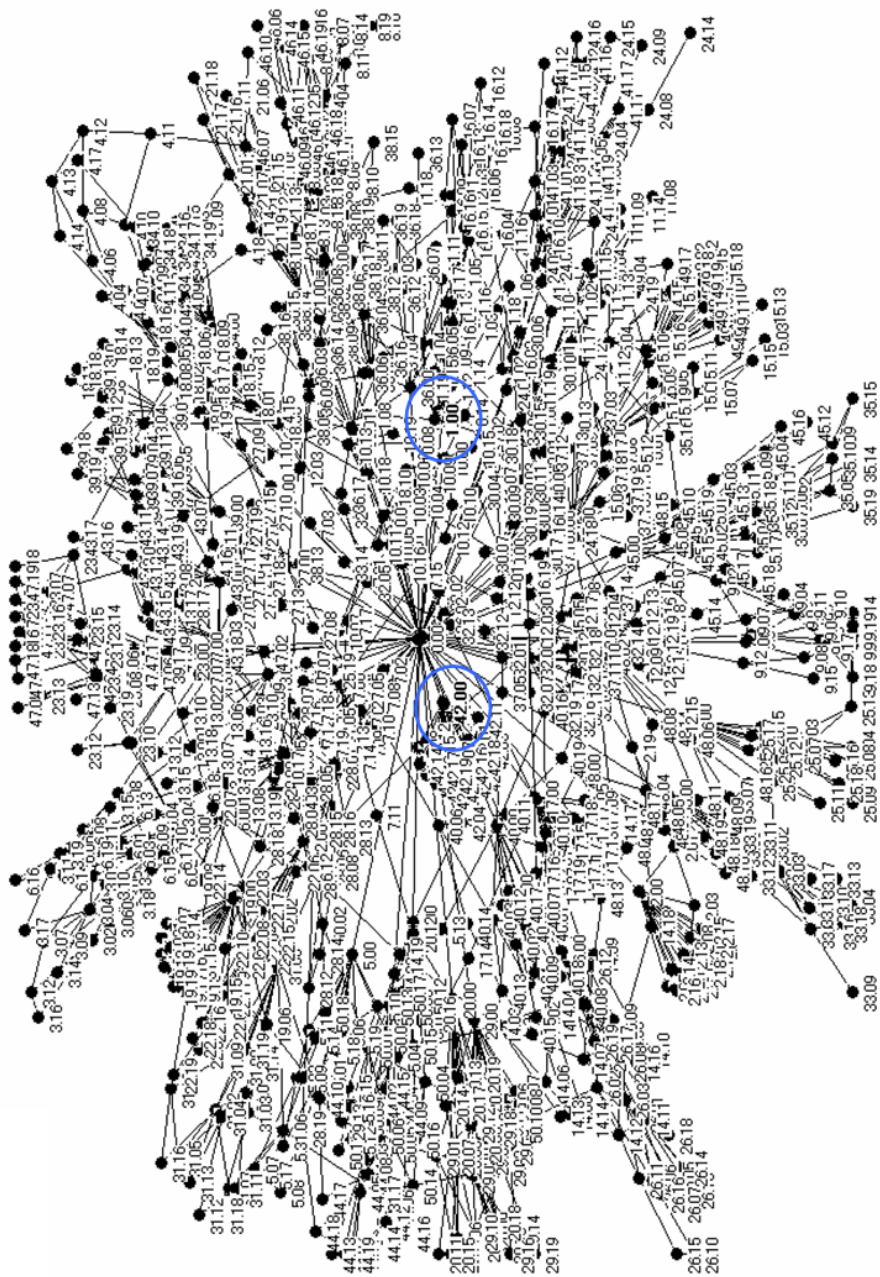


図 11: 組織のモデルとして、テストデータの生成に使用する 1,000 ノードのネットワーク。次数 13 のノード ID=1.00 は、平均次数の 4 倍を持つハブノードの例である。次数 21 のノード ID=42.00 は、平均次数の 6 倍を持つ、大規模なハブノードの例である。これらを円で囲んだ。

1. 手順 1: あるノードを起点として、周辺ノードへ伝播するコミュニケーションに含まれるノード群をひとつコミュニケーションバスケットとするデータを生成する。ある起点ノードから一定の距離以下のノード群をコミュニケーションへの参加ノード群としてバスケット化する。起点ノードに相当する人物が、ある話題を提供し、その周辺ノードが会話に参加する状況がひとつのバスケットに対応する。この話題の伝播範囲は、リンクのホップ数で数えた距離で表現できる。評価では、2 ~ 4 ホップで試行ごとに変更できるパラメータとした。起点ノードは、1,000 ノードのなかの任意のノードであり、バスケットごとにランダムに選ばれる。評価では、250 ~ 1,000 バスケットを生成し、試行ごとに変更できるパラメータとした。例えば、ノード ID=1.00 を起点ノードとするバスケットは、 $\beta_0 = \{0.00, 1.01, \dots, 2.00, 2.01, \dots, 3.00, \dots\}$  となる。ノード ID=1.00 の発言に、周辺のノードが応え、それらの会話への参加者をひとつのコミュニケーション履歴としてまとめたものに相当する。
2. 手順 2: 手順 1 で生成したコミュニケーションバスケットから特定のノードを消去することで、潜在構造（見えないノード）を織り込んだテストデータとする。バスケットから、ノード ID=1.00 を消去する。消去されたノードは、観測にかからない見えない構造に対応する。例えば、ノード ID=1.00 を見えない構造とする評価では、手順 1 のバスケットの例は、 $b_0 = \{1.01, \dots, 2.00, 2.01, \dots, 3.00, \dots\}$  となる。ノードが複数のバスケットに出現していれば、複数のバスケットに見えない構造が織り込まれる。ここで、ネットワークにおいてではなく、データにおいて、ノードを消去したことに注意する。消去によらず、ネットワークは同一で、コミュニケーションのパターン  $\beta_i$  は変化しない。同じネットワークで、特定のノードに関する情報のみが観測にかからなくなることに対応する。見えない重要人物が電話で指示を与える場合に、他の人物のメールの監視から得られるコミュニケーションの履歴が該当する。上記のバスケットがアルゴリズムの入力となる。アルゴリズムは、すべてのバスケットをランキングし、見えない構造に関係しているバスケットとして、 $b_i \neq \beta_i$  となっているバスケットを取り出すよう動作する。複数のバスケットからノードが消去されていれば、アルゴリズムは、これらすべてを取り出すことを目標とする。

#### 4.1.2 コミュニケーション強度を特徴付ける指標の導入

組織内のコミュニケーションを特徴付けるふたつの尺度を導入する．データにおけるコミュニケーションのきっかけを作る発信ノードの比率を表す発信者強度を式 (26) で定義する．発信されたコミュニケーションへの参加ノードの平均比率を表す参加者強度を式 (27) で定義する．どちらも，値が1に近いほど，コミュニケーションの強度が強く活発である．参加者強度は，コミュニケーションの伝播ホップ数  $h$  と相関がある．式 (28) で近似できる．

$$R_{in} \equiv \frac{\text{コミュニケーションの発信ノード数}}{\text{全ノード数}}. \quad (26)$$

$$R_{pa} \equiv \frac{\text{コミュニケーションの参加ノード数}}{\text{全ノード数}}. \quad (27)$$

$$R_{pa} \propto \sum_{n=0}^h \bar{d}^n = \frac{\bar{d}^{h+1} - 1}{\bar{d} - 1}. \quad (28)$$

#### 4.1.3 精度のランキング関数に対する依存性

3.3 節で述べたランキング関数が，アルゴリズムの精度にどう影響するかを調べる．アルゴリズムがランキングし，取り出したダミーノード (バスケット) の精度を測定した．テストデータには，クラスタ数に関する事前理解が含まれていない．しかし，アルゴリズムは，クラスタ数に敏感には依存しないので，およその想定値で代用する．次数の大きなハブノード数と同オーダーとし，クラスタ数  $|c| = 100$  とした．図 11 のノード ID=1.00 を観測できない見えないノードとして，観測データから消去して，アルゴリズムで復元する試行を行った． $I_{nu}$  ,  $I_{av}$  ,  $I_{sd}$  ,  $I_{tp}$  を用いて測定したプレジジョンを図 13 に示す．横軸は，取り出したバスケット数の全バスケットの数に対する比率  $m^{ret}/|b|$  である．取り出したバスケットにおいて，ノード ID=1.00 が消去されていれば，正解とみなす．横軸は，全バスケットの数に対して，取り出したランキング上位のバスケットの数の割合である．多数のバスケットを取り出すと，不正解のバスケットが増えていくので，プレジジョンは低下していく．右下がりのグラフになる．ランキング関数 (a) $I_{nu}$  , (b) $I_{av}$  , (c) $I_{sd}$  , (d) $I_{tp}$  を用いて測定した．

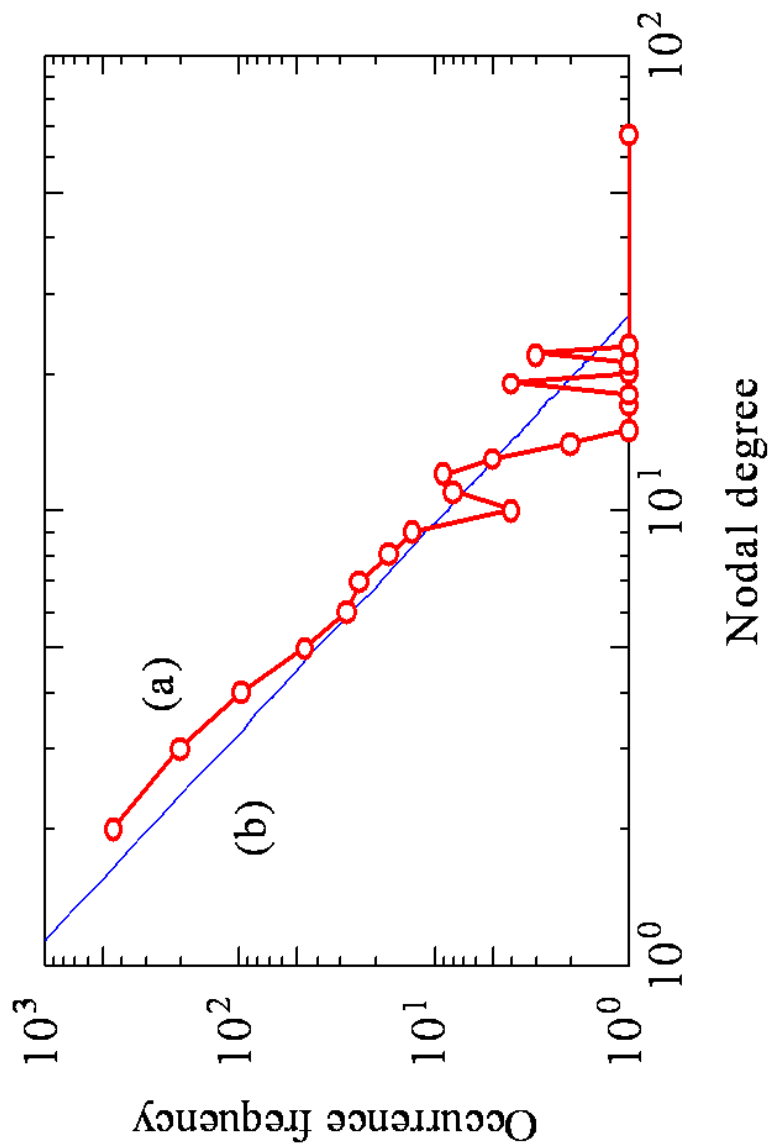


図 12: (a) 図 11 に示すネットワークの次数の分布 (スケールフリー性),  
 (b)  $\text{Freq}(d) \propto d^{-2.4}$ .



$I_{nu}$  が最も性能がよいことがわかった．以後の測定では， $I_{nu}$  を使用する． $I_{nu}$  の値が同一の場合には， $I_{av}$  を併用する．

#### 4.1.4 精度のノードに対する依存性

ネットワーク上のノードの位置が，アルゴリズムの精度にどう影響するかを調べる．観測できない見えないノード ID=1.00 と ID=46.00 に対して，測定したプレジジョン  $p$  とリコール  $r$  を図 14 に示す．プレジジョンは式 (25)，リコールは式 (29) で定義される．

$$r = \frac{\sum_{i=0}^{m^{\text{ret}}-1} B(b_i^{\text{ret}} \neq \beta_i)}{\sum_{i=0}^{|b|-1} B(b_i^{\text{ret}} \neq \beta_i)} \quad (29)$$

横軸は，取り出したバスケット数の全バスケットの数に対する比率  $m^{\text{ret}}/|b|$  である．アルゴリズムが上位 5% にランキングしたバスケット  $b_i^{\text{ret}}$  は，プレジジョンが 100% とすべて正しく，これだけで 80% のリコールを達成している．ただし，残り 20% を見つけるのは難しいことがわかる．ノードの違いによる性能差は，ほとんど見られない．ノード 46.00 のような大規模ハブだけでなく，ノード 1.00 のような小規模ハブも同じように復元できる．

#### 4.1.5 精度のコミュニケーション強度に対する依存性

4.1.2 節で定義したコミュニケーション強度が，アルゴリズムの精度にどう影響するかを調べる．組織内のコミュニケーションを特徴付ける発信者強度と参加者強度に対して，プレジジョンの依存性を測定した．観測できない見えないノード ID=1.00 に対して，発信者強度  $R_{in}$  に対する依存性を図 15 に示す．条件は， $R_{pa} = 0.02$  ( $h = 2$ )，(a)  $R_{in} = 1$ ，(b)  $R_{in} = 0.5$ ，(c)  $R_{in} = 0.25$  とした．コミュニケーションの伝播ホップ数を  $h$  とする．差異は少なく，復元精度の高い推定をするのに，発信者強度はあまり重要でないことがわかった．ノード ID=1.00 において，参加者強度  $R_{pa}$  に対する依存性を図 16 に示す．条件は， $R_{in} = 1$ ，(a)  $R_{pa} = 0.02$  ( $h = 2$ )，(b)  $R_{pa} = 0.11$  ( $h = 3$ )，(c)  $R_{pa} = 0.49$  ( $h = 4$ ) とした．参加者強度が大きければ，プレジジョンは非常に高くなる．参加者強度が 2% 程度で，コミュニケーションの伝播距離が平均 2 ホップ程度と短くても，見えない重要人物を探るのに有効な情報が得られる．

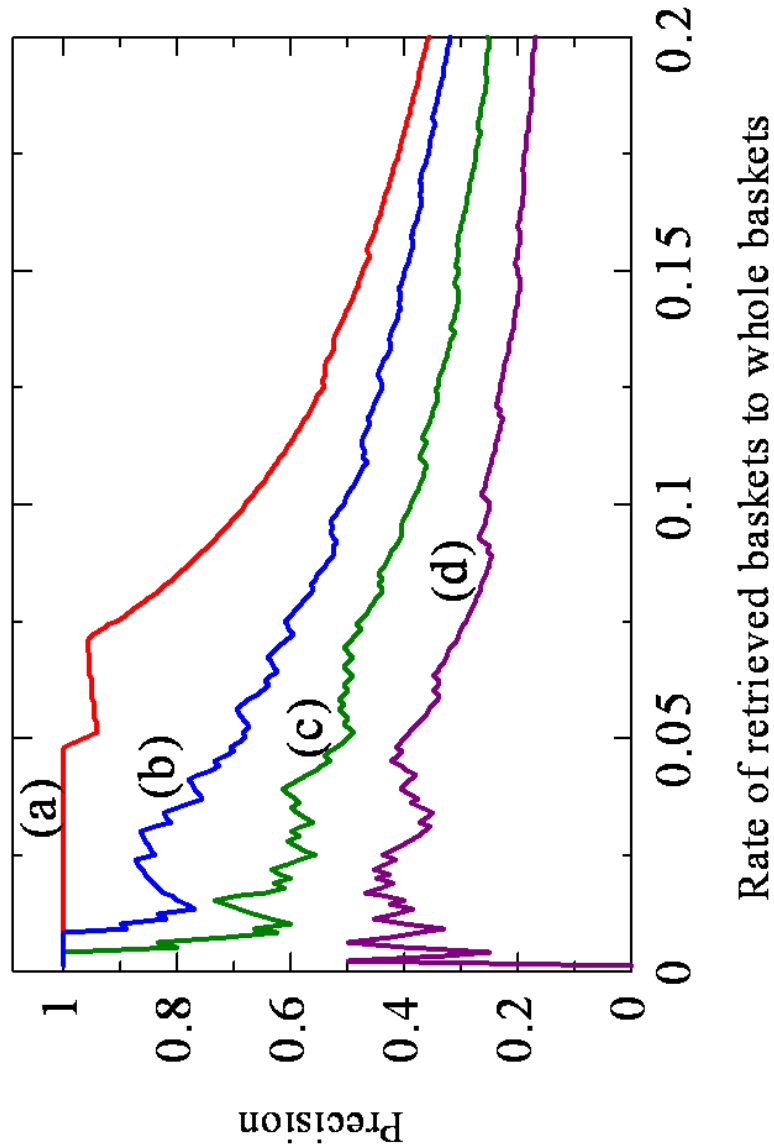
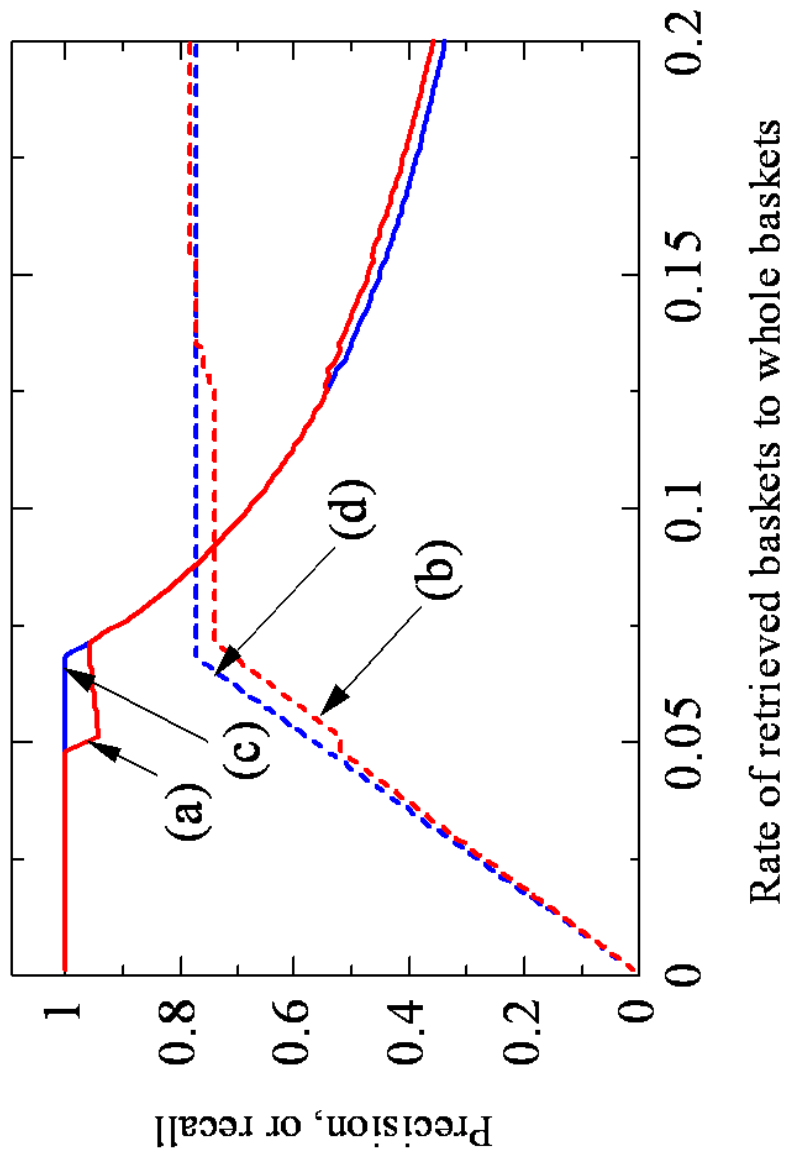


図 13: 図 11 において設定した, 観測できない見えないノード ID=1.00 に関連するバスケットを特定する実験で, ランキング関数 (a)  $I_{nu}$ , (b)  $I_{av}$ , (c)  $I_{sd}$ , (d)  $I_{tp}$  を用いて測定したプレジジョン. 横軸は, 取り出したバスケット数の全バスケットの数に対する比率  $m^{ret}/|b|$  である.  $I_{nu}$  が最も性能がよいことがわかる.



s

図 14: 図 11 において設定した観測できない見えないノード ID=1.00 に関連するバスケットを特定する試行において測定した (a) プレジジョン  $p$ , (b) リコール  $r$ , 観測できない見えないノード ID=46.00 に関連するバスケットを特定する実験において測定した (c) プレジジョン  $p$ , (d) リコール  $r$ . 横軸は, 取り出したバスケット数の全バスケットの数に対する比率  $m^{\text{ret}}/|b|$  である. 上位 5% にランキングしたバスケット  $b_i^{\text{ret}}$  は, プレジジョンが 100% とすべて正しく, これだけで 80% のリコールを達成している.

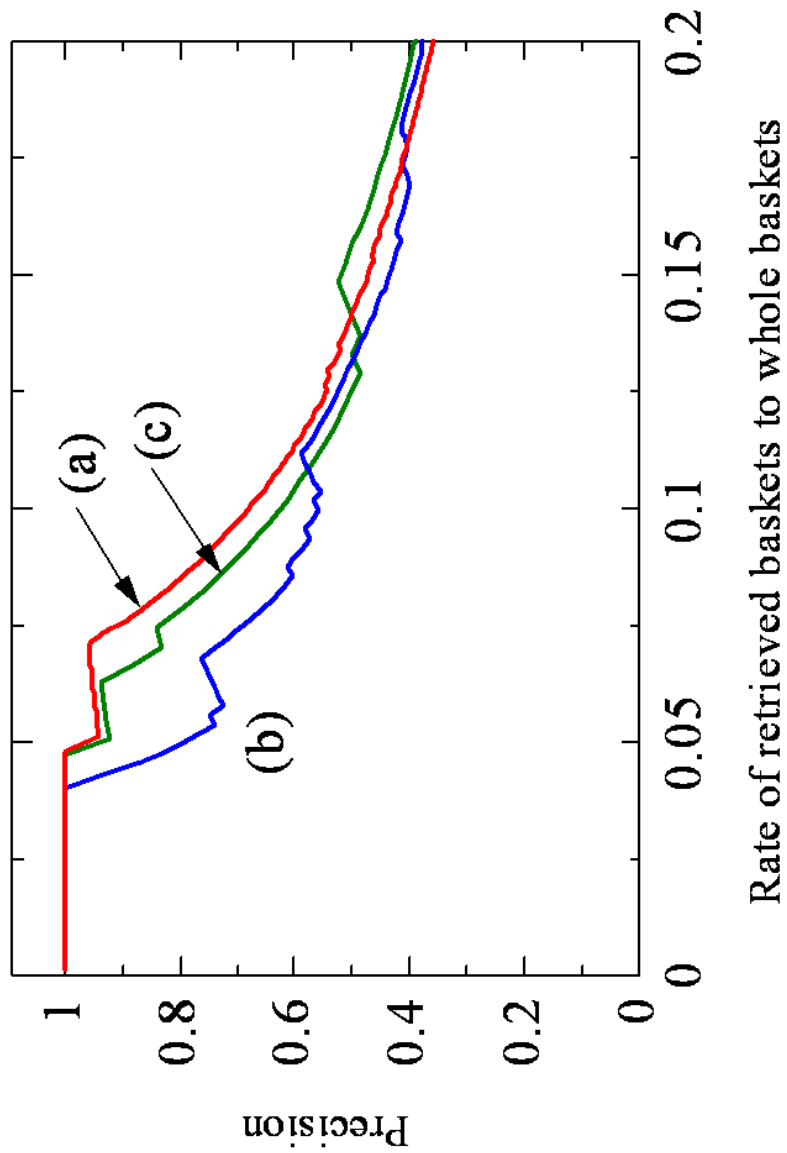


図 15: 観測できない見えないノード ID=1.00 に関連するバスケットを特定する試行において測定した, 発信者強度  $R_{in}$  に対するプレジジョン  $p$  の依存性.  $R_{pa} = 0.02$  ( $h = 2$ ), (a)  $R_{in} = 1$ , (b)  $R_{in} = 0.5$ , (c)  $R_{in} = 0.25$ . 横軸は, 取り出したバスケット数の全バスケットの数に対する比率  $m^{\text{ret}}/|b|$  である. 復元精度の高い推定をするのに, 発信者強度はあまり重要でないことがわかる.

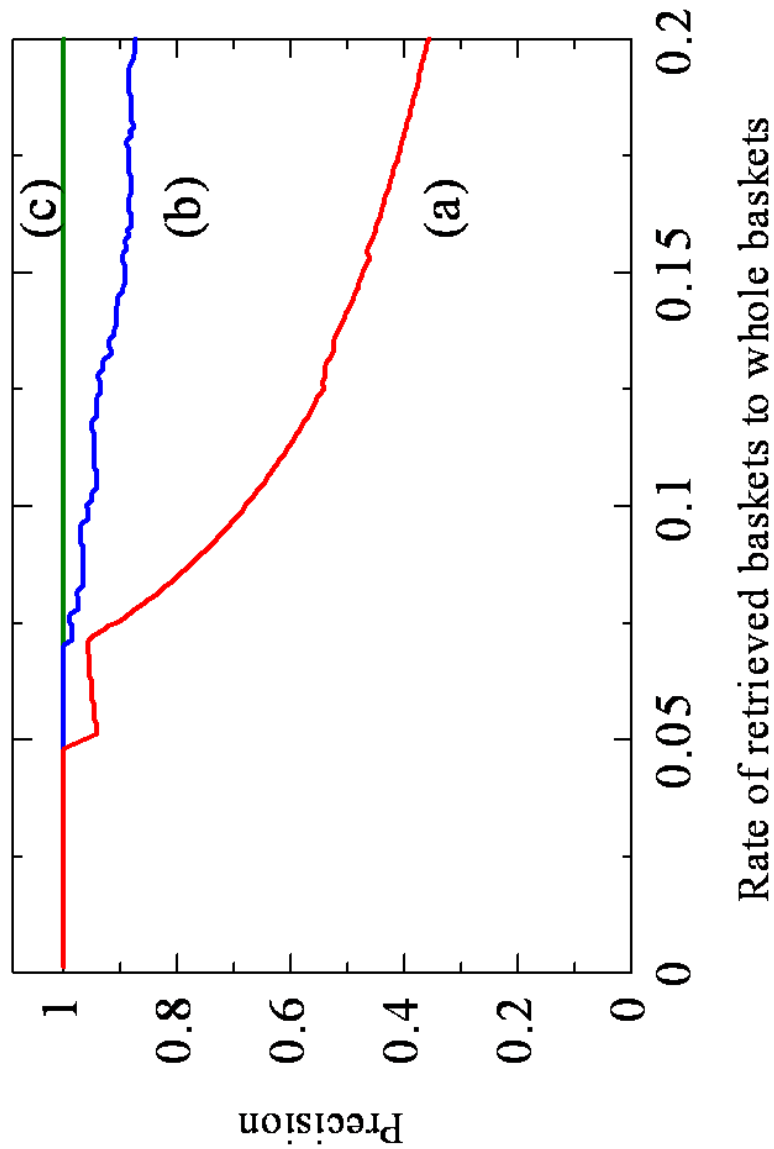


図 16: 観測できない見えないノード ID=1.00 に関連するバスケットを特定する試行において測定した, 参加者強度  $R_{pa}$  に対するプレジジョン  $p$  の依存性.  $R_{in} = 1$ , (a)  $R_{pa} = 0.02$  ( $h = 2$ ), (b)  $R_{pa} = 0.11$  ( $h = 3$ ), (c)  $R_{pa} = 0.49$  ( $h = 4$ ). 横軸は, 取り出したバスケット数の全バスケットの数に対する比率  $m^{ret}/|b|$  である. 参加者強度が 2%程度で, コミュニケーションの伝播距離が平均 2 ホップ程度と短くても, 見えない重要人物を探るのに有効な情報が得られることがわかる.

## 4.2 均一な組織構造

### 4.2.1 ネットワークモデル

4.1.1 節と同じ手順で、均一な組織構造について、評価を行う。995 ノードの均一な組織構造のモデルを図 17 に示す。左上のグラフは、995 ノードの次数の存在確率の分布を示す。横軸は、平均値で規格化した次数 ( $d/\mu(d)$ ) である。図 11 の不均一なネットワークとは異なり、次数のばらつきは小さい。図の中のノード [a] (赤) は、もっとも次数が大きなノードである。ノード [b] (青) は、他のノードへの距離 (ホップ数) のばらつき (標準偏差の値) が最も小さいノードである。ノード [c] (緑) は、他のノードへの距離の平均値が最も小さいノードである。これらの 3 つのノードを 3 回分のプレジジョン測定の試行に使用する。

### 4.2.2 精度のノードに対する依存性

ネットワーク上のノードの位置が、アルゴリズムの精度にどう影響するかを調べる。図 17 において設定した、観測できない見えないノード [a], [b], [c] に関連するバスケットを特定する試行で測定したプレジジョン  $p$  を図 18 に示す。横軸は、取り出したバスケット数  $m^{\text{ret}}$  である。

もっとも次数が大きなノード [a] の見つける場合には、それほど大きなプレジジョンを得ることができない。均一なネットワークでは、ノードごとの次数の差が小さいため、ノード [a] の欠落を検出することは難しい。しかし、 $m^{\text{ret}}$  が小さくところでは、大きなプレジジョンが得られている。次に、他のノードへの距離 (ホップ数) のばらつき (標準偏差の値) が最も小さいノード [b] と、他のノードへの距離の平均値が最も小さいノード [c] では、大きなプレジジョンを得ることができた。

## 4.3 まとめ

不均一な組織構造を均一な組織構造とのモデルとなる、ネットワーク上の観測できない見えないノードを発見する問題に取り組んだ。1,000 ノードの不均一な組織構造では、精度のランキング関数に対する依存性を測定し、 $I_{\text{nu}}$  によるランキングが最も優れていることがわかった。コミュニケーション強度を特徴付ける指標について、発信者強度  $R_{\text{in}}$  に対する依存性は少ない。参加者強度  $R_{\text{pa}}$  が組織全体の 2% 程度であっても、大きな

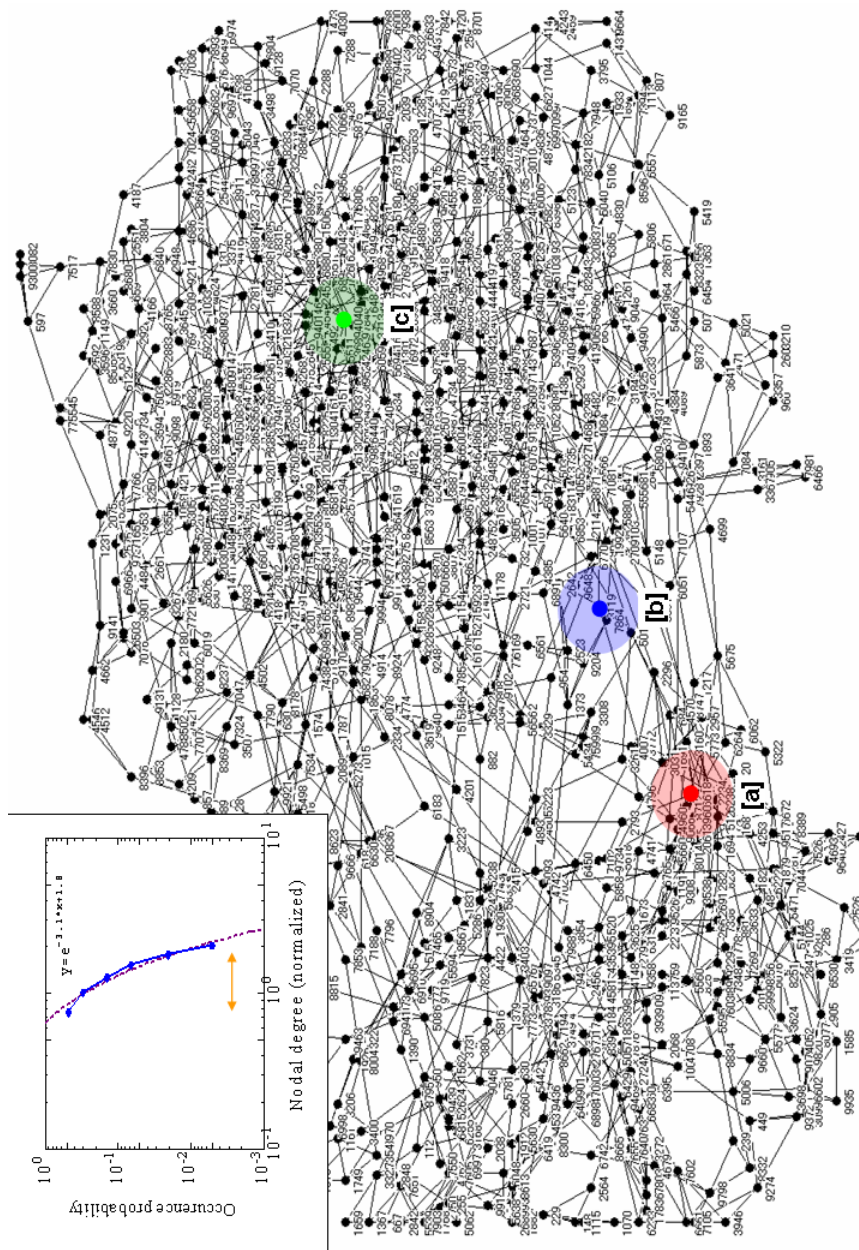


図 17: 995 ノードの均一なネットワーク. 左上は, ノードの次数の存在確率の分布を示す. ノード [a] (赤) は, もっとも次数が大きなノードである. ノード [b] (青) は, 他のノードへの距離 (ホップ数) のばらつき (標準偏差の値) が最も小さいノードである. ノード [c] (緑) は, 他のノードへの距離の平均値が最も小さいノードである.

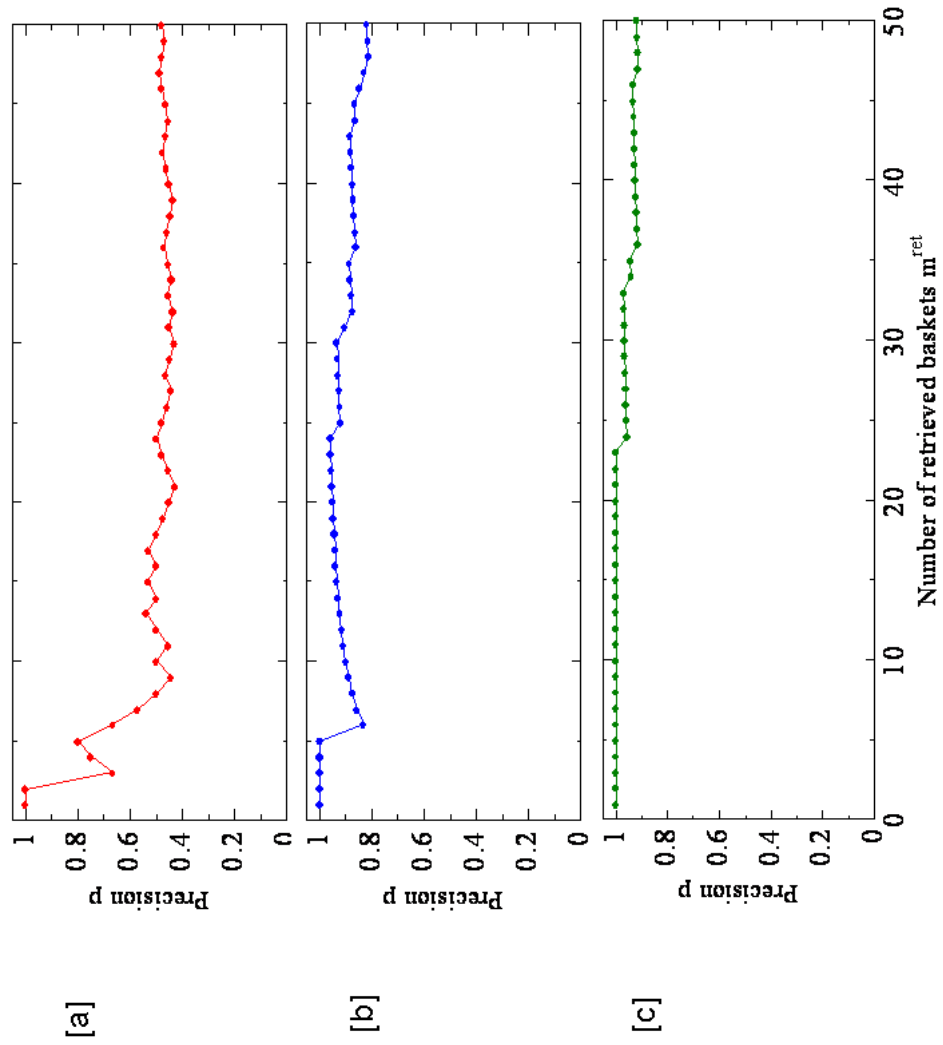


図 18: 図 17 において設定した, 観測できない見えないノード [a], [b], [c] に関連するバスケットを特定する試行で測定したプレジジョン  $p$ . 横軸は, 取り出したバスケット数  $m^{\text{ret}}$  である.



プレジジョンを達成することができ，見えない重要なノードを発見するヒントが得られた．995 ノードの均一な組織構造の場合でも，大きなプレジジョンを達成することができ，見えない重要なノードを発見するヒントが得られた．

今後の課題についても述べる．コミュニケーションに付随する発生タイミングや発生順序といった時系列情報の活用が重要になるだろう．バッチ処理ではなく，オンライン処理も重要である．実用面を考慮すると，大量の観測データを効率的に処理することも必要になるだろう．

## 5 意思決定の隠れた黒幕を探る実験

### 5.1 電子メールによる組織の意思決定の方法

現実的な環境下で、見えない重要人物を探る実験を行った。ヒューマンネットワークを表すシナリオマップから、見えない黒幕についての仮説を創出し、検証する。組織の意思決定を目的とした電子メールによるコミュニケーションを観測データとした。研究室のバーベキュー大会のためのメーリングリストを用意し、企画の内容、参加者の募集、会場や食材の準備、作業分担などを決定する。

5.2 節の予備実験では、意思決定を誘導する黒幕を設定せず、4.1.1 節のシミュレーションと同じ内容を実際に実行した。後で、議論の流れから自然に生まれた、最も影響力のあった人物が誰か選定する。この人物が、テストデータを生成する手順で消去するハブノードとなる。一方、5.3 節の本実験では、あらかじめ、意思決定を誘導する黒幕を設定する。黒幕はメーリングリストを使わずに指示を与え、観測データに現れることはない。

5.3 節の本実験では、参加者 20 人で、電子メールの交換を行った。参加者は、組織の意思決定として、期日までにバーベキューの詳細なプランを決定する。事前に、役割分担や議論の方法は、設定されていない。前提のない状態から議論を始める。この他に、あらかじめ設定された、意思決定を誘導する見えない黒幕がいる。黒幕はメーリングリストを使わずに指示を与え、観測データに現れることはない。黒幕は、4 章における、グループ間を結びつけるハブノード、ノード間をつなぎ合わせるハブノードのような役割を担っていると想定する。「黒幕を見つける問題 = 見えない重要なハブを見つける問題」と読みかえて、3 章で提案したプロセス、アルゴリズムを適用する。

黒幕の任務は、期日までに、組織の総意として、参加者にプランを決定させることである。黒幕は、参加者が結論に至ることを優先し、適切な話題、効果的な議事進行、積極的な発言などを指示する。決定内容を操作することはなく、モデルプランを用意して、決定内容を誘導するものでもない。状況に応じて、被指示者として、複数の参加者に同時に指示を与えてもよい。しかし、いつ、誰に、どんな文面のメールを送信するか、といった具体的な行動は、参加者が自律的に判断する。参加者は、黒幕からの指示に、服従する必要はない。指示を参考にしながら、適切な議論の方向性や結論への道筋を判断する。この実験条件では、シミュ

レーションとは異なり，コミュニケーションの解釈の多義性，状況依存的な心理の起伏といった現実の条件を反映することになる．

1ヶ月の実験で，57個の subject について，220通のメールが交換された．メールリストからメールを採取し，subject により分類されたメール群から，バスケットデータを作成する．バスケットが含む事象は，subject に対して，メールを送信，返信した人物の識別子 (PID) である．例えば，バーベキュー開催，というメールを送信して議論を呼びかけた人物と，それに応えて，re: バーベキュー開催，というメールを返信した人物の集合である．分類されたメール群からひとつのバスケットを作成するため，ダミーノードとメール群の subject (その subject で交換されたメールの内容) とは，1対1に対応する．

メールの観測データから計算したコミュニケーションの強度は， $R_{in} = 0.86$ ,  $R_{pa} = 0.14$  だった．参加者強度が充分強く，アルゴリズムが機能する条件にある．図 16 から， $R_{pa} = 0.11$  の場合，ランキング上位 20% のバスケットで，プレジジョンが 90% 程度であることがわかる．この知見を参考に，ランキング上位 15 個のダミーノードを確からしい隠れた重要人物の候補として，分析することにした．

## 5.2 影響力のある見えない重要人物を探る実験

参加者 11 人で，電子メールの交換を行った．メールの観測データから計算したコミュニケーションの強度は， $R_{in} = 0.45$ ,  $R_{pa} = 0.32$  だった．参加者強度が充分強く，アルゴリズムが機能する条件にあると見込まれる．発言力が大きいと感じられたのは，人物  $PID_1$  だった． $PID_1$  をバスケットデータから消去し，見えない重要人物とした．

この人物を復元したシナリオマップを図 19 に示す．最も高くランキングされたダミーノード  $DE_{10}$  を示した．見えない重要人物は， $DE_{10}$  に対応するバスケット  $b_{10}$  に含まれ， $PID_9$ ,  $PID_{18}$ ,  $PID_{20}$  との間に活発なコミュニケーションがあるとの仮説を導いた．実際に，バスケットデータを作成する際に，バスケット  $b_{10}$  から人物  $PID_1$  を削除しており，この仮説は正しかった．バスケットデータのどの部分で，人物  $PID_1$  を削除したのか，特定することができ，見えない重要人物に関するヒントが得られた．

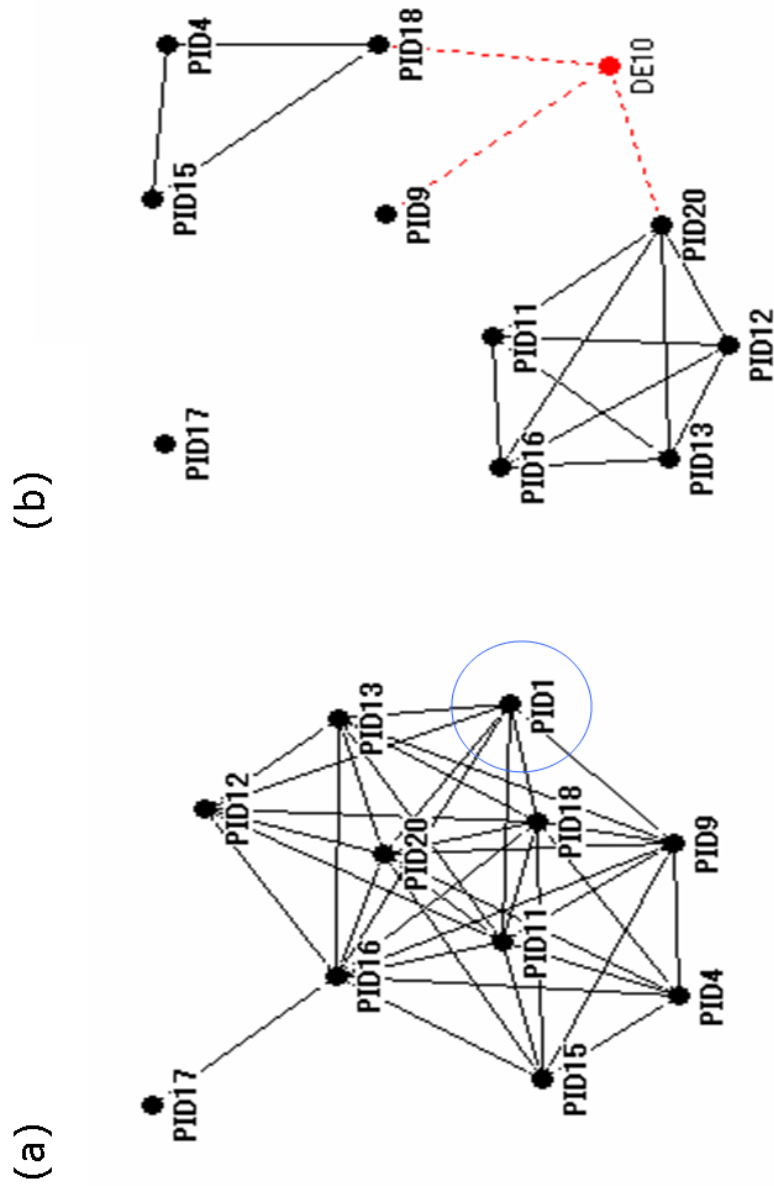


図 19: 消去した, 議論に最も影響力があった人物 PID<sub>1</sub> を復元したシナリオマップ (ヒューマンネットワーク). (a) 人物 PID<sub>1</sub> を消去する前, (b) 人物 PID<sub>1</sub> を消去した後, 復元したシナリオマップ.

クラスタ	解釈された特徴
1	時折，重要な議題を提出し議論を活性化する
2	新しい参加者で，議論に乗れていない
3	発言が活発だが議論の方向付けが不得手である
4	結論につながる明快な発言がなされる
5	重要な決定事項にのみ参加する傾向がある

表 1: クラスタに属する人物の特徴の解釈．

### 5.3 組織の意思決定での見えない黒幕を探る実験

#### 5.3.1 黒幕からの指示内容の一致性

ヒューマンインタラクティブアニメーションの3回目の繰り返し後のシナリオマップを図20に示す．図が煩雑なため，ダミーノードについては，クラスタへのリンクを2本ずつ図示した．人間の解釈として，クラスタに属する人物の議論の特徴を表1にまとめた．クラスタ1は，時折，重要な議題を提出し議論を活性化する，と解釈された．クラスタ3は，発言が活発だが議論の方向付けが不得手である，と解釈された．クラスタ4は，結論につながる明快な発言がなされる，と解釈された．

被験者はすべて研究室の関係者で，彼らの特徴を熟知した上で，シナリオマップを解釈する前提となっている．それぞれのクラスタははっきりとした特徴がある．議論の内容だけでなく，人物の考え方や行動特性，研究室との係わり（所属学生，卒業生，外部からの招待者などのステータス）の面でも，まとまり感のあるクラスタ群である．

ダミーノードが最も多く表出しているクラスタ1と4とに注目する．ヒューリスティクスとして，クラスタ間の空間でダミーノード数がより多い領域から着目するよう推奨されている [Ohsawa 2006]．特に，4つのダミーノードにつながったクラスタ1のPID<sub>4</sub>に着目する．図21に，PID<sub>4</sub>につながる4つのダミーノードDE<sub>8</sub>, DE<sub>35</sub>, DE<sub>39</sub>, DE<sub>52</sub>について，クラスタへのリンクをすべて図示した．4つのダミーノードすべてが，クラスタ1, 3, 4をつないでいる．表1の解釈によると，独立性の高そうに見えるクラスタ1, 3, 4をまとめ上げ，一貫性のある組織の意思決定を誘導する，見えない黒幕が存在する可能性が窺える．

見えない黒幕が誘導している可能性がある意思決定の詳細を明らかにするため，4つのダミーノードが対応するバスケット  $b_8, b_{35}, b_{39}, b_{52}$  に分類

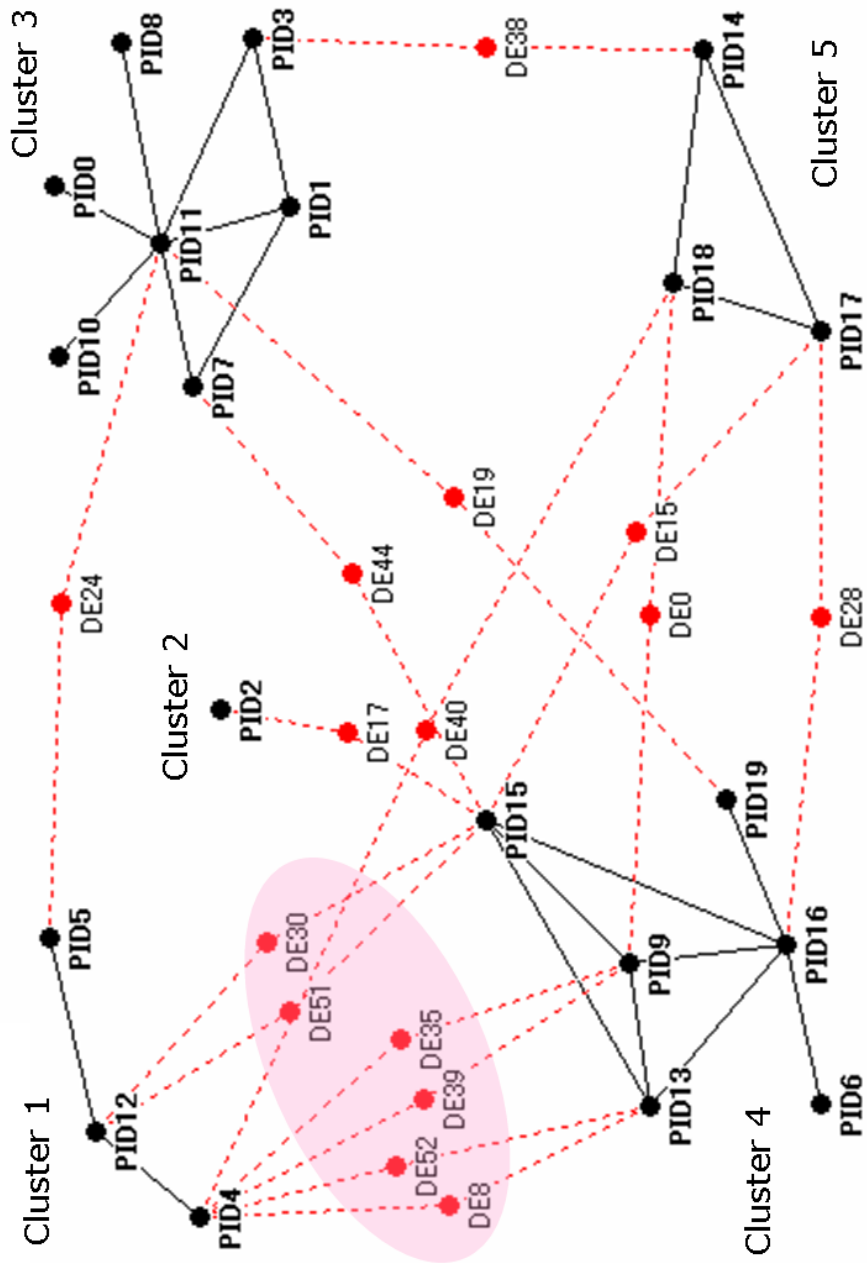


図 20: ヒューマンインタラクティブアニーリングの3回目の繰り返し後において、見えない黒幕の候補を可視化したシナリオマップ（ヒューマンネットワーク）。

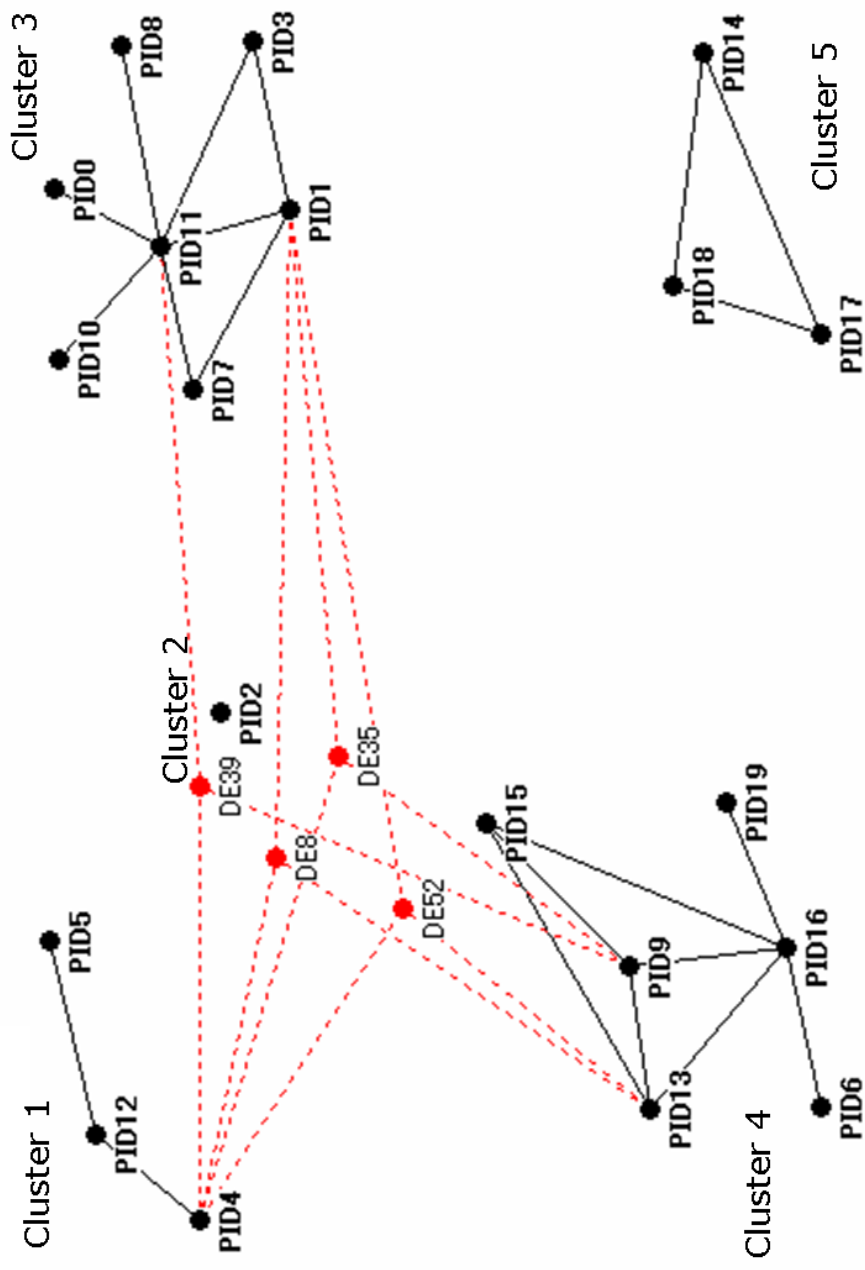


図 21: 見えない黒幕の候補として、図 20 でダミーノードが最も多く表出している、クラスタ 1 と 4 との間の 4 つのダミーノードと、4 つのダミーノードにつながったクラスタ 1 の人物 PID<sub>4</sub> に着目したシナリオマップ (ヒューマンネットワーク)。

バスケット	メールの subject	メール送受信者 PID
$b_8$	役割分担を決めよう	2, 4, 11, 12, 13, 15
$b_{35}$	場所決め	1, 4, 9, 11, 12, 16
$b_{39}$	設営版業務連絡	4, 9, 11, 12
$b_{52}$	企画の投票開始	1, 4, 11, 12, 13, 15, 16

表 2: 図 21 の 4 つのダミーノードに対応するバスケットに分類されたメール群 .

されたメール群の内容を調べた . バスケット  $b_8, b_{35}, b_{39}, b_{52}$  に分類されたメール群の内容を可視化した . 図 22 に , KeyGraph を示す . バーベキューの場所の選択 場所に適したプランの作成 プランに沿った分担の決定 , と議論の流れが表現されている . 地理的な環境要因を決定してから , バーベキュー当日までの準備作業をブレイクダウンしている . ダミーノードは , 前提から決定へ向かう議論の方向付けを巧みに演出しているように見える .

表 2 は , 4 つのバスケットの subject と送受信者をまとめたものである . メールの内容によると , バーベキューの場所の選択 場所に適したプランの作成 プランに沿った分担の決定 , と議論が流れていた . 地理的な環境要因を決定してから , バーベキュー当日までの準備作業をブレイクダウンしている . ダミーノードは , 前提から決定へ向かう議論の方向付けを巧みに演出しているように見える . この分析から , 次の仮説を導いた .

- 黒幕は , 発散的な特性を持つ人物群 (活発に論点を提供するクラスタ 1 と , 発言が多く継続的な議論に寄与するクラスタ 3) を収束的な特性を持つ人物群 (結論の方向性を明確にするクラスタ 4) に結びつけ , 議論を誘導し , 意思決定への収束を促していた .

仮説を検証するため , 黒幕へヒアリングを行い , 評価を行った . すべての指示内容と被指示者を記録してもらい , テキストで提出してもらった . 表 3 は , 12 個の指示である . 指示の内容と被指示者が , 表 2 の subject が対応するメール群とメール送受信者に一致するか比較した . 内容について , 黒幕が , メール群が指示を反映してなされた議論だと同意できるか , で判定した . 8 個の指示がダミーノードに対応するメール群で示されていることを確認した . 精度について , プレジジョンは 100% (4/4) , リコールは 67% (8/12) といえる .





	黒幕からの指示内容	内容 の一致	被指示者 PID の一致 (比率)
1	メーリングリストを 告知せよ	No	
2	外部から参加者を 招待せよ	No	
3	自己紹介せよ	No	
4	議題ごとのチームを 編成せよ	Yes ( $b_8$ )	1, <u>4</u> , 7, 9, <u>11</u> , <u>12</u> , <u>13</u> , 16, 17, 18 (0.4)
5	チームリーダーは役割に 集中せよ	Yes ( $b_8$ )	1, 7, <u>12</u> (0.33)
6	分担を割り当てる議論を 始めよ	Yes ( $b_8$ )	<u>12</u> , <u>13</u> (1.0)
7	特定の議題に集中せよ	No	
8	開催場所について 議論せよ	Yes ( $b_{35}$ )	<u>12</u> (1.0)
9	レシピについて結論せよ	Yes ( $b_{52}$ )	9, <u>16</u> (0.5)
10	分担の割り当てを 結論せよ	Yes ( $b_{52}$ )	<u>12</u> , <u>13</u> (1.0)
11	設営について告知せよ	Yes ( $b_{39}$ )	<u>4</u> (1.0)
12	詳細を告知せよ	Yes ( $b_{39}$ )	<u>4</u> , <u>11</u> (1.0)

表 3: 黒幕からの指示内容, 被指示者 (指示を与えられた参加者) と図 2  
との一致度合い. 下線で示す PID が, 図 2 の送受信者 PID と一致.

指示回数	被指示者 PID の一致 (比率)
7	<u>12</u> (1.0)
5	<u>13</u> (1.0)
4	<u>4</u> , <u>16</u> (1.0)
3	7 (0.0)
2	<u>1</u> , <u>9</u> , <u>11</u> , 17, 18 (0.6)
1	0, <u>2</u> , 3, 6, 8, 10, 14, 19 (0.13)

表 4: 黒幕からの指示回数ごとの被指示者 (指示を与えられた参加者) と表 2 との一致度合い。下線で示す PID が、表 2 の送受信者 PID と一致。

指示 5, 8, 10 のように、ダミーノードで示された指示内容は、議論を結論に収束させる上で重要である。表 1 で述べた参加者の特性を考慮すると、的確な指示だったようだ。一方、指示 3 の自己紹介、指示 7 の議題に集中といった、メールの subject を立てて組織で議論する発展性に乏しい指示の特定は難しい。

### 5.3.2 黒幕からの被指示者の一致性

被指示者について、黒幕が記録した被指示者が、表 2 のメール送受信者に含まれる比率で評価した。被指示者の数が多い指示 4 では比率が低い、他の多くの指示では 100% である。表 4 に、被指示者への指示の回数ごとに、被指示者がメール送受信者に含まれる比率を示す。

上位 4 位の被指示者  $PID_4$ ,  $PID_{12}$ ,  $PID_{13}$ ,  $PID_{16}$  は、メール送受信者に含まれる。5 位の  $PID_7$  が含まれないのは、指示にもかかわらず、メール送信回数が 3 回と少なかったためだろう。指示の回数が 1 回の被験者を特定するのは、難しい。表 2 に現れる回数の多い、 $PID_1$ ,  $PID_4$ ,  $PID_9$ ,  $PID_{11}$ ,  $PID_{12}$ ,  $PID_{13}$  などは、表 4 で複数の指示の回数を記録している。これらの人物は、黒幕と組織とのゲートウェイの役割を果たしており、黒幕からクラスタへの経路として注目すべき、との結果が得られた。黒幕を突き止めるには、これらの人物を監視し、より詳細な観測データを収集すべきことを示唆している。

組織の意思決定を目的とした電子メールによるコミュニケーションを観測データとする、組織の見えない黒幕を探る実験において、有効な仮説を創出できた。事前知識によると独立性の高そうなクラスタ間に、ク

ラスタを結びつけるダミーノードが出現した。このヒントを解釈し、事前理解と観測データへの認知とのずれから気づきを獲得し、事前知識には欠けていた仮説を創出した。ヒューマンインタラクティブアニーリングが、組織のコミュニケーションと意思決定を誘導する、見えない黒幕を探るための有効な情報を与えることを示した。

## 5.4 まとめ

組織の見えない黒幕を探る問題に取り組んだ。組織の意思決定を目的とした電子メールによるコミュニケーションを観測データとする、組織の見えない黒幕を探る実験において、有効な仮説を創出できた。事前知識によると独立性の高そうなクラスタ間に、クラスタを結びつけるダミーノードが出現した。このヒントを解釈し、事前理解と観測データへの認知とのずれから気づきを獲得し、事前知識には欠けていた仮説を創出した。ヒューマンインタラクティブアニーリングが、組織のコミュニケーションと意思決定を誘導する、見えない黒幕を探るための有効な情報を与えることを示した。

今後の課題についても述べる。見えない重要人物が存在しない組織との対比が必要だろう。実験が示すように、観測データの中から、コミュニケーションや組織の機能に重要な見えない事象が強く関連する部分を取り出すことに成功した。しかし、それが、人物としての黒幕の存在を保証するとは限らない。例えば、グループの境界を越えて共有された危機意識や、意思決定にバイアスをかけるメディアや流行かもしれない。一方、テロリスト対策では、不完全な仮説に基づく過剰な警報が発信されても、人命に係わる重要事象を見過ごすよりもよいだろう。不完全な仮説の創出の抑制よりも、新たな観測や多様な人間の知見を取り込みやすくして、迅速な仮説の検証を促進する方が、有益だと考えられる。

## 6 ケーススタディに基づく共謀者の探知

見えない重要人物を探る問題の 1 例として, 9/11 同時多発テロの組織において実行犯を支援した共謀者を探る問題を取り上げる. 公開されている 9/11 テロ組織の情報をもとに, 共謀者の特徴を分析した上で, 組織行動をシミュレーションし, 3 章の手法を適用する.

### 6.1 9/11 テロ組織のネットワーク構造の分析

[Krebs 2002], [Morselli 2007] は, 9/11 テロにおいて, 4 機の旅客機 (AA11, AA77, AA175, UA93) に登場した 19 人のハイジャック実行犯と 18 人の共謀者の社会ネットワークを研究した. ネットワークを図 23 に示す.

ここで, いくつかの指標を定義する. Peripherality  $e(i)$  と partiality  $a(i)$  は, ノードのコミュニケーション効率への寄与を表す. Peripherality は, ある人物  $i$  から他の人物  $j$  への最短距離  $l_{ij}$  の平均値である. 距離をホップ数で計測し, 式 (30) で定義する. Closeness centrality [Freeman 1979] に近い. 小さな peripherality は, 他の人物に近いことを意味する. Partiality は, ある人物  $i$  から他の人物  $j$  への最短距離  $l_{ij}$  の標準偏差である. 式 (31) で定義する. 小さな partiality は, バイアスの少ない中立的な位置を意味する. 他方, nodal degree  $d(i)$  と clustering coefficient  $c(i)$  [Watts 1998] は, 露出の度合い (目に留まり易さ) を表す. なお, 本研究では, 1 本のリンクのみ持つノードの clustering coefficient は 1 と定義する. 一部の文献では, 0 と定義している. 大きな nodal degree は, 観測し易い強い結びつきや活性を示す. 大きな clustering coefficient は, 調査者の関心を弾きやすいクレークの中心的な位置をにあることを意味する.

$$e(i) \equiv \frac{1}{n-1} \sum_{j \neq i} l_{ij}. \quad (30)$$

$$a(i)^2 \equiv \frac{1}{n-1} \sum_{j \neq i} (l_{ij} - e(i))^2. \quad (31)$$

9/11 テロネットワークは, スケールフリーネットワークである Barabási-Albert モデル [Barabási 1999a] と比較してみる. は大きく異なる. Barabási-Albert モデルのパラメータは, peripherarity, partiality, nodal degree の Gini 係数がほぼ同じになるよう調整した. 上記の 4 つの指標を表 5 に示す. 9/11 テロネットワークは, nodal degree が 1.2 倍大きく, clustering

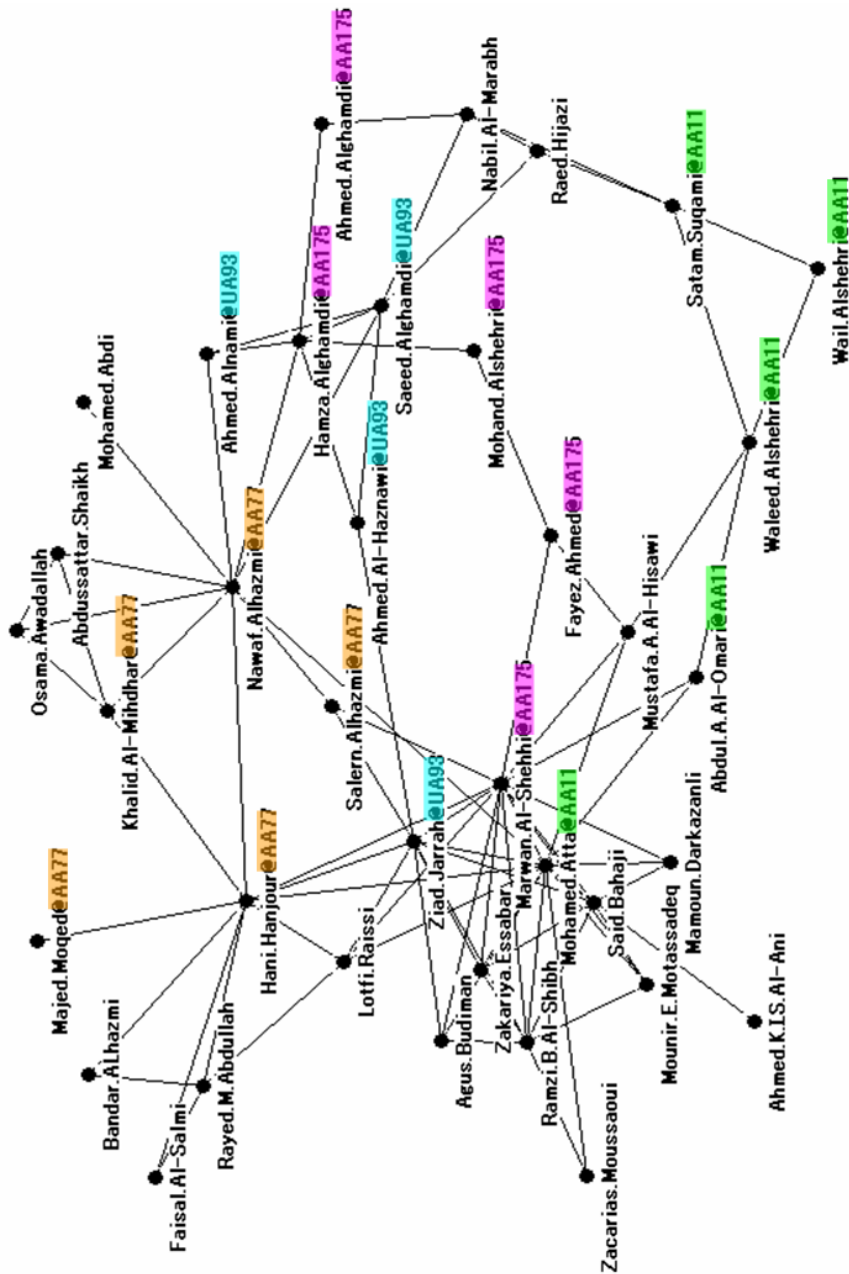


図 23: 9/11 テロにおける 19 人のハイジャック実行犯と 18 人の共謀者の社会ネットワーク [Krebs 2002] . ハイジャック実行犯には , ハイジャックした 4 機の旅客機 ( AA11, AA77, AA175, UA93 ) に対応する 4 つのグループがある .

	9/11 terrorist network	Barabási-Albert model
Nodal degree $\mu(d)$	4.6	3.7
Clustering Coeff. $\mu(c)$	0.60	0.19
Peripherality $\mu(e)$	2.6	2.6
Partiality $\mu(a)$	0.96	0.91
Gini Coeff.	0.33	0.33

表 5: テロリストネットワークと Barabási-Albert モデルとの比較．平均  $\mu(d)$  などは，すべての人物に関する値である．

coefficient が 3.2 倍大きい．これは，9/11 テロネットワークが，ハブノードを設けずに，コミュニケーションの効率が同等となるような強さの接続性を確保しようとするためである．ハブノードがないことは，Barabási-Albert モデルの欠点を補う．欠点とは，ハブノードへの攻撃に弱いこと [Albert 2000]，ネットワークでの効率的な検索を可能にしてしまうこと [Adamic 2001] である．

9/11 テロネットワークは，コンパクトな構造ではない．大きな clustering coefficient は，多数のクラスタ群の存在を示唆する．[Klerks 2002] が定性的に議論したように，9/11 テロネットワークは，center-and-periphery 構造ではなく，cluster-and-bridge 構造のようだ．Bridge は，ひとりの人物の場合もあれば，グループの場合もある．Bridge は，クラスタをつなぎあわせネットワークを構成するための重要な部品となっている．9/11 テロネットワークは小さなネットワークであるため，統計的な不確実性が大きいですが，ネットワークの構造と成長についての有用な示唆を与える．

## 6.2 人物のグローバル・ローカルな重要性指標の導入

コミュニケーションの効率と露出の度合いをネットワーク中のグローバルな重要度，ローカルな重要度という指標で最定義する．重要度は，0 から 1 までの値に規格化され，相互に比較可能である．

グローバルな重要度は，peripherality が partiality に基づく，トポロジから見たグローバルな重要性を表現する．例えば，ネットワーク全体に速やかに同報的に情報をブロードキャスト（マルチキャスト）配信できるノードは，グローバルに重要である．グローバルな重要性は，ローカルに観測可能な量から導くことはできない．式 (32) で定義する．平均値

$\mu$  と標準偏差  $\sigma$  を用いて，規格化され，無次元の相対的な重みに変換する．次に，重みをシグモイド関数  $s(x) \in [0, 1]$  で変換する（式 (33)）．例えば， $s(\pm 1) = 0.5 \pm 0.23$ ， $s(\pm 2) = 0.5 \pm 0.38$  である．小さな peripherality，小さな partiality を持つノードは，より重要で，1 に近い値が割り当てられる．

$$R_{\text{gl}}(i) = R_e(i) \equiv s\left(\frac{\mu(e) - e(i)}{\sigma(e)}\right), \text{ or } R_a(i) \equiv s\left(\frac{\mu(a) - a(i)}{\sigma(a)}\right). \quad (32)$$

$$s(x) \equiv \frac{1}{1 + e^{-x}} = \frac{1}{2}(1 + \tanh(\frac{x}{2})), \quad (0 \leq s(x) \leq 1). \quad (33)$$

ローカルな重要性は，nodal degree か clustering coefficient に基づく，ローカルな重要性を表現する．例えば，コミュニケーションルータとして多数の隣接ノードを結びつけるノードは，ローカルに重要である．式 (34) で定義する．大きな degree，大きな clustering coefficient を持つノードは，より重要で，1 に近い値が割り当てられる．ローカルな重要なノードは，観測しやすく，露出や検出に対するセキュリティという点では，ウィークポイントになるだろう．最後に，重要性比率 (global-to-local relevance ratio) を式 (35) で定義する．

$$R_{\text{lo}}(i) = R_d(i) \equiv s\left(\frac{d(i) - \mu(d)}{\sigma(d)}\right), \text{ or } R_c(i) \equiv s\left(\frac{c(i) - \mu(c)}{\sigma(c)}\right). \quad (34)$$

$$R(i) \equiv \frac{R_{\text{gl}}(i)}{R_{\text{lo}}(i)}. \quad (35)$$

### 6.3 重要性指標による共謀者の分類

重要性比率をすべてのテロリストについて計算した．結果を表 6 にまとめる． $R_e/R_c$  と  $R_a/R_c$  によって，実行犯と共謀者が明確に分離できることがわかる．実行犯は，大きな比率を持っている． $R_e/R_c$  の順番で，上位 19 人に 15 人 (79%) の実行犯が含まれる． $R_a/R_c$  では，16 人 (84%) が含まれる．[Morselli 2007] によると，betweenness centrality と closeness centralities では，14 人 (74%)，13 人 (68%) が含まれる．Nawaf Alhazmi, Mohamed Atta, Hani Hanjour が上位 3 位にくる．実行犯は，大きなコ



コミュニケーション効率と小さな露出によって、特徴付けることができそう  
うだ。他方、 $R_c/R_d$ と $R_a/R_d$ からは、有用な情報は得られない。

次に、共謀者の方に目を転じる。注意深く表を見ると、2人の共謀者  
が、実行犯と同程度の重要性比率を示している。Mustafa A. Al-Hisawi  
は、 $R_e/R_c$ で8位、 $R_a/R_c$ で7位である。Lotfi Raissiは、 $R_e/R_c$ で17  
位、 $R_a/R_c$ で20位である。他の共謀者とは、性質が異なるようだ。異  
なる機能を提供する2種類の共謀者があるようだ。表7は、2種類の共謀  
者と実行犯の特性をまとめた。タイプ1の共謀者は、もっと大きな $R_e$ と  
 $R_a$ を持つ。タイプ1は、実行犯グループ同士を結びつける役割を担っ  
ているのだろう。他方、タイプ2の共謀者は、最も小さい $R_e$ と $R_a$ 、大き  
な $R_c$ を持っている。タイプ2は、少数のグループと外部世界を結びつけ  
る役割を担っているのだろう。

これまでの考察をまとめ、9/11テロ組織と共謀者の特徴を述べる。Cluster-  
and-bridge構造を基礎とし、クラスタ間の隙間をbridgeとなる共謀者が  
埋めるような形で、ネットワークが成長する。クラスタは、相対的に壊  
れにくく、変化しにくい構造であろう。一方、共謀者は流動的で、環境へ  
の適応に対して重要な役割を担っているだろう。タイプ1の共謀者は、文  
字通り、bridgeの役割を行う。タイプ2の共謀者は、外部世界とのbridge  
の役割を担うことになる。図24は、これらの考察を模式的に示した、  
ネットワークの成長の様子である。クラスタ間の構造を組成・解体しな  
がら進化するネットワークである。[Popp 2006]で、Al Qaedaはクリーク  
が柔軟に相互接続したものである、との分析と一致している。

## 6.4 隠れた共謀者を探るシミュレーション

9/11テロネットワークをもとに、テストデータを生成し、見えない構  
造としての隠れた共謀者についてのシミュレーション実験を行う。

図23のネットワークをもとに、コミュニケーション履歴に相当するテ  
ストデータを生成する。コミュニケーションバスケットは、以下の2ス  
テップで生成する。2ステップ目で、共謀者を履歴から消去すること  
によって、見えない構造を織り込んだテストデータとする。見えない構造  
を織り込む作業は、ネットワーク中のコミュニケーションパターンを変  
えることはなく、アルゴリズムの入力となる観測データに相当するテ  
ストデータのみを変更する。

ステップ1では、コミュニケーション履歴を作成する。コミュニケー

Name	Local		Global		Ratio (ranking)	
	$R_d$	$R_c$	$R_e$	$R_a$	$R_e/R_c$	$R_a/R_c$
Nawaf.Alhazmi@AA77	0.83	0.23	0.86	0.86	3.74 (1)	3.74 (1)
Mohamed.Atta@AA11	0.96	0.27	0.89	0.82	3.26 (2)	3.00 (3)
Hani.Hanjour@AA77	0.83	0.27	0.83	0.69	3.11 (3)	2.60 (4)
Marwan.Al-Shehhi@AA175	0.94	0.30	0.82	0.36	2.76 (4)	1.23 (15)
Hamza.Alghamdi@AA175	0.60	0.28	0.62	0.58	2.21 (5)	2.08 (10)
Mohand.Alshehri@AA175	0.32	0.15	0.32	0.56	2.11 (6)	3.66 (2)
Ziad.Jarrah@UA93	0.83	0.38	0.78	0.56	2.08 (7)	1.50 (13)
Mustafa.A.Al-Hisawi	0.46	0.32	0.66	0.78	2.08 (8)	2.46 (7)
Salern.Alhazmi@AA77	0.38	0.32	0.66	0.78	2.08 (9)	2.46 (6)
Saeed.Alghamdi@UA93	0.60	0.32	0.65	0.72	2.03 (10)	2.26 (9)
Abdul.A.Al-Omari@AA11	0.38	0.32	0.63	0.81	1.98 (11)	2.54 (5)
Ahmed.Al-Haznawi@UA93	0.38	0.32	0.58	0.74	1.83 (12)	2.32 (8)
Fayez.Ahmed@AA175	0.38	0.32	0.52	0.55	1.62 (13)	1.74 (12)
Ahmed.Alghamdi@AA175	0.32	0.15	0.22	0.31	1.41 (14)	2.03 (11)
Waleed.Alshehri@AA11	0.46	0.22	0.31	0.30	1.36 (15)	1.34 (14)
Ramzi.B.Al-Shibh	0.73	0.51	0.62	0.30	1.22 (16)	0.59 (30)
Lotfi.Raissi	0.53	0.57	0.65	0.48	1.13 (17)	0.84 (20)
Said.Bahaji	0.67	0.58	0.60	0.34	1.03 (18)	0.59 (31)
Khalid.Al-Mihdhar@AA77	0.46	0.55	0.55	0.49	1.00 (19)	0.90 (18)
Nabil.Al-Marabh	0.46	0.32	0.27	0.19	0.83 (20)	0.60 (28)
Rayed.M.Abdullah	0.46	0.43	0.35	0.26	0.82 (21)	0.60 (29)
Zakariya.Essabar	0.53	0.76	0.57	0.44	0.75 (22)	0.58 (32)
Agus.Budiman	0.46	0.76	0.55	0.49	0.73 (23)	0.65 (25)
Ahmed.Alnami@UA93	0.38	0.76	0.53	0.73	0.70 (24)	0.96 (16)
Mounir.E.Motassadeq	0.46	0.76	0.53	0.49	0.70 (25)	0.65 (26)
Mamoun.Darkazanli	0.38	0.76	0.52	0.55	0.68 (26)	0.73 (22)
Zacarias.Moussaoui	0.32	0.76	0.45	0.57	0.59 (27)	0.75 (21)
Ahmed.K.I.S.Al-Ani	0.26	0.76	0.43	0.64	0.57 (28)	0.84 (19)
Abdussattar.Shaikh	0.38	0.76	0.40	0.54	0.53 (29)	0.71 (24)
Osama.Awadallah	0.38	0.76	0.40	0.54	0.53 (30)	0.71 (23)
Mohamed.Abdi	0.26	0.76	0.37	0.69	0.48 (31)	0.90 (17)
Raed.Hijazi	0.38	0.55	0.25	0.24	0.46 (32)	0.45 (35)
Satam.Suqami@AA11	0.46	0.32	0.14	0.06	0.43 (33)	0.19 (36)
Bandar.ALhazmi	0.32	0.76	0.32	0.38	0.42 (34)	0.50 (34)
Faisal.Al-Salmi	0.32	0.76	0.32	0.38	0.42 (35)	0.50 (33)
Majed.Moqed@AA77	0.26	0.76	0.31	0.46	0.40 (36)	0.61 (27)
Wail.Alshehri@AA11	0.32	0.76	0.07	0.07	0.09 (37)	0.10 (37)

表 6: すべてのテロリストのグローバル・ローカルな重要性・重要性比率  $R = R_e/R_c$  の大きさによって並べた。

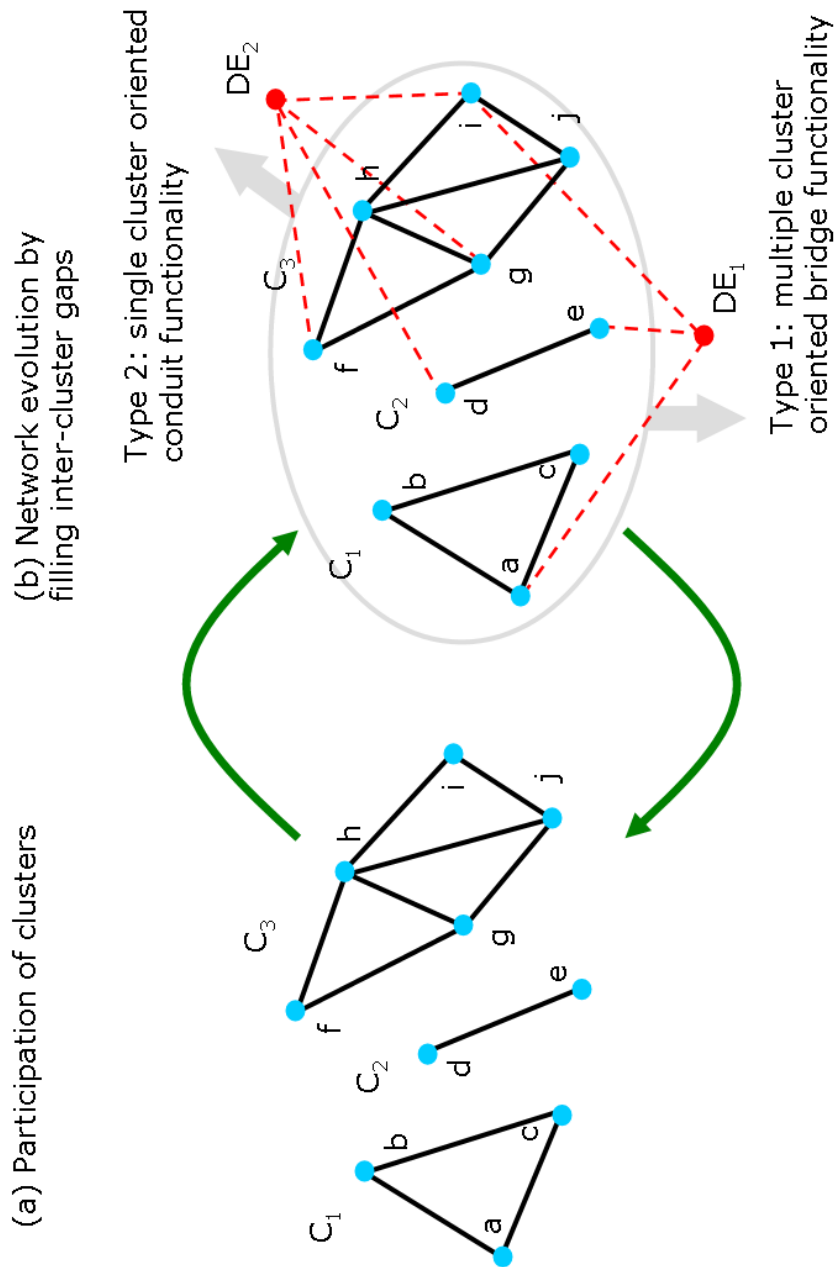


図 24: クラスタ間の構造を組成・解体しながら進化するネットワーク。クラスタ間の構造には、2種類ある。タイプ1の共謀者は、文字通り、bridgeの役割を行う。タイプ2の共謀者は、外部世界とのbridgeの役割を担う。

Type	Local		Global		Ratio	
	$\mu(R_d)$	$\mu(R_c)$	$\mu(R_e)$	$\mu(R_a)$	$\mu(R_e/R_c)$	$\mu(R_a/R_c)$
Hijacker	0.53	0.37	0.54	0.55	1.8	1.86
Type 1 conspirator	0.49	0.44	0.65	0.63	1.61	1.65
Type 2 conspirator	0.42	0.67	0.43	0.44	0.67	0.65

表 7: テロリストの3つのカテゴリの比較．平均  $\mu(R_d)$  などは，カテゴリごとに算出した．

シヨンは，発話者から一定の距離まで伝播すると想定する．発話者と発話者に応答した参加者によるひとつの会話に対応する．コミュニケーションは，確率  $t$  でリンクを伝播する．これは，コミュニケーションの強さとも言える．1 ホップで，平均  $t \times \mu(d)$  人に到達する．1 ホップは，リンク1本の距離である．ここでは，最大伝播距離を2 ホップに制限した．ネットワークが小さいため，3 ホップ以上のコミュニケーションは，ほとんどのノードに到達してしまうためである．発話者はランダムに選ぶ．コミュニケーションが伝播した人物の名前を要素としたバスケットを作成する． $t = 0.4, 0.6, 0.8, 1.0$  において，バスケットに含まれる平均の人数は6.5, 10.1, 13.7, 17.1 である．バスケットの数は， $|b| = 10|e| = 370$  とした．下記の例は，Abdul A. Al-Omari, Mustafa A. Al-Hisawi, Waleed Alshehri, Fayez Ahmed を発話者とするコミュニケーションである．

- $b_1 = \{\text{Abdul A. Al-Omari, Marwan Al-Shehhi, Mohamed Atta, Waleed Alshehri}\}.$
- $b_2 = \{\text{Mustafa A. Al-Hisawi, Marwan Al-Shehhi, Mohamed Atta, Fayez Ahmed, Waleed Alshehri}\}.$
- $b_3 = \{\text{Waleed Alshehri, Abdul A. Al-Omari, Mustafa A. Al-Hisawi, Wail Alshehri, Satam Suqami}\}.$
- $b_4 = \{\text{Fayez Ahmed, Mohand Alshehri, Hamza Alghamdi}\}.$

ステップ2では，見えない構造として隠れた共謀者をテストデータに織り込む．ステップ1のテストデータから，設定する隠れた共謀者を消去する．その結果，消去された共謀者と共謀者のリンクについての情報がテストデータから失われ，見えない構造ができあがる．これが，アルゴリズムに入力となる．このテストデータは，いわば，隠れた共謀者が

らの口頭のコミュニケーションが欠落した，eメールの監視履歴に相当する．下記の例は，Mustafa.A.Al-Hisawi が隠れた共謀者となって，テストデータである．アルゴリズムは，見えない構造に関するヒントを含むバスケットとして， $b'_2$  と  $b'_3$  を取り出すことが期待される．これらのバスケットを取り出すことができれば，両方に含まれる Waleed.Alshehri の周囲を調べることで，隠れた Mustafa.A.Al-Hisawi についてのヒントを見つけられるだろう．

- $b_1 = \{\text{Abdul A. Al-Omari, Marwan Al-Shehhi, Mohamed Atta, Waleed Alshehri}\}.$
- $b_2' = \{\text{Marwan Al-Shehhi, Mohamed Atta, Fayez Ahmed, Waleed Alshehri}\}.$
- $b_3' = \{\text{Waleed Alshehri, Abdul A. Al-Omari, Wail Alshehri, Satam Suqami}\}.$
- $b_4 = \{\text{Fayez Ahmed, Mohand Alshehri, Hamza Alghamdi}\}.$

バスケットについてのランキング関数には，3.3 節と同様の計算式を用いる．ダミー事象とクラスタとの間の Jaccard 係数の平均（式 (36)），上位 2 つの合計（式 (37)），標準偏差（式 (38)）を用いる．関数  $T(x_j, l)$  は， $x_j$  を値の大きな順に並べて， $l$  番目の要素を取り出す． $I_{tp}$  は，2 つのクラスタ間の結びつきに注目する． $I_{av}$  と  $I_{tp}$  は，大きな peripherality を持つ人物を探す． $I_{sd}$  は，大きな partiality を持つ人物を探す．大きな  $I_{av}(i)$ ， $I_{tp}$ ，小さな  $I_{sd}(i)$  をより上位にランク付けし，取り出す．ランクの上位のものから，見えない構造を含むバスケットとして，取り出される．

$$I_{av}(i) \equiv \frac{1}{|c|} \sum_{1 \leq j \leq |c|} \max_{h_k \in C_j, b_i} J(\text{DE}_i, h_k). \quad (36)$$

$$I_{tp}(i) \equiv \frac{1}{2} (T(\max_{h_k \in C_j, b_i} J(\text{DE}_i, h_k), 1) + T(\max_{h_k \in C_j, b_i} J(\text{DE}_i, h_k), 2)). \quad (37)$$

$$I_{sd}(i)^2 \equiv \frac{1}{|c|} \sum_{1 \leq j \leq |c|} (\max_{h_k \in C_j, b_i} J(\text{DE}_i, h_k) - I_{av}(i))^2. \quad (38)$$

#### 6.4.1 ランキング関数に対する依存性

プレジジョン  $p$  , リコール  $r$  に加えて , F value を用いて , アルゴリズムの性能を評価する . F value は , プレジジョンとリコールの幾何平均で , 式 (39) で定義する .

$$F \equiv \frac{1}{\frac{1}{2}(\frac{1}{p} + \frac{1}{r})} = \frac{2pr}{p+r} \quad (39)$$

さまざまな条件下で , アルゴリズムの性能を評価した . 図 25 は ,  $|c| = 4$  ,  $p = 0.8$  で , Mustafa A. Al-Hisawi が隠れた共謀者の場合のプレジジョンとリコールとを示す . 横軸は , 全バスケットの数に対する取り出したバスケットの数の比率である . 上位 10% にランク付けされたバスケットは , 隠れた共謀者が削除されていたバスケットを正しく取り出している . ランキング関数  $I_{sd}$  は ,  $I_{av}$  ,  $I_{sd}$  よりも少し良い性能を示す . 以後は ,  $I_{sd}$  を使用して評価を進める . 全バスケットを取り出した際のプレジジョンは 45% である . この問題は , 多くの正解を含み , 易しい問題である . これは , ネットワークが小さいためである .

#### 6.4.2 ネットワーク構造の事前知識に対する依存性

図 26 は ,  $|c| = 2, 4, 8$  ,  $p = 0.8$  におけるプレジジョンとリコールを示す .  $|c|$  の値は , ネットワークの構造についての事前知識に依存する . 4 機の旅客機がハイジャックされたことから ,  $|c| = 4$  は自然な選択である .  $|c| = 4$  が最も良い性能を示している . 誤った事前知識 ( $|c| = 2$ ) では , 性能は劣化する .  $|c| = 8$  での性能の劣化は小さい . これは , 実際のグループ数が , 4 に近い少し大きな値とするのが妥当なためであろう .

#### 6.4.3 コミュニケーションの伝播距離に対する依存性

図 27 は ,  $|c| = 4$  ,  $t = 1.0, 0.8, 0.6, 0.4$  での F value ゲインを示す . ゲインは , ランダムな取り出しと比較した場合の F value の比率である .  $t = 1.0, 0.8$  では , 性能は安定しており , カーブはスムーズである .  $t = 1.0$  では , 長距離のコミュニケーションが問題を簡単にするため , ゲインは小さい .  $t = 0.6$  では , 性能は不安定になりはじめ , カーブはゆらぎはじめる .  $t = 0.4$  では , アルゴリズムは機能しなくなる . これは , コミュニケーションの伝播距離が短く , 隠れた共謀者を媒介したクラスタ間の関係性を抽出できなくなるためである .

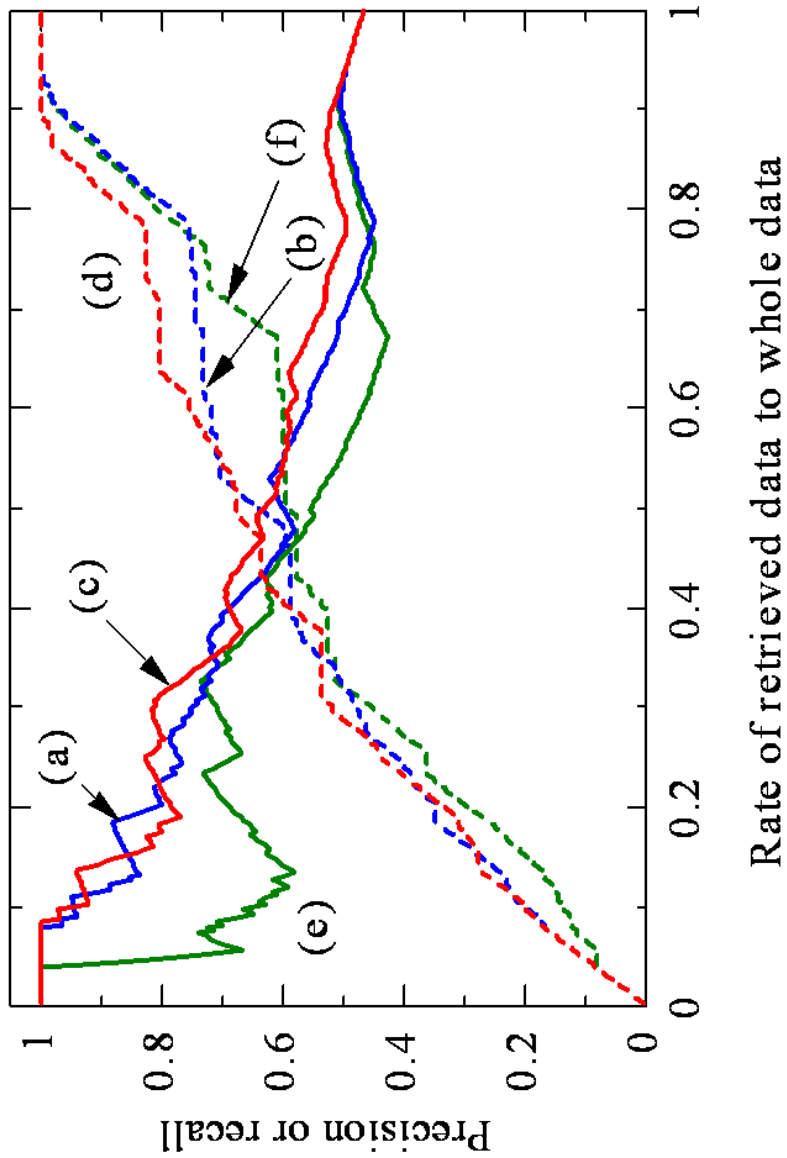


図 25: Mustafa A. Al-Hisawi が隠れた共謀者である場合に、取り出したバスケットに対するプレジジョン  $p$  , リコール  $r$  . (a)  $p$  using  $I_{av}$ , (b)  $r$  using  $I_{av}$ , (c)  $p$  using  $I_{sd}$ , (d)  $r$  using  $I_{sd}$ , (e)  $p$  using  $I_{tp}$ , and (f)  $r$  using  $I_{tp}$ .  $|c| = 4$  and  $t = 0.8$ . 横軸は、バスケット全体に対する取り出したバスケットの数の比率である .

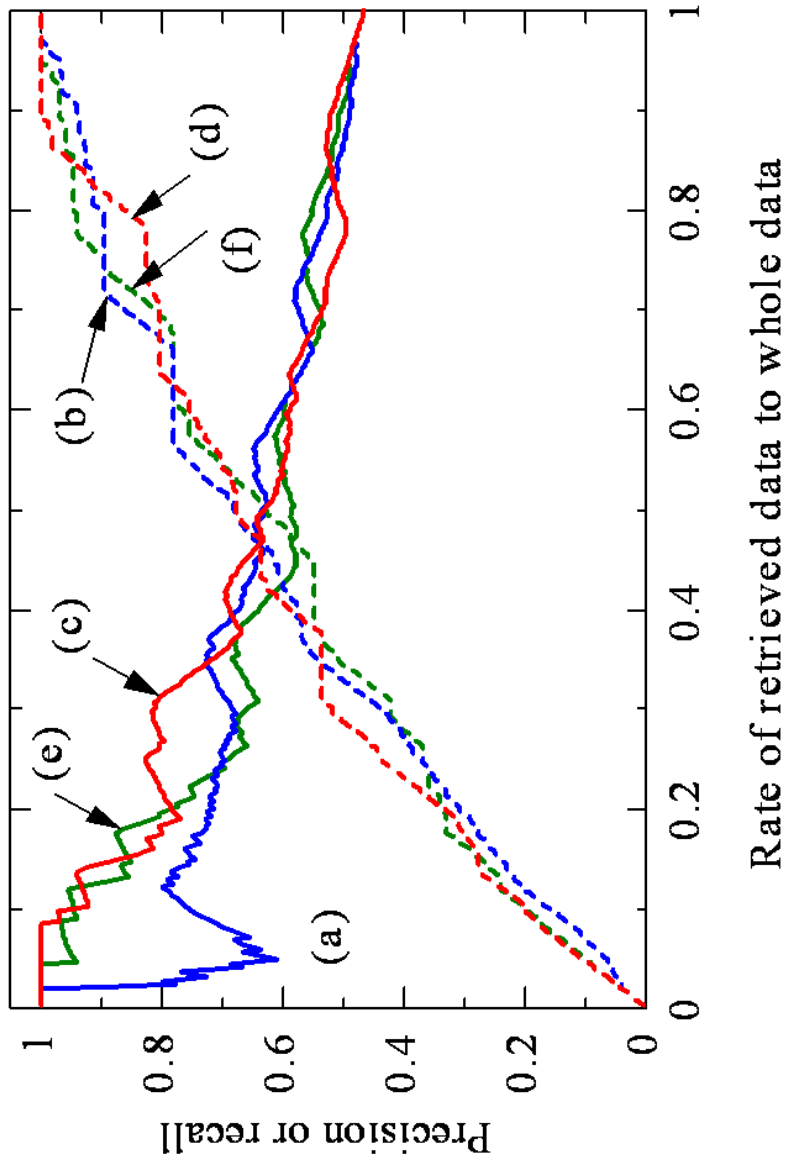


図 26: Mustafa A. Al-Hisawi が隠れた共謀者である場合に , 取り出したバスケットに対するプレジジョン  $p$  , リコール  $r$  . (a)  $p$  at  $|c| = 2$ , (b)  $r$  at  $|c| = 2$ , (c)  $p$  at  $|c| = 4$ , (d)  $r$  at  $|c| = 4$ , (e)  $p$  at  $|c| = 8$ , and (f)  $r$  at  $|c| = 8$ .  $t = 0.8$  and  $I_{sd}$  is used.



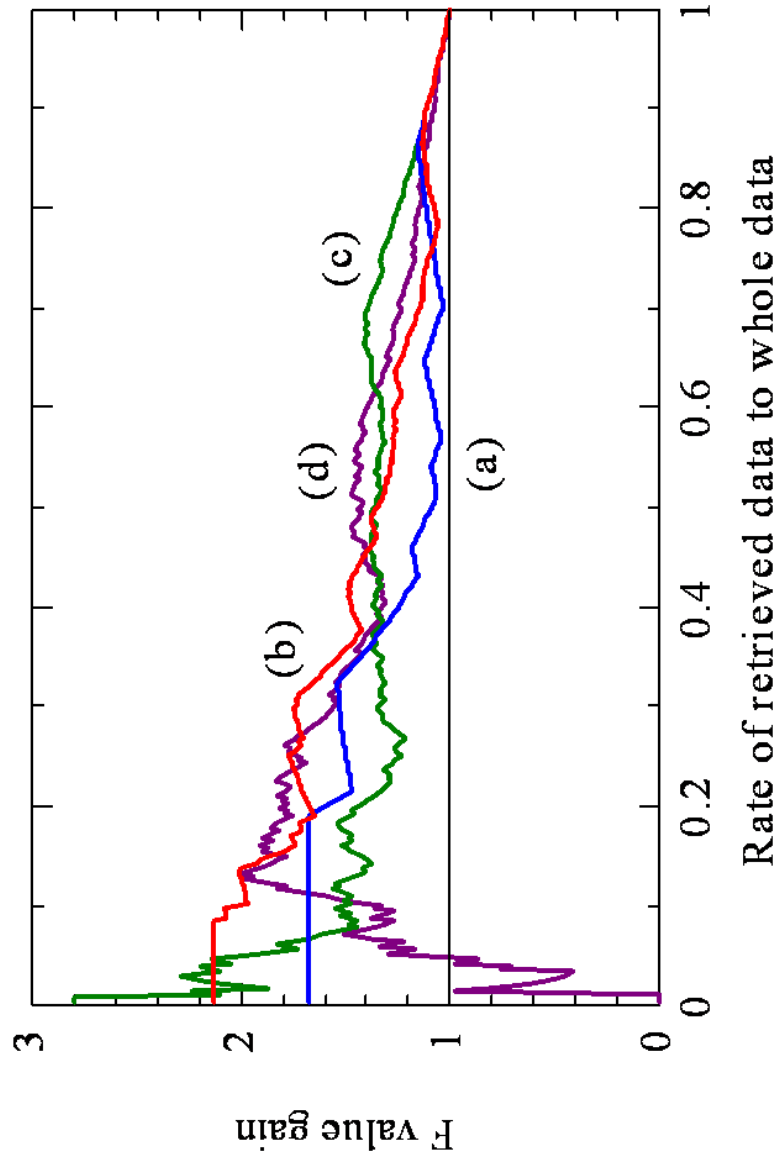


図 27: Mustafa A. Al-Hisawi が隠れた共謀者である場合に，取り出したバスケットに対する F value ゲイン . (a)  $t = 1.0$ , (b)  $t = 0.8$ , (c)  $t = 0.6$ , and (d)  $t = 0.4$ .  $|c| = 4$  and  $I_{sd}$  is used. ゲインは，ランダムな取り出しと比較した場合の F value の比率である .

#### 6.4.4 共謀者の分類に対する依存性

図 28 は、さまざまな隠れた共謀者に対する F value ゲインを示す。Lotfi Raissi に対しては、Mustafa A. Al-Hisawi と同様にアルゴリズムは機能する。Rayed M. Abdullah に対しても機能する。Rayed M. Abdullah の位置は、Lotfi Raissi の位置に近い。Ramzi B. Al-Shibh, Said Bahaji に対しても機能するが、Ramzi B. Al-Shibh に対しては劣化が起こる。Osama Awadallah, Raed Hijazi に対しては、性能が悪く不安定である。表 6 で、Osama Awadallah は、 $R_e/R_c$  で 30 番目、 $R_a/R_c$  で 23 番目である。Raed Hijazi は、 $R_e/R_c$  で 32 番目、 $R_a/R_c$  で 35 番目である。アルゴリズムは、タイプ 2 の共謀者に対しては、限界があることがわかる。図 29 は、Raed Hijazi に対して、 $I_{av}$  and  $I_{tp}$  を用いた場合の F value ゲインの大きさを示す。不安定ではあるが、 $I_{tp}$  の方がよい性能を示す。タイプ 2 の共謀者に対しては、2 つのクラスタに着目して、 $I_{tp}$  を用いることで、性能が向上する可能性がある。

### 6.5 仮説の創出に向けたシナリオマップ

提案するアルゴリズムが、コミュニケーション履歴から隠れた共謀者を発見するのに、有効に機能することを示した。特に、タイプ 1 の共謀者に有効である。ここで、発見の意味合いを検討する。一般的に、ノードの発見問題はリンクの発見問題よりも、あいまいな側面がある。仮に、1 章の式 (6) のような、厳密な数理的な解法によって、未知のノードを発見できたとしよう。そのノードが持つ意味を解釈できるだろうか？得られた知見は、ノードが何か新しいものというだけで、ノードの実体に関する情報は依然として欠けている。

明らかに、単なるデータの処理から、背景のコンテキストや知識を活用して発見の意味合いを解釈する段階に進む必要がある。ノードが表現する、ネットワークの見えない構造に起因する機会や脅威についての具体的な仮説に到達する必要がある。そのためには、数値的な結果だけでなく、結果を地図 [Shen 2006] の形に可視化して直感的洞察を刺激することは有用である。事前の理解や誤解を意識し、観測に対する認知、認知されている観測 [Krackhardt 1987] を見直す契機となる。これが、ヒューマンインタラクティブアニメーションの狙いである。

図 30 は、Mustafa A. Al-Hisawi が隠れた共謀者となる場合について、コミュニケーションから復元したネットワークである。図 25 と同じ条件で

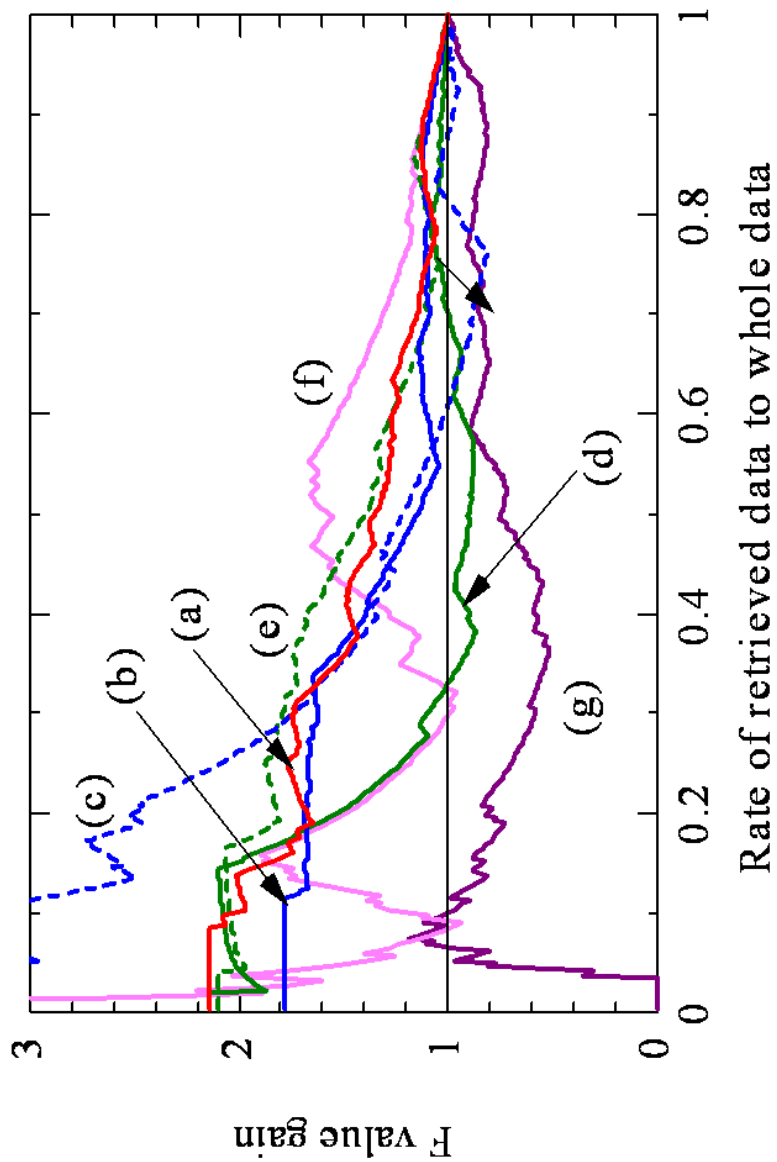


図 28: さまざまな隠れた共謀者に対して, 取り出したバスケットに対する F value ゲイン . 隠れた共謀者 : (a) Mustafa A. Al-Hisawi, (b) Lotfi Raissi, (c) Rayed M. Abdullah, (d) Ramzi B. Al-Shibh, (e) Said Bahaji, (f) Osama Awadallah, and (g) Raed Hijazi.  $|c| = 4$ ,  $t = 0.8$ , and  $I_{sd}$  is used.

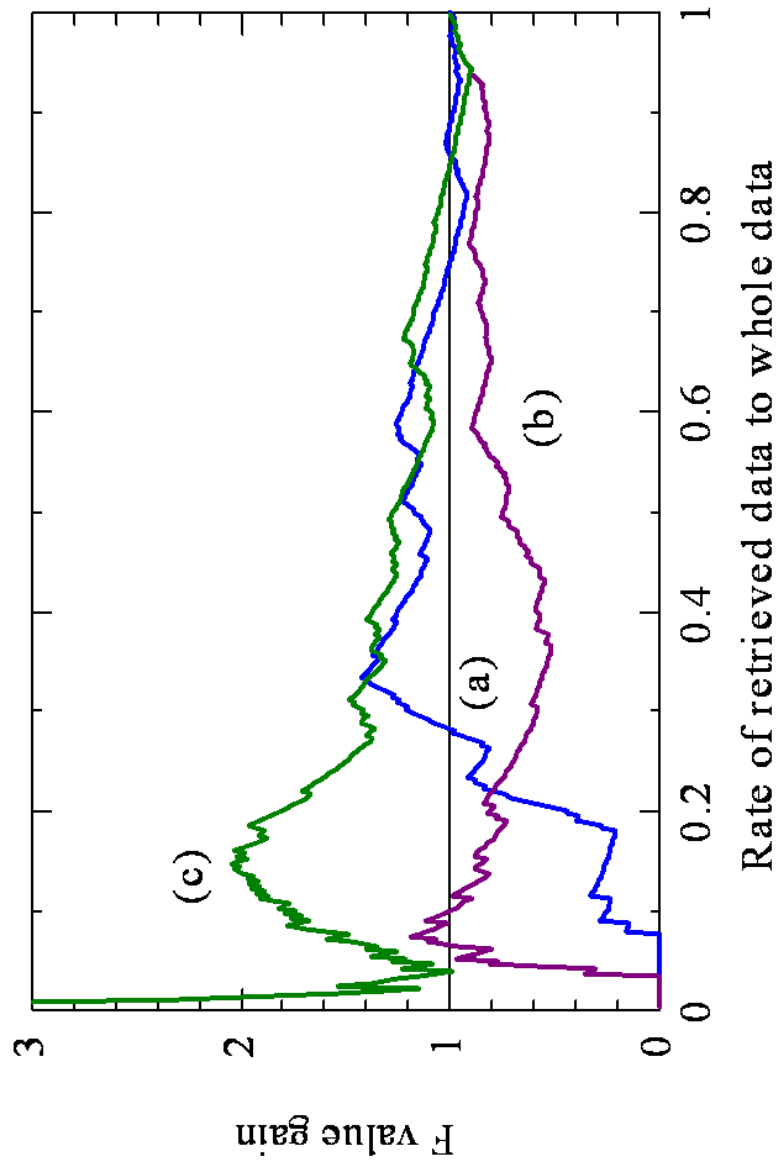


図 29: Raed Hijazi が隠れた共謀者である場合に , 取り出したバスケット に対する F value ゲイン . (a)  $I_{av}$ , (b)  $I_{sd}$ , and (c)  $I_{tp}$ .  $|c| = 4$  and  $t = 0.8$ .

ある。4つのクラスタが、ダミーノード  $DE_i$  で示された仮想的な Mustafa A. Al-Hisawi で、相互に接続されている。Jaccard 係数の性質から、頻繁に出現する人物を重要視しないことになる。これは、問題ではなく、良い性質である。しばしば、頻繁で目に留まりやすい事象に目を奪われ、予期しなかったヒントを見落とすことがある。

他方、頻度が低い事象は、予期しなかったヒントであることもあれば、ノイズであることもあるので、諸刃の刃でもある。ダミーノード  $DE_i$  で示された仮想的な Mustafa A. Al-Hisawi が接続しているノードの中で、Majed Moqed, Mohamed Abdi, Ahmed K. I. S. Al-Ani はノイズであろう。図 23 で、Mustafa A. Al-Hisawi から遠く離れている。Jaccard 係数で人物間の距離を計測しているため、そもそも出現頻度の少ない人物が強調される傾向があり、ノイズの原因となっている。一方、Waleed Alshehri と Mohand Alshehri がダミーノード  $DE_i$  で示された仮想的な Mustafa A. Al-Hisawi に接続されていることは、注目に値する。図 23 で、ダミーノード  $DE_i$  で示された仮想的な Mustafa A. Al-Hisawi と隣接しており、隠れた共謀者の調査を始める際の重要な出発点を示唆している。

## 6.6 まとめ

9/11 同時多発テロの組織において、実行犯を支援した共謀者についてのネットワークを分析した。トポロジから見たグローバル (peripherarity, partiality) ・ローカル (degree, clustering coefficient) な重要性指標を導入した。重要性比率によって、共謀者が 2 種類に分類できることを見出した。タイプ 1 の共謀者は、文字通り、クラスタ間の bridge の役割を行う。タイプ 2 の共謀者は、外部世界との bridge の役割を担う。

このネットワークに基づくシミュレーションを行い、プレジジョン、リコール、F value を評価した。特に、タイプ 1 の共謀者について、 $I_{sd}$  を用いたアルゴリズムが効果的に機能することを示した。一方、タイプ 2 の共謀者を探ることは難しく、 $I_{tp}$  で性能は向上するものの、不完全であることがわかった。シナリオマップの形式で出力することで、仮説にもとづき共謀者の調査を始めるきっかけを見出せることを議論した。

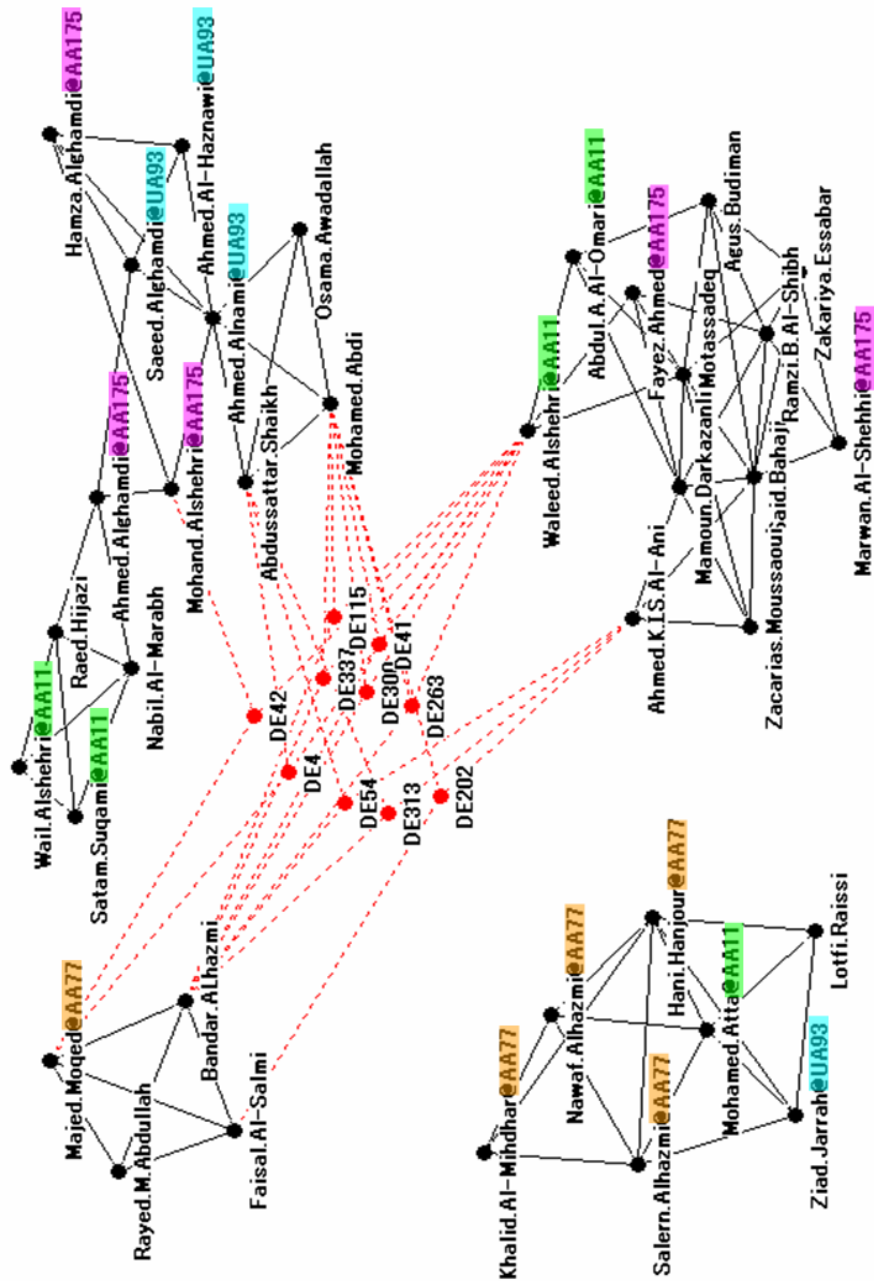


図 30: Mustafa A. Al-Hisawi が隠れた共謀者であるデータから復元したシナリオマップ. シナリオマップには, 4つのクラスターと10個のランキングの高い  $DE_i$  (ダミーノードで示された仮想的な Mustafa A. Al-Hisawi) とを含む.

## 7 グループ討議から読み解く個人の好み

ここでは、ヒューマンインタラクティブアニメーリングの応用展開として、組織的な意思決定とも関連する、グループ討議によって個人の理解を深める手法を開発する。アートの好みに対するアンケートを題材として、問題に特有の要因を考慮して拡張する。この問題では、二部グラフ構造を持ったアンケートの回答結果を取り扱う。このデータ構造に合わせて、アルゴリズムを改良する。さらに、グループ討議によって個人の理解を深める手法として、ヒューマンインタラクティブアニメーリングのプロセスの改良も施す。いずれも、コミュニケーションだけでなく、マーケティングや消費者分析に利用できる有用な応用である。

### 7.1 二部グラフ構造へのアルゴリズムの改良

ヒューマンインターラクティブアニメーリングを個人の行動・心理の理解を深めることに応用する。好みに関するアンケートの回答を基礎データとする。被験者 + {好むアイテム群} というデータのリストで表すことができる。または、アイテム + {アイテムを好む被験者群} というデータのリストで表すことができる。3.3節で述べた、バスケット形式の式(4)を拡張した、式(40)のようなサブジェクト付きバスケットの形式となる。サブジェクト  $s_i$  と事象群  $e_j$  から構成される。事象群は、サブジェクトのもとで、関連性の高い事象の一覧である。

$$b_i = s_i : \{e_j\} \quad (0 \leq i \leq |b| - 1, 0 \leq j \leq |b_i| - 1). \quad (40)$$

ひとりの被験者が1回しか回答しないとすると、行方向をアイテム、列方向を被験者とする行列において、好む=1、好まない=0を割り当てることで表現することができる。これは、二部グラフ (bipartite graph) の構造になる。

3.3節で述べた、単純バスケット形式の入力データに対するアルゴリズム(式(16)や式(18))を式(40)の入力データに対応できるよう、改良を施す。データが、アイテム + {アイテムを好む被験者群}を表すものとして、説明する。まず、3.3節と同じk-medoidsアルゴリズムを用いて、事象群  $e_{ij}$  を  $|c|$  個のグループにクラスタリングする。クラスタは、好みの観点から類似性のある被験者群である。好みの似た被験者をまとめたものと言える。クラスタリングには、3.3節と同じく、k-medoid法や階層化

クラスタリングを使用する．被験者の関係性を可視化したものという意味で，ヒューマンネットワーク，社会ネットワークの一種とも言える．

ひとりの被験者が，複数回しか回答する場合もあることを前提とする．1回しか回答しない場合には，計算は簡単になる．サブジェクト  $s_i$  が最も強く関連しているプライマリクラスタとプライマリクラスタのゲートウェイ事象を選ぶ．プライマリクラスタ  $c_P(i)$  を式 (41) で計算する．

$$c_P(i) = \arg_n \max_{e_k \in c_n} \frac{\sum_{s_i \in b_l} B(e_k \in b_l)}{\sum_l B(e_k \in b_l)} \quad (41)$$

サブジェクト  $s_i$  のプライマリクラスタ  $c_P(i)$  に対して，ゲートウェイ事象  $e_{\text{gtw}(P)}(i)$  を式 (42) で計算する．

$$e_{\text{gtw}(P)}(i) = \arg_k \max_{e_k \in c_P(i)} \frac{\sum_{s_i \in b_l} B(e_k \in b_l)}{\sum_l B(e_k \in b_l)} \quad (42)$$

セカンダリクラスタは，サブジェクト  $s_i$  が弱く関連しているクラスタである．セカンダリクラスタ  $c_S(i)$  を式 (43) で計算する．

$$c_S(i) = \arg_{c_n \neq c_P(i)} \max_{e_k \in c_n} \frac{\sum_{s_i \in b_l} B(e_k \in b_l)}{\sum_l B(e_k \in b_l)} \quad (43)$$

サブジェクト  $s_i$  のセカンダリクラスタ  $c_S(i)$  に対して，ゲートウェイ事象  $e_{\text{gtw}(S)}(i)$  を式 (44) で計算する．

$$e_{\text{gtw}(S)}(i) = \arg_k \max_{e_k \in c_S(i)} \frac{\sum_{s_i \in b_l} B(e_k \in b_l)}{\sum_l B(e_k \in b_l)} \quad (44)$$

セカンダリクラスタには，サブジェクト  $s_i$  が最も弱く関連しているクラスタを選ぶこともできる．式 (45) で計算する．

$$c_S(i) = \arg_{c_n \neq c_P(i)} \min_{e_k \in c_n} \frac{\sum_{s_i \in b_l} B(e_k \in b_l)}{\sum_l B(e_k \in b_l)} \quad (45)$$

シナリオマップへの可視化の際には，サブジェクト  $s_i$  に対応したダミーノード  $DE_i$  でプライマリクラスタとセカンダリクラスタのゲートウェイノード ( $e_{\text{gtw}(P)}(i)$ ,  $e_{\text{gtw}(S)}(i)$ ) を接続し，ダミーノード  $DE_i$  に，サブジェクト  $s_i$  を関連付ける．サブジェクトは，作品である．作品が，異なる好みの被験者クラスタをどのように結びつけるのが可視化したものができる．異なる好みの被験者クラスタの間のコミュニケーションを媒介するものと捉えることもできる．セカンダリクラスタに結びつきが最も弱いクラスタを用いると，意外性の高い結びつきが得られる可能性がある．これが，



シナリオマップとなる．上記の手順を模式的に，図 31 に示す．サブジェクト  $s_3$  に対して，プライマリクラスタ  $c_P(3) = c_1$ ，プライマリクラスタへのゲートウェイノード  $e_{\text{gtw}(P)}(3) = e_1$ ，セカンダリクラスタ  $c_S(3) = c_2$ ，セカンダリクラスタへのゲートウェイノード  $e_{\text{gtw}(S)}(3) = e_4$  となる．

上記は，入力データがアイテム：{ アイテムを好む被験者群 } の場合であったが，被験者：{ 好むアイテム群 } の場合についても同様にシナリオマップを作成する．入力データの形式は，式 (40) と同じになる．クラスタとの結びつきの計算には，式 (41)，式 (42)，式 (43)，式 (44) を流用する．シナリオマップに可視化する際には，図 31 とは異なり，クラスタにサブジェクトを関連付けた上で，ダミーノードを描画する手順とする．これを模式的に，図 32 に示す．サブジェクト  $s_3$  に対して，プライマリクラスタ  $c_P(3) = c_1$ ，プライマリクラスタへのゲートウェイノード  $e_{\text{gtw}(P)}(3) = e_1 (e_2, e_3)$ ，セカンダリクラスタ  $c_S(3) = c_2$ ，セカンダリクラスタへのゲートウェイノード  $e_{\text{gtw}(S)}(3) = e_4$  となる．まず，サブジェクト  $s_i$  をプライマリクラスタのゲートウェイノード  $e_{\text{gtw}(P)}(i)$  に接続する．被験者が最も好む作品クラスタに対応付けることに相当する．次に，サブジェクトに対応するダミーノード  $DE_i$  を通して，セカンダリクラスタのゲートウェイノード  $e_{S,\text{gtw}}(i)$  に接続する．ダミーノードが何らかのきっかけになって，被験者の関心が強い好みから弱い好みにスイッチする場合を図示したものである．

## 7.2 コミュニケーションの活性化シナリオの創出

被験者が，アンケートを可視化したシナリオマップを用いて議論することにより，行動の変容や予測に役立つシナリオを創出することを試みる．ここでは，好み異なるグループ間でのコミュニケーションの活性化についてのシナリオの創出を行う．好みに関するアンケートの題材として，美術作品を利用する．肖像画，風景画，抽象画などの 50 枚の作品を用意した．30 人の被験者に，好きだと感じる作品を任意数選んでもらった．これがアンケートの回答で，作品 ID (識別子): { 作品を好きだと感じる被験者 ID (識別子) } の形式となっている．

図 33 に，図 31 の形式で作成したシナリオマップを示す．8 つの被験者クラスタを図示して例である．ここで，3 つの場合について，ダミー事象を選択してシナリオマップに記載した．できるだけ多くの被験者クラスタを接続するダミー事象を選択し，関連付けられたサブジェクトであ

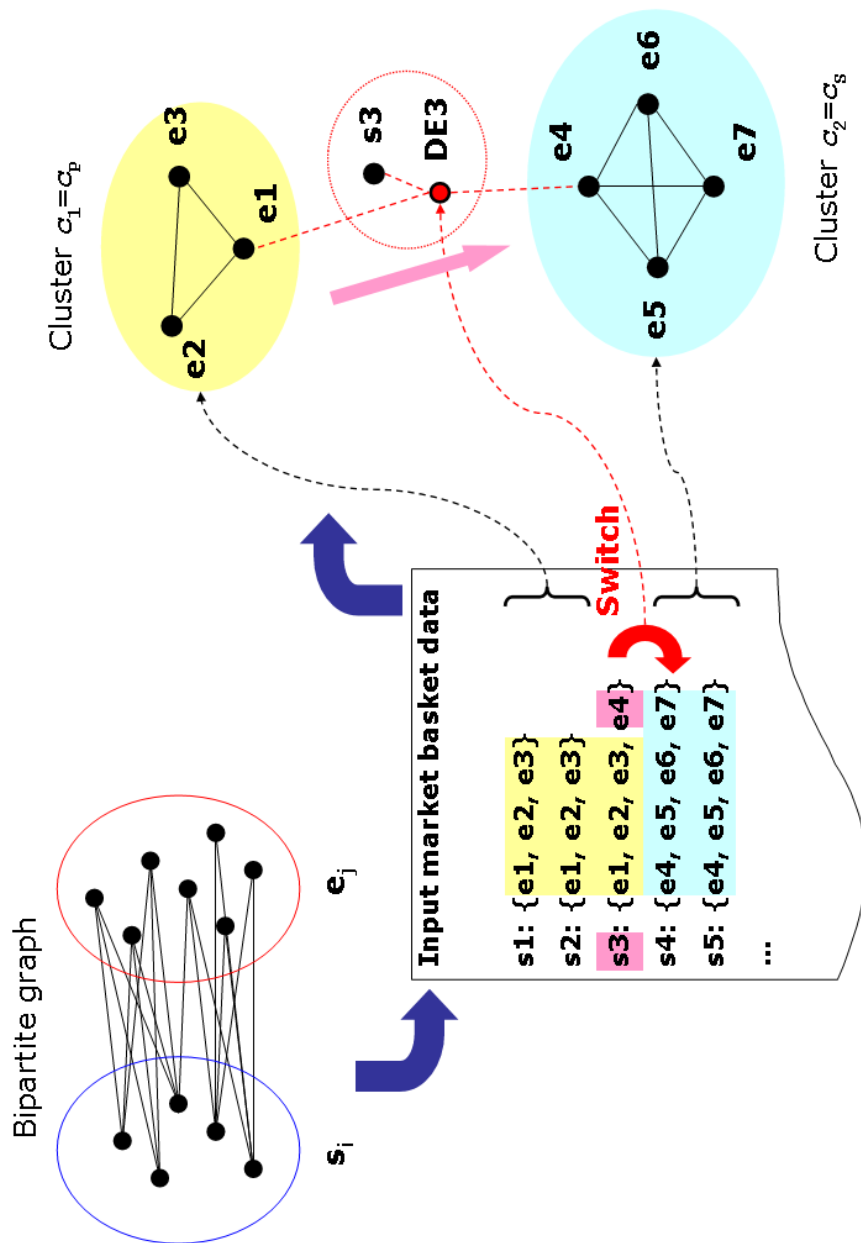


図 31: 二部グラフ構造 (Bipartite graph) のサブジェクト付きバスケットデータを入力とする, 人物クラスタを含むシナリオマップの生成. サブジェクト  $s_3$  に対して, プライマリクラスタ  $c_p(3) = c_1$ , , プライマリクラスタへのゲートウェイノード  $e_{gtw(P)}(3) = e_1$ , セカンダリクラスタ  $c_s(3) = c_2$ , , セカンダリクラスタへのゲートウェイノード  $e_{gtw(S)}(3) = e_4$  となる.

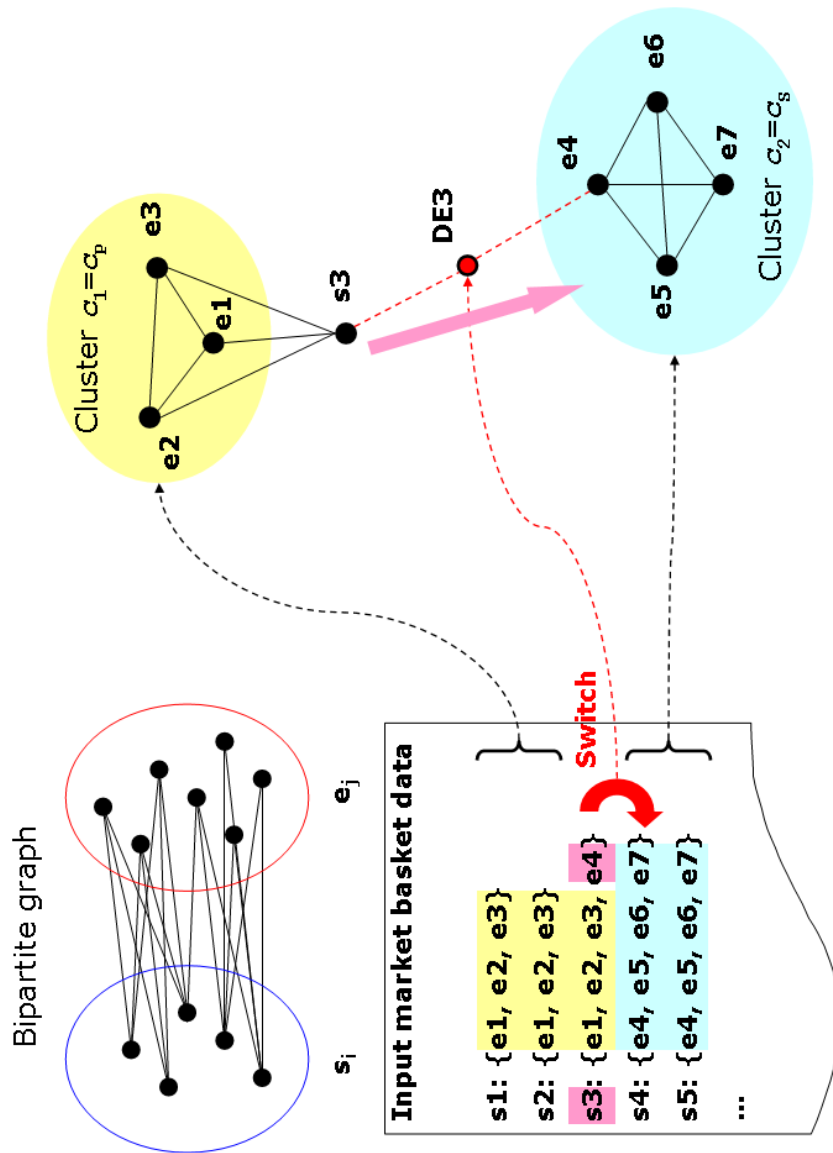


図 32: 二部グラフ構造 (Bipartite graph) のサブジェクト付きバスケットデータを入力とする, 作品クラスタを含むシナリオマップの生成. サブジェクト  $s_3$  に対して, プライマリクラスタ  $c_P(3) = c_1$ , , プライマリクラスタへのゲートウェイノード  $e_{\text{gtw}(P)}(3) = e_1$  ( $e_2, e_3$ ), セカンダリクラスタ  $c_S(3) = c_2$ , , セカンダリクラスタへのゲートウェイノード  $e_{\text{gtw}(S)}(3) = e_4$  となる.

Cluster	Keywords for the member persons
1	Fashion, trendy, lowbrow, imitation
2	Health, brave, independent, non group-oriented
3	Commonsense, balance
4	Cheerful, hidden abnormality
5, 6	Autonomous, scholar, abnormal, dark
7	Premature, growing, talent
8	Calm, adult, businessperson

表 8: クラスターの解釈内容 .

る作品と共に図示した (図 33 [a]) . 2つの被験者クラスターを最も強く接続するダミー事象を選択し, 関連付けられたサブジェクトである作品と共に図示した (図 33 [b]) . 最も孤立した被験者クラスターを他の被験者クラスターと接続するダミー事象を選択し, 関連付けられたサブジェクトである作品と共に図示した (図 33 [c]) .

図 33 のシナリオマップから, 好み異なる被験者クラスター間のコミュニケーションを媒介するためのシナリオの創出を試みた . ここでは, 試験者のうちの 4 名が, 1 時間に渡って, シナリオマップを見ながら, 自分の好み, 他人の好み, 好みの違いなどをグループ討議することで, シナリオを創出した . おおよそ, グループ討議の前半がクラスターの解釈, 後半がダミー事象を含む具体的なシナリオの創出である .

まず, 被験者クラスターの特徴を解釈した . 特徴を表すキーワードとして表現した . 結果を表 8 にまとめる . 次に, 被験者クラスターの特徴をふまえた上で, ダミー事象によって媒介される被験者クラスター間のコミュニケーションについて, シナリオを創出した . シナリオは, 新しいコミュニケーションの可能性を探るもので, コミュニケーションを活性化するための媒介となる人物や環境に着目した .

図 33[a] では, ダミー事象の次数が最も大きいものを選択した . 森の脇でじっと佇んでいる鹿を描いた作品 (DE<sub>38</sub>), モノクロームで描かれた花瓶に挿された花を描いた作品 (DE<sub>39</sub>) は, 静けさを志向している . 創出したシナリオは, 「寂しさを共有できる成熟したメンタリティを持つ人物が媒介となれば, 多くの被験者クラスター間のコミュニケーションが促進される」である .

図 33[b] では, 2つのクラスター間に最も多くのダミー事象が表出したク

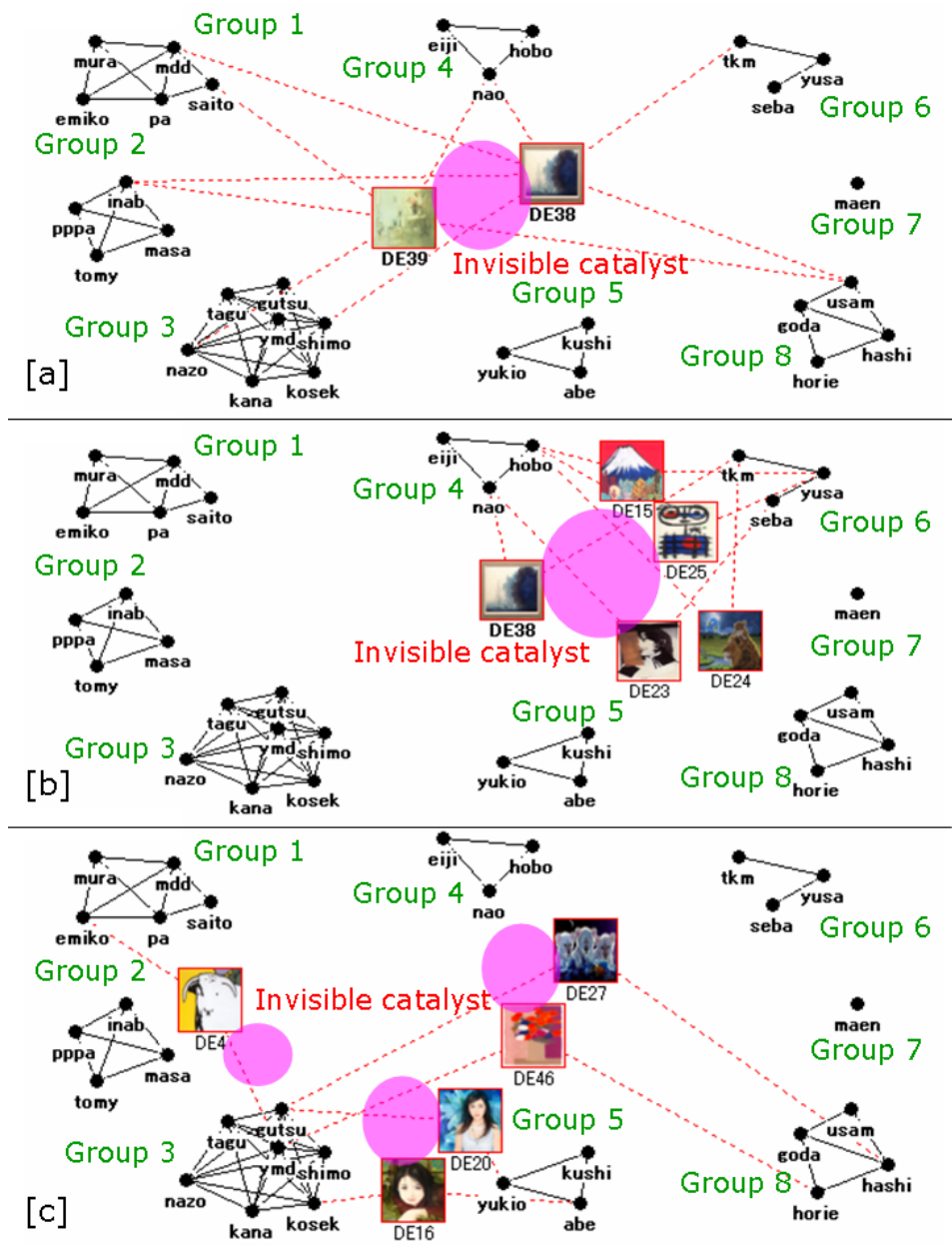


図 33: 好みの異なる 8 つのグループ. ダミー事象  $DE_i$  で示唆される, グループ間のコミュニケーションを媒介する人物. コミュニケーションを媒介する人物には, ダミー事象のサブジェクト  $s_i$  に対応する作品を貼り付けた. [a] 最も大きな次数を持つ媒介人物. [b] 2 つのクラスタを最も強く媒介する人物. [c] 最も孤立したクラスタと他のクラスタを媒介する人物.

Scenario	Proposability	Unnoticeability	Growability	Total
a	Moderate	Moderate	Bad	Bad
b	Good	Moderate	Moderate	Moderate
c	Good	Good	Moderate	Good

表 9: Evaluation of the three scenarios.

ラストを選択し，その間のダミー事象を選択した．選ばれたクラスタは，クラスタ4と6である．カラフルな山々を描いた作品 (DE<sub>15</sub>)，人間のような抽象的な作品 (DE<sub>25</sub>) は，快活で，強い感覚を表現している．ユニークな笑った表情を描いた作品 (DE<sub>23</sub>)，象やしまうまを見つめているライオン (DE<sub>24</sub>) は，刺激的な印象を与える．創出したシナリオは「人々を刺激し勇気付ける個性の強い人物が媒介となれば，被験者クラスタ2と4のコミュニケーションが促進される」である．

図 33[c] では，最も孤立したクラスタ3と他のクラスタを接続するダミー事象を選択した．ダミー事象は，3つの方向で，クラスタを接続する．それぞれに対応して創出したシナリオは「豪華で生き生きとした人物が媒介となれば，被験者クラスタ3と5のコミュニケーションが促進される」(DE<sub>27</sub>, DE<sub>46</sub>) 「女性的な魅力を感じさせる人物が媒介となれば，被験者クラスタ3と8のコミュニケーションが促進される」(DE<sub>16</sub>, DE<sub>20</sub>) である．

次に，創出したシナリオを3つの観点から評価した．観点は，提案可能性（具体的な行動計画になる可能性），気づきにくさ（気づくのが容易かどうか），成長可能性（実際に実行に移される可能性）である．評価の結果を表9にまとめる．評価の高い[c]で創出されたシナリオは興味深い．クラスタ3や8の穏やかな人物が，快活な人物を媒介として，コミュニケーションを活性化する．クラスタ3の規範的な人物が，クラスタ5のユニークな学者的な人物と，女性的な魅力を感じさせる人物を媒介として，コミュニケーションを活性化する．これらのシナリオは，気づきにくいが実行できそうな重要なシナリオであると，合意された．

[a], [b], [c]の順番で，仮説としての魅力が増加する．クラスタの全体構成をうまくつなぐシナリオを考えるよりも，個々のクラスタに着目して，細かい粒度で具体性を追求した方がよいことを示唆している．

Cluster	Interpretation of tastes indicated by the cluster	Keyword
1	colorful, oriental, symbolic	soul
2	transparent, sky-high, clear, calm	peace
3	wonder, paranormal, ghostly	genius
4	abnormal, grotesque	false genius
5	stirring, uplifting, boiling, radical	oppression

表 10: 図 35 のクラスタの解釈結果 .

### 7.3 潜在的な好みを表出するシナリオの創出

7.2 節と同様に、好みに関するアンケートの題材として、50 枚の美術作品を使用する。ここでは、被験者 ID (識別子): { 好きだと感じる作品 ID (識別子) } の形式とする。ここでは、ダミー事象は、弱い好みに対応し、より潜在的、無意識的な好みへのスイッチに相当すると解釈できる。サブジェクト  $s_i$  に対応するスイッチ  $DE_i$  を含むシナリオマップの例を図 34 に模式的に示す。左上のクラスタは、強い好みで、より顕在的、意識的な好みである。右下のクラスタは、弱い好みで、より潜在的、無意識的な好みである。ダミー事象は、これらの好みの中のスイッチに相当する役割を担う。ダミー事象についてのシナリオを想起することで、潜在的、無意識的な好みにアプローチする。

アンケートから得られたシナリオマップを図 35 に示す。直感的な理解を助けるために、個々の作品のイメージをシナリオマップに貼り付けた。5 つのクラスタが示されている。クラスタには、類似性のある作品のグループとそれを好む被験者が含まれている。被験者には 3 種類がある。第一に、強い顕在的な好み (黒リンク)、弱い潜在的な好み (ダミー事象を介した赤リンク) の両方を持つ被験者がいる。幾人か (tkm, and ymd) は、複数の好みを持っている。第二に、強い顕在的な好み (黒リンク) しか持たない被験者 (shimo, and gutsu) がいる。第三に、弱い潜在的な好み (ダミー事象を介した赤リンク) しか持たない被験者 (saito, and emiko) がいる。

まず、美術作品クラスタの特徴を解釈した。特徴を表すキーワードとして表現した。結果を表 10 にまとめる。次に、クラスタの特徴をふまえた上で、ダミー事象によって媒介されるクラスタ間の好みのスイッチについて、シナリオを創出した。18 個のシナリオを創出し、12 個につい

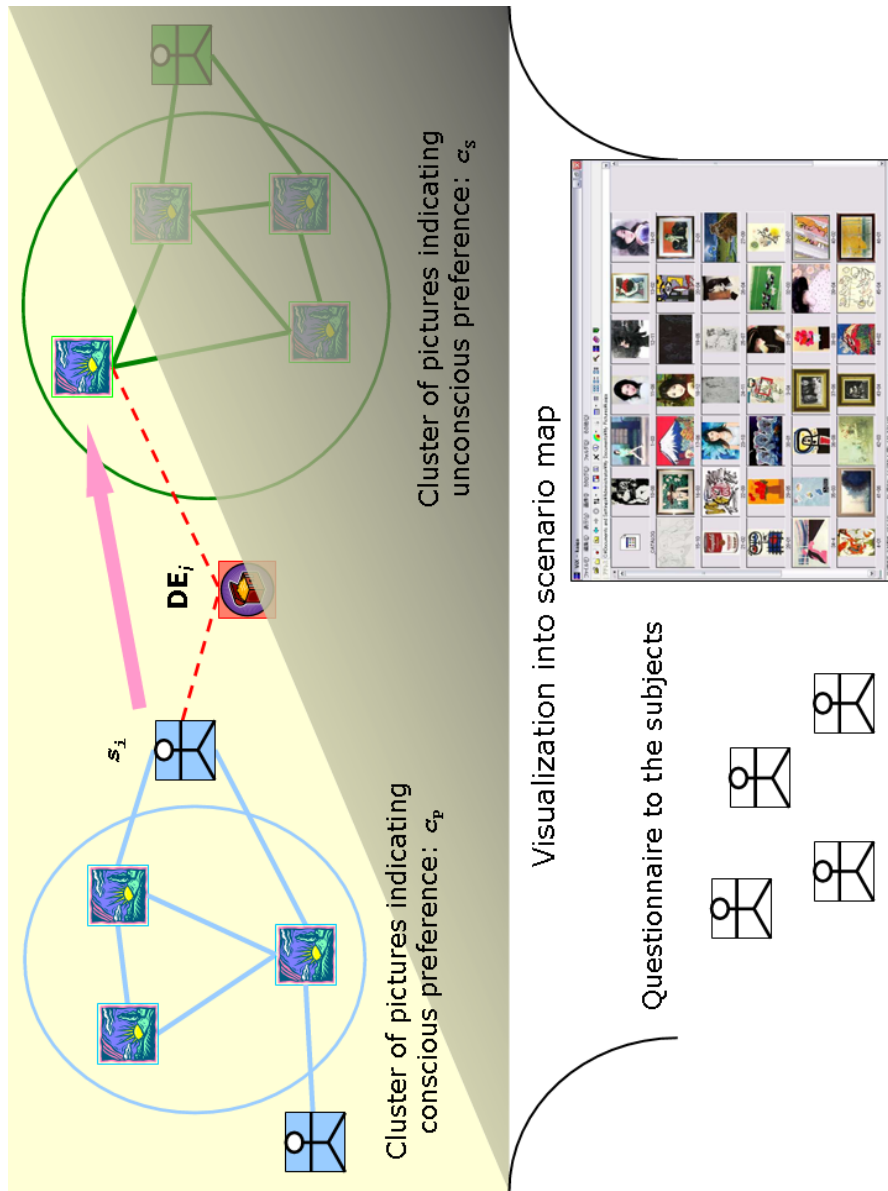


図 34: サブジェクト  $s_i$  に対応するスイッチ  $DE_i$  を含むシナリオマップ。左上のクラスターは、強い好みで、より顕在的、意識的な好みである。右下のクラスターは、弱い好みで、より潜在的、無意識的な好みである。



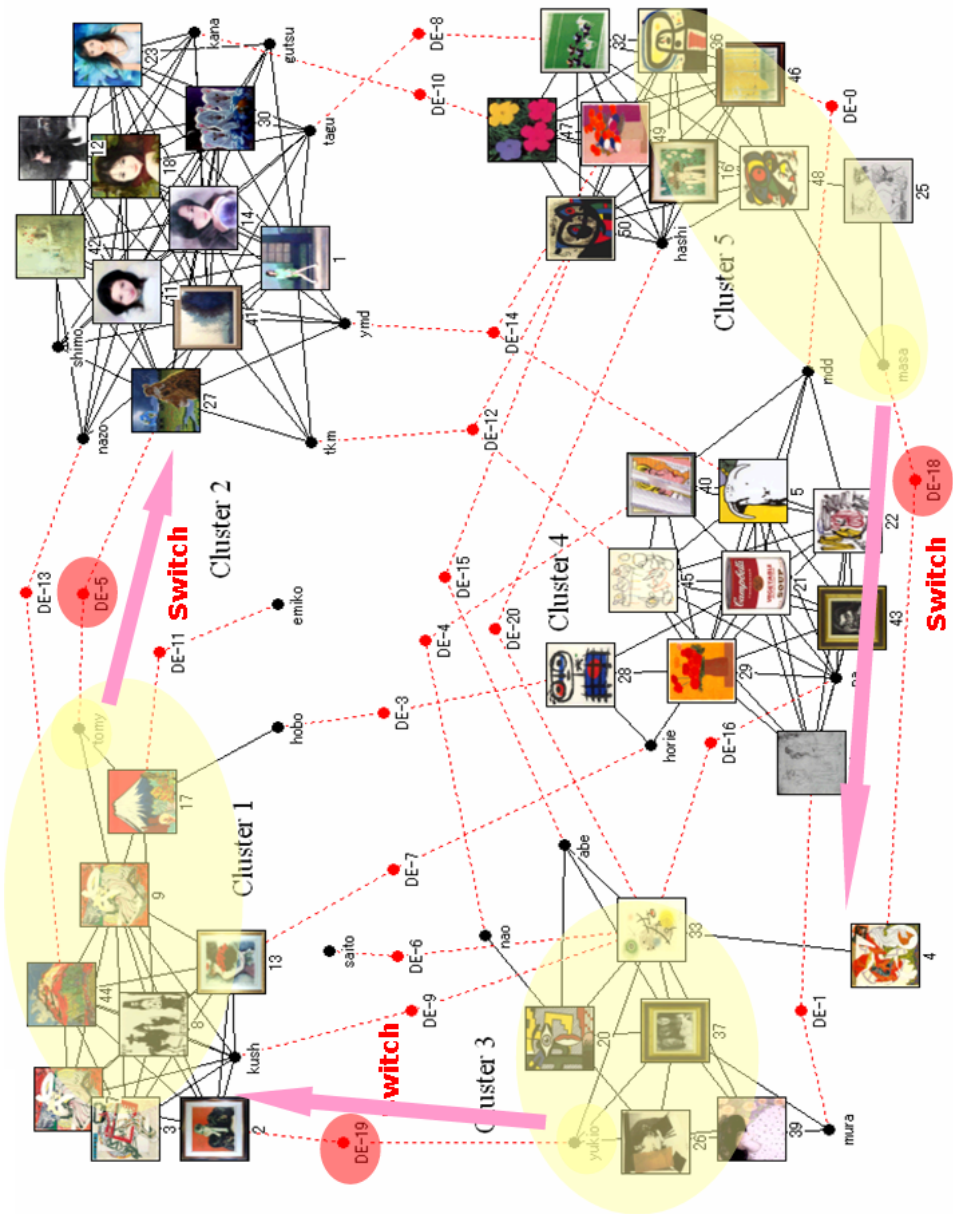


図 35: アンケートから得られたシナリオマップ. 5つのクラスタが示されている. クラスタには, 類似性のある作品のグループとそれを好む被験者が含まれている.

Evaluation item	Score
Number of triggers whose identity is inferred ( $Y_1$ )	12
Number of triggers whose identity is agreeable ( $X_1$ )	10
Rate $X_1/Y_1$	<b>0.83</b>
Number of invented scenarios ( $Y_2$ )	18
Number of scenarios which are agreeable ( $X_2$ )	14
Rate $X_2/Y_2$	<b>0.78</b>

表 11: 創出されたシナリオの評価結果 .

ては、好みのスイッチの実体についての仮説を導いた。以下は3つの例である。

- masa は、自分自身を鼓舞し成功している。しかし、彼は無意識のうちにストレスから逃れたいと思っている。ダミー事象  $DE_{18}$  として、癒しを与えると、スイッチが切り替わり、行動が変容するだろう。
- yukio は、天才的な資質に惹かれている。しかし、彼はゾクッとするような精神的な高揚感にもあこがれている。ダミー事象  $DE_{19}$  として、冒険やギャンブルのおもしろさを知ると、リスクと積極的にリスクを取るような行動をするだろう。
- tomy は、高い目標を掲げ成長しようと努力している。しかし、彼にはありのままの自分でいたいという思いもある。ダミー事象  $DE_5$  として、マニアックな趣味やホビーに関心が向くと、素の自分のままの行動へ傾くだろう。

次に、創出したシナリオを提案可能性（具体的な行動計画になる可能性）、気づきにくさ（気づくのが容易かどうか）、成長可能性（実際に実行に移される可能性）をふまえて、合意に値するものかどうか評価した。シナリオと特定されたダミー事象の実体としてのスイッチについて、合意できる or 合意できないを判定した。評価の結果を表 11 にまとめる。シナリオの 78% が合意され、ダミー事象の 83% が合意された。この結果から、ヒューマンインタラクティブアニメーションの方法が、質の高いシナリオを創出するのに、一定の効果があることがわかる。

このようなシナリオは、マーケティングにおけるブランドのスイッチやコミュニケーションにおける異質グループとの新しいコンタクトを促

すことが期待できる。消費者が無意識のうちに思考や行動に何かをインプットするサブリミナル広告などとは、まったく異なる手法である。この方法は、変化を促すよう、消費者にすでに存在している潜在的な傾向を刺激するものである。刺激が意識化を促し、行動の変容に結びつくことを期待している。

## 7.4 グループ討議によって深める好みの理解

### 7.4.1 グループ討議のプロセス

シナリオマップをツールとして、視覚化・言語化 [石井 2004] を行いながら、グループ討議を進める。以下の説明は、アート作品 [三浦 2003], [尾崎 2004] に対する好みを題材とする前提で行う。気になる商品といった題材でも使えるし、嫌な感じのする広告といった題材へも応用できる。全体プロセスを図 36 に示す。事前調査として、被験者に作品を提示して、好きな作品をいくつでも選んでもらう (質問 Q1)。実験者は、回答から、粒度  $|c|$  を変えた複数のマップを作成して、被験者のグループ討議に向けた準備を行う。ヒューマンインタラクティブアニメーションでは、利用者が事前理解に基づいて、粒度を選択する。しかし、グループ討議では、被験者ごとに事前理解が多様であるため、複数のマップを比較できるようにした。グループ討議の後、実験者が録音した討議の内容からプロトコル分析し、討議で交わされた論点を抽出する。この論点をもとに、被験者に事後調査 (質問 Q7, Q8) を行う。これらの作業は、5 で述べる、実験を分析し検証を行うために設けている。

グループ討議のプロセスを図 37 に示す。まず、被験者に、自分の好みを言葉で書いてもらう (質問 Q2)。グループ討議に入る。グループ討議では、実験者がシナリオマップを提示し、クラスタなどの基本的な構成要素の見方を説明する。複数のシナリオマップを比較しながら、自分の好みや他者の好みについて、自由に討議する。討議は、何らかの結論を求めるものではない。マップに視覚化された、事象や環境の特徴や関係性を捉えようと努める作業は、構成的知覚における現象学的知覚 [諏訪 2004], [諏訪 2005] に相当する。討議が一段落したら、被験者が気づいたことを言葉で書いてもらう (質問 Q3)。これは、知覚した特徴や関係性に意味付けを行う作業であり、構成的知覚における外的表象化に相当する。メタ認知の言語化である。

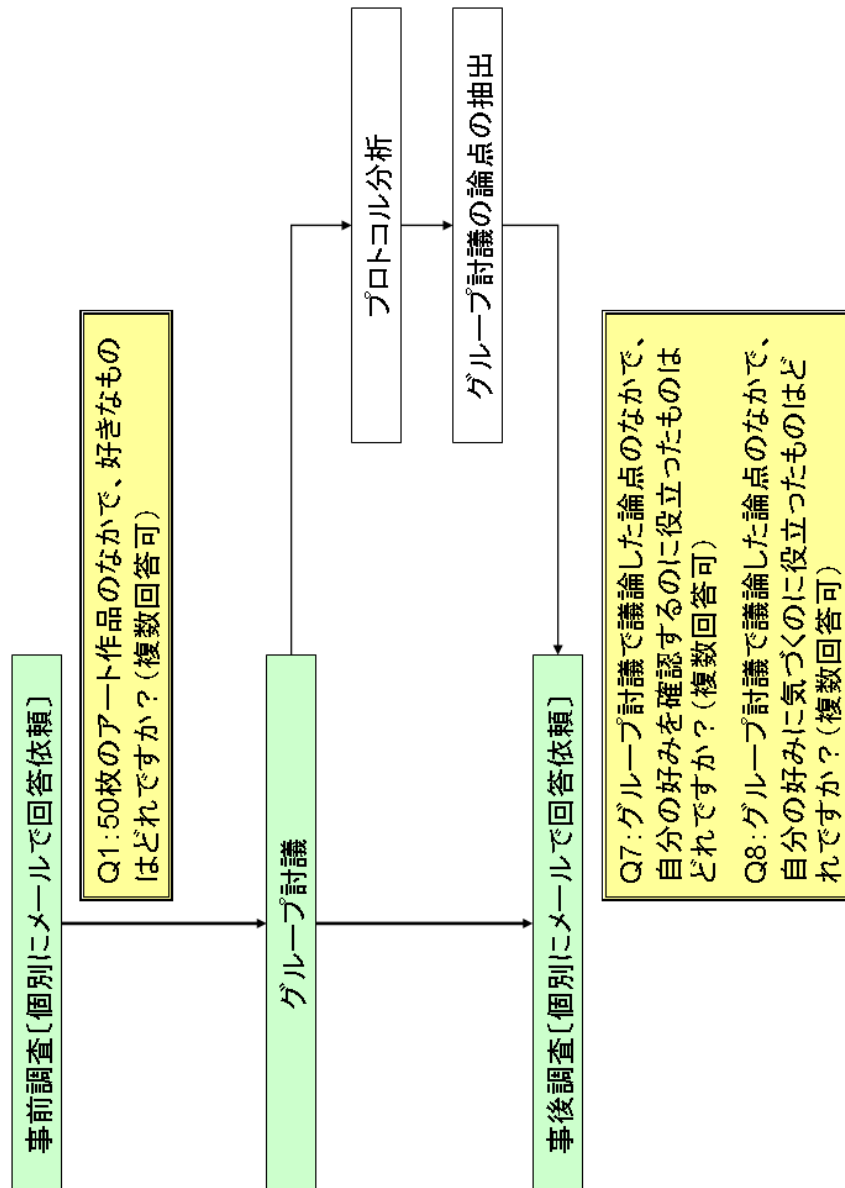


図 36: 全体プロセスと被験者への質問内容。

図 37 では、討議を前半・後半の 2 回としている。前半では、スイッチを含まないものを用いる。ここでは、要素の視覚的特徴、要素と要素との関係性が討議されるのではなか、と想定した。後半では、スイッチを含むものを用いる。ここでは、視覚的特徴、関係性に加えて、要素の間の空白地帯が討議されるのではないかと想定した。前半・後半ともに、粗粒度 (小さなクラスタ数  $|c|$ ) と細粒度 (大きなクラスタ数) との複数のシナリオマップを作成し、同時に見ながら比較する。討議の最後に、被験者が自分の好みを理解するのに、マップが役立ったかどうか答えてもらう (質問 Q5, Q6)。前半・後半の 2 回の討議や質問 Q5~Q8 は、手法の検証、観測事例の蓄積のための観測点として導入している。自分への理解を磨く本来の目的では、これらの手順は付随的な項目であり、省略してもよい。一般には、実験者が作成するマップの種類、討議の盛り上がり、題材に応じた論点の多様性に応じて、討議の回数を設定することになる。

#### 7.4.2 実験条件

図 37 のプロセスに従い、実験を行った。題材となるアート作品は、人物画、風景画、抽象画など、著名な古典作品から現代のポップアートまで多岐に渡る 50 枚の絵画である。著者の環境において利用できる作品を用いた。実験者として、特定の意図があるわけではない。事前調査に応えた被験者は、32 名である。その中で、グループ討議に参加した被験者は、11 名である。グループ討議は 3 回実施し、それぞれに 4 名、2 名、5 名の被験者が参加した。質問への回答も含めて、60~90 分で行った。討議の参加者は、被験者の個人属性を考慮したものではなく、被験者の都合で偶発的に決まったものである。実験者として、特定の意図があるわけではない。

討議の前半では、スイッチを含まないシナリオマップを用いる。粗粒度 (クラスタ数  $|c| = 3, 5$ ) と細粒度 (クラスタ数  $|c| = 7, 8$ ) の 4 枚を用いた。図 38 に、実験で用いた粗粒度  $|c| = 5$  のマップを示す。5 つのクラスタ  $c_1 \sim c_5$  を含む。被験者にとってのプライマリクラスタに相当する。

討議の後半では、前半の 4 枚に対応する、スイッチを含む 4 枚のシナリオマップを用いた。図 39 に、図 38 にスイッチを追加した粗粒度  $|c| = 5$  のマップを示す。煩雑で見にくくならないよう、討議に参加した 11 名の被験者に関するスイッチとセカンダリクラスタへのリンクを図示した。図 40 に、細粒度  $|c| = 8$  のマップを示す。

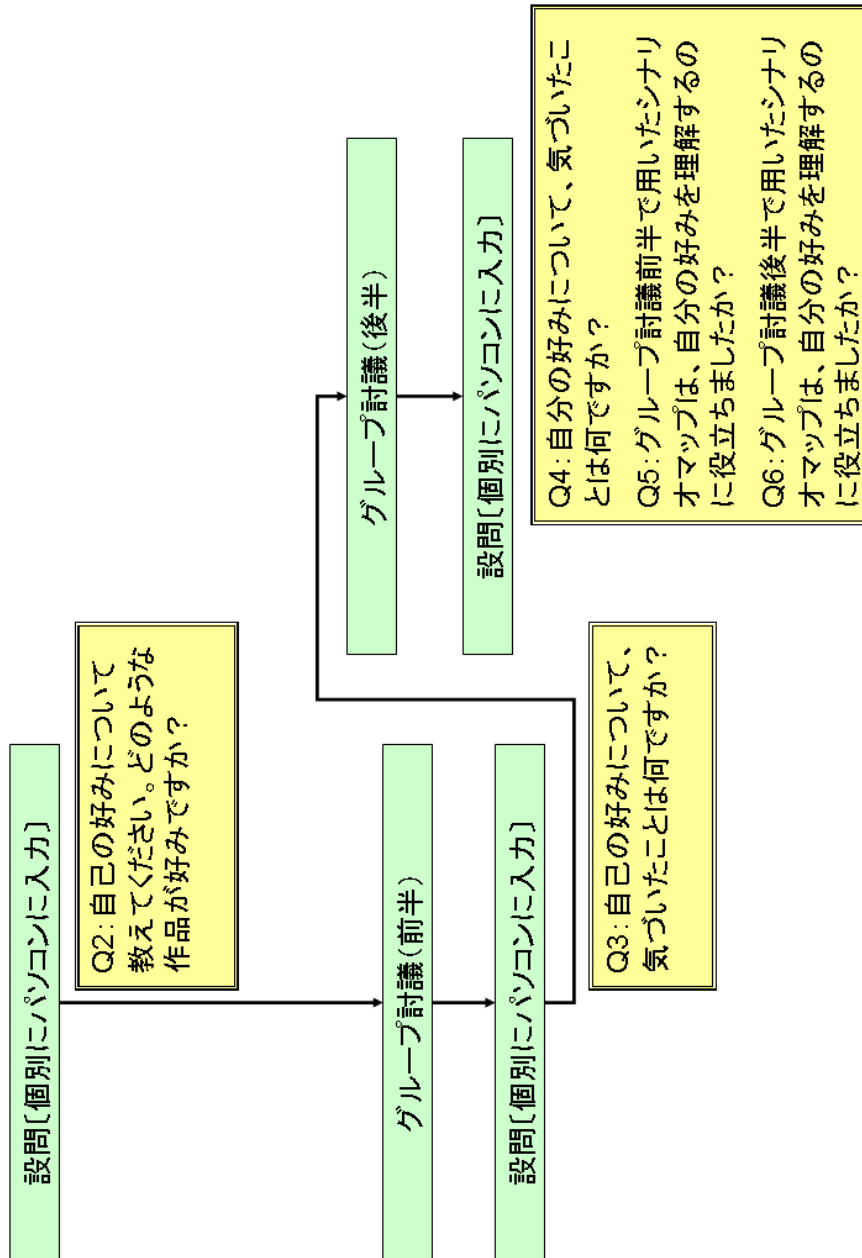


図 37: グループ討議のプロセスと被験者への質問内容。

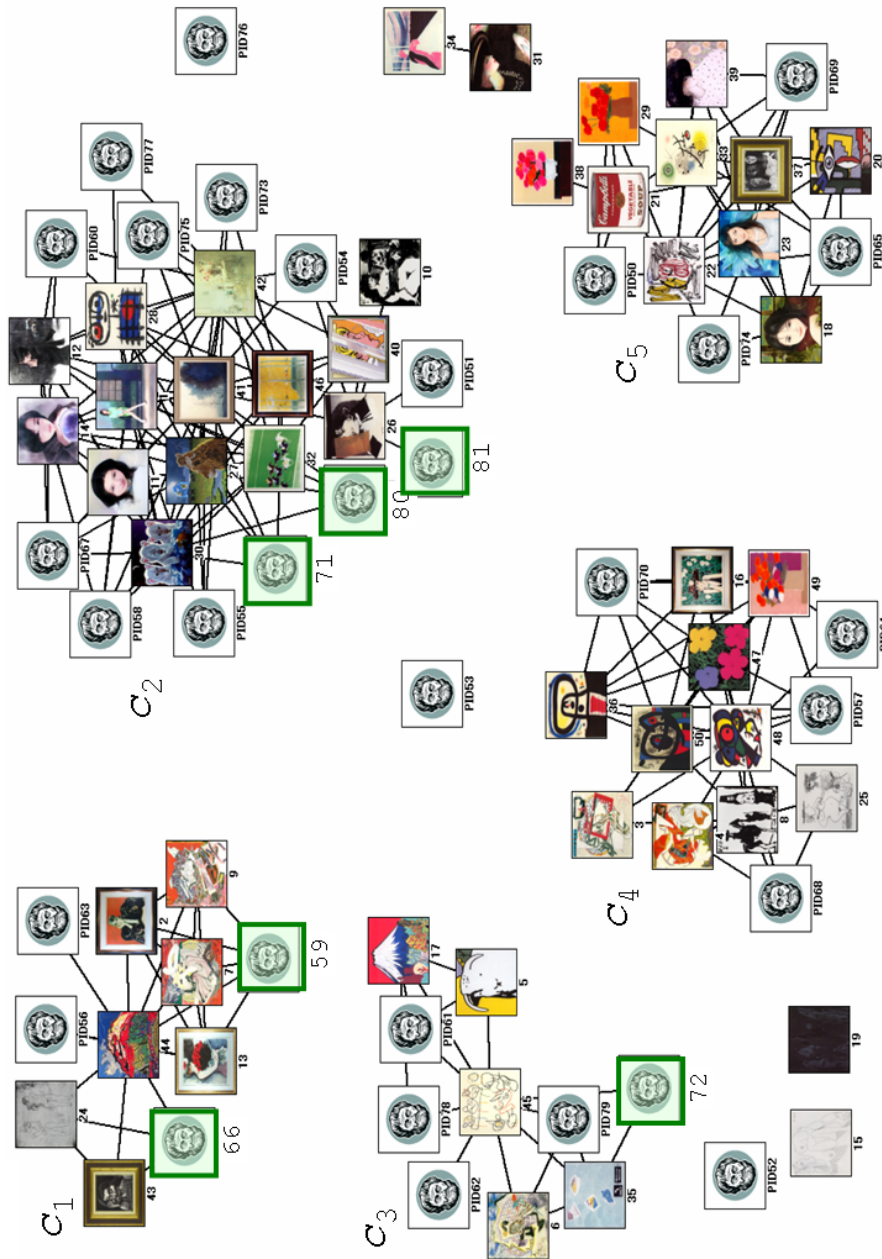


図 38: グループ討議の前半で用いた粗粒度のシナリオマップ . クラスタ数  $|c| = 5$  . (図 38 ~ 図 40 の個々の作品の著作権は , 作者に帰属 . )

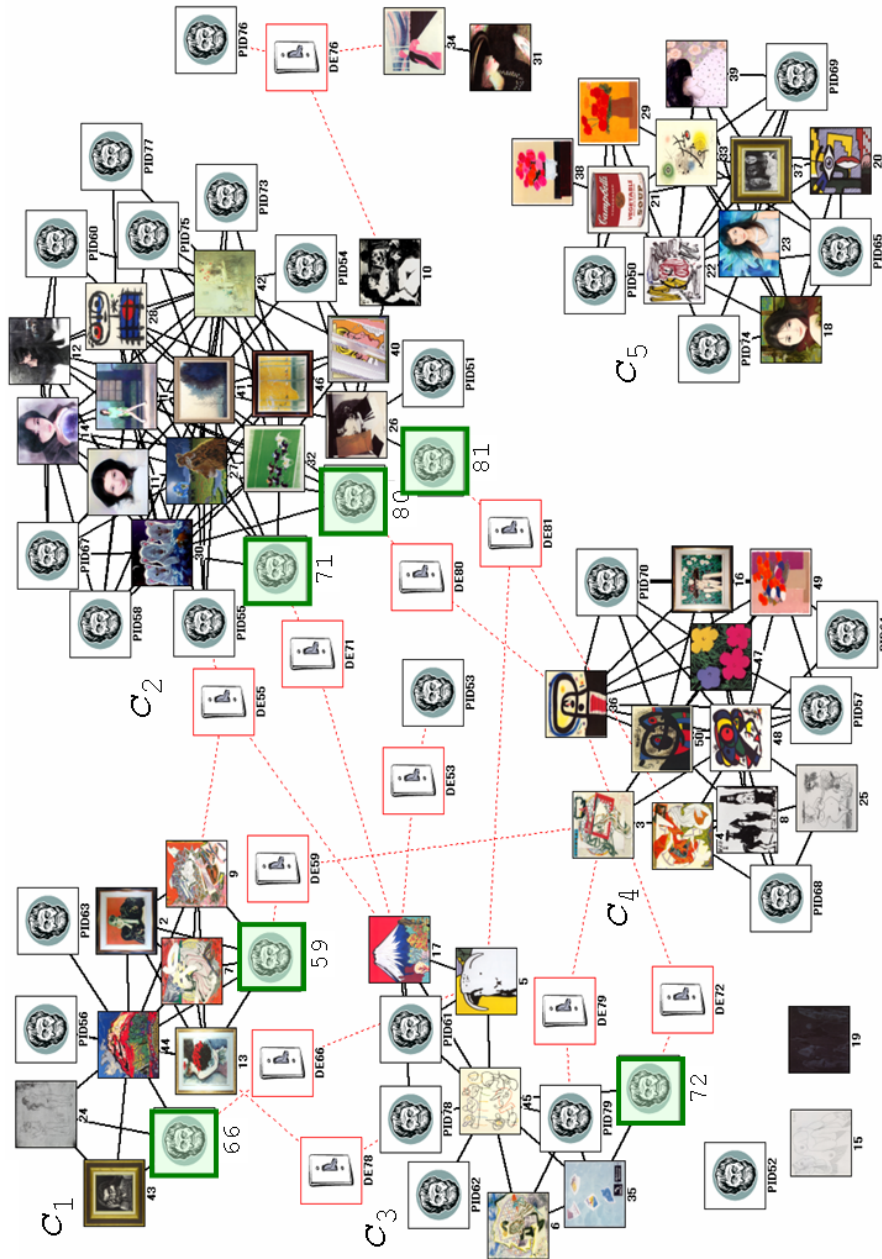


図 39: グループ討議の後半で用いたスイッチを含む, 粗粒度のシナリオマップ. クラスタ数  $|c| = 5$ .



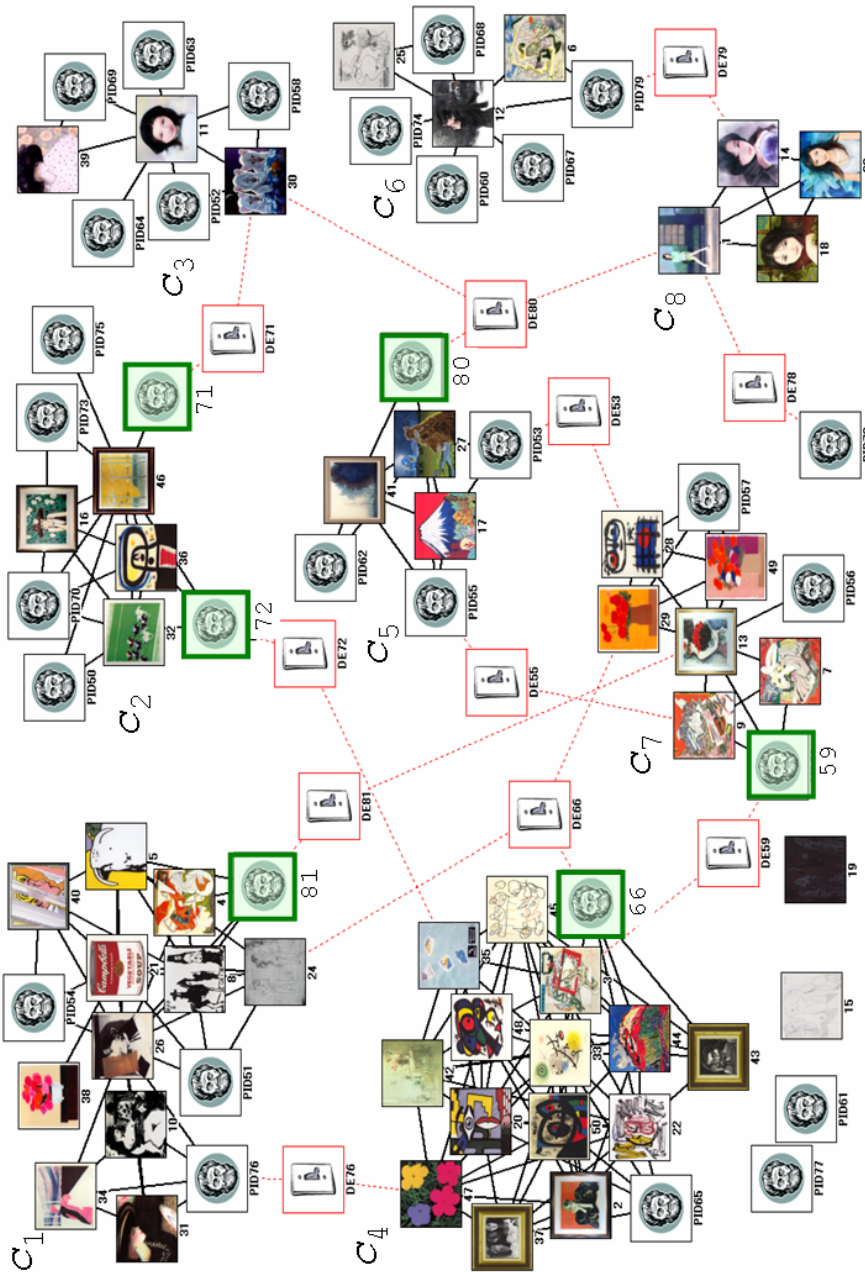


図 40: グループ討議の後半で用いたスイッチを含む，細粒度のシナリオマップ．クラスタ数  $|c| = 8$  .

質問	「はい」の得票数	得票比率
Q5	11	100%
Q6	9	82 %

表 12: 質問 Q5, Q6 への回答結果 .

### 7.4.3 プロセスの評価

グループ討議の参加者への質問 Q2~Q8 の回答を集計し、実験の結果を分析する。プロセス全体の評価 (質問 Q5, 6 への回答)、気づきに役立った討議の論点 (プロトコル分析と質問 Q7, 8 への回答)、参加者ごとの気づきの特徴 (プロトコル分析と質問 Q2~4 への回答) の順に、分析を詳細化する。ここでは、プロセス全体を評価するため、討議におけるシナリオマップが被験者に役立ったか分析する。質問 Q5, Q6 の回答結果を表 7.4.3 にまとめる。Q5 において、すべての参加者が、討議の前半で用いた 4 枚のマップが自分の好みを理解するのに役立ったと回答した。マップの有用性を示す結果である。

Q6 の後半の 4 枚のシナリオマップについては、82%の参加者が役立つと回答したが、役立たないと回答した参加者もいた。この違いの原因を探るため、7.4.4 では、異なるマップを用いた前半と後半との討議の内容の質的な違いを分析する。

### 7.4.4 気づきに役立った論点

気づきに役立った討議の内容を分析する準備として、図 36 のプロセスで録音したプロトコルを分析し、グループ討議の論点を抽出した。録音を再生し、実験者が、被験者が活発に討議した論点の一覧を作成した。実験者の主観に左右される懸念はあるが、活発に討議した論点とは 5 分程度討議が継続した話題とした。その結果、9 個の論点を抽出した。同じ論点が、時系列上で時間をおいて再び現れることもあった。それらは 1 個と数えた。そのうち、5 個は前半の討議に現れたもので、クラスタの解釈が主な論点となった。4 個は後半で観測したもので、スイッチの解釈が主な論点となった。シナリオマップに表現された構造の違いから、これらは自然な論点だと考えられる。

9 個の論点を表 7.4.4 にまとめる。論点に関して、討議を活発化させた具体的な発言の観測例を疑問文の形式で示す。例えば、プロトコル中の論

討議	論点 (番号)	討議のきっかけとなった疑問の例
前半	1. クラスタに属する作品の似ている点	a. どんな色彩, コントラスト, モチーフ, 内容, 意味合いを含むか? b. 抽象・具象, 単純・複雑, 写真・絵, どんな対比で捉えられるか? c. どのクラスタが理解しやすいか?
	2. クラスタに属する人物の似ている点	a. どんな人物が多数派か, 少数派か, 孤立しているか? b. 自分と同じクラスタに属するのは誰か? 納得できるか? 意外か? c. 好みと個性には, どんな関係があるか?
	3. クラスタ単位での好き・嫌い	a. 好きになれそうなクラスタ, 絶対に嫌いなクラスタはどれか? なぜか?
	4. 自分が似ていると思う作品群とクラスタとの差異	a. 似ていると思うものと, クラスタとがなぜ一致しないか? b. 解釈, 事前の予想と違って, クラスタに意外性を感じるのはどこか?
	5. 粒度を変えた時にクラスタに起こる変化	a. 粒度が変化すると, クラスタのどんな組換えが面白い? 意外か? b. 粒度の変化が, 好みの深さや判断の基準と関係しているか? c. 好みの幅の広い人物, 狭い人物が読み取れるか? d. 表層的な好みと深層的な好みは, 人物の個性と関係あるか?
後半	6. 粒度を変えた時に変化しないクラスタ	a. あまり変化しないクラスタについて, 好みの特徴はなにか? b. 人物と作品の組み合わせが変化しないことが起こるのは, なぜか?
	7. スイッチが表すもの	a. どんなスイッチによって, ひとりの人物の好み分散するのか? b. スイッチが多いクラスタと少ないクラスタとの違いは何か? c. スイッチが少ないクラスタの孤立性を解消できるか?
	8. スイッチが繋ぐクラスタの好き・嫌い	a. どの作品まで好きになれるか? ここまでという境界はどこか? b. スイッチでつながったクラスタは, 強い好みと似ているか? 意外か? c. 強い好みと弱い好みとに, 心理的な矛盾があるか?
	9. 粒度を変えた時のクラスタの役割の変化	a. 粒度を変えると, どんな好みの可能性を開拓できそうか? b. 好みの深さと広さについて, 何が読み取れるだろうか?

表 13: プロトコル分析から抽出した 9 個のグループ討議の論点 .

点 5 においては, シナリオマップの粒度の違いから受ける印象の違い (a, b) の議論から, 粒度の違いによって表出したクラスタ構造の変化と個々の被験者の好みの幅 (c) や深さ (d) との関連性の議論へ発展する様子が観測された. 論点 8 では, 弱い好みに相当するセカンダリクラスタを構成する個々の作品の特徴に関心が集まり (a), 被験者の強い好みとの関係性の議論 (b, c) に移行する様子が観測された. 9 個の論点に共通して, 参加者間の作品や被験者の好みについての解釈のくい違い, 個々の参加者がマップを見る前の予想との違いを表明し, 意見を交換する様子が観測された. これらの論点の他に, 一般には考えられるが今回のグループ討議では話題に上らなかった論点もある. これらについては, ここでの分析対象とはしない.

参加者が, 表 7.4.4 に抽出した 9 個の論点から, 自分の好みを確認するのに役立つ論点 (質問 Q7) と自分の好みに気づくのに役立つ論点 (質問 Q8) とを選んだ結果を集計した. 参加者からの得票数を表 7.4.4 にまとめる.

まず, 自分の好みを確認するのに役立つ論点は, 論点 3, 2, 1 などの討議の前半に現れたものが上位を占めた. クラスタ単位での類似性, 好き嫌いの討議が役立つと言える. 討議の後半に現れる論点 6 は順位が低く, 論点 7, 8, 9 は全く選ばれなかった. 討議の前半で論じた, プライマリクラスタとして現れる強い好みは, 好みを確認することに役立っていることがわかる. 著者は, 「参加者は, 自分の好みについての事前理解が, 共通的な好み具体化したクラスタとして視覚化されれば, 自分の

論点番号	質問			
	Q7		Q8	
	得票数	得票順位	得票数	得票順位
1	6	2	3	5
2	5	3	2	8
3	7	1	5	2
4	4	4	3	5
5	3	6	3	5
6	4	4	1	9
7	0	(7)	4	4
8	0	(7)	7	1
9	0	(7)	5	2

表 14: 質問 Q7, Q8 への回答結果 .

理解が他者の観点とずれていないと納得する」と推測している . 今後さらに , 他者の観点の受けとめ方を分析する必要がある .

次に , 自分の好みに気づくのに役立つ論点は , 論点 8 , 9 , 7 などの討議の後半に現れたものが上位を占めた . 論点 3 のクラスタ単位の好き・嫌いの議論は , 自分の好みを確認することにも , 気づくことにも寄与した . しかし , 論点 2 , 1 は , 気づくのに役立つ論点での得票順位を落とした . クラスタに属する作品群や人物群の詳細は , 気づくことにあまり大きな影響を与えないことがわかる . 表 7.4.4 に示すように , 得票順位 1 位の論点 8 では , 弱い好みを表すセカンダリクラスタと強い好みを表すプライマリクラスタとの関連性を議論していた . スイッチを含むシナリオマップから , 弱い好みを視覚化・言語化する作業が , 自分の好みに気づくのに役立つことを示している . 著者は , 「強い好みとの視覚的な対比は , 言語化する際の尺度が捉えやすくなり , 弱い好みの知覚の促進に有効である」と推測している . 今後さらに , 対照実験による分析が必要である .

上記の結果に , 表 7.4.3 で討議の後半で気づきを得られなかった参加者がいたことを加味すると , 弱い好みに着目して視覚化・言語化する行為が気づきを必ず保証するわけではないが , スイッチを含むシナリオマップの方が気づきに寄与する , と言える . 弱い好みが気づきに寄与する有用な着眼点である点が , 本実験から得られた第一の知見である . では , 参

被験者 PID	質問	回答
59	Q2	モチーフとして、人物は好きではない。
	Q3	自分の好みには、抽象度（解釈可能な抽象性）、色（暖色）、モチーフ（風景）が、この順で重要である。
	Q4	質問 3 と同じ。
66	Q2	色使いより構図で好み分けられる。
	Q3	モチーフは好き嫌いに関係ない。強過ぎる作品、広告のような作品は好きではない。
	Q4	線の動き、筆の使い方などで心地良さが決まる。
71	Q2	特に、好みはない。嫌いな作品は思いつかない。
	Q3	意外と好みがあることがわかった。赤や黄色の原色は好きではない。女の子の作品が気になる。
	Q4	粒度が細かいほど、自分の直感を表現していると感じる。
72	Q2	風景画、描写がシンプルでごちゃごちゃしていない作品が好き。
	Q3	シンプルなやさしい癒される作品が好き。書き込み過ぎの作品、強過ぎる作品は嫌い。
	Q4	質問 3 と同じ。
80	Q2	自然描写で主張を感じさせる作品、ほっとさせる作品が好き。
	Q3	自然描写が好きなのは変わらないが、抽象画のシンプルさもひとつの主張だと感じてきた。
	Q4	自然描写だけだと疲れてしまう自分に気づいた。
81	Q2	面白い作品、わかりやすい作品、印象に残る作品、かわった作品が好き。
	Q3	自分のクラスに属す女の子の作品は意外だが、同じクラスターの静物画は好き。
	Q4	弱い好みは、気分によって好きだと思う作品、人に影響されて好きになった作品だと思う。

表 15: 質問 Q2, Q3, Q4 への回答例。

加者がどのような気づきを得たのか、分析を深める必要がある。そこで、7.4.5 では、参加者ごとの気づきの特徴を分析する。

#### 7.4.5 気づきの特徴

参加者ごとの気づきの特徴として、自分の好みについての事前の理解（質問 Q2 への回答）と、参加者が気づいたこと（討議の前半後の質問 Q3、後半後の Q4）との例を表 7.4.5 に示す。表 7.4.3 において、討議の後半が気づきを得るのに役立たなかったと回答した参加者と、役立ったと回答した参加者とに分けて述べる。

まず、討議の後半が気づきを得るのに役立たなかったと回答した参加者を探りあげる。人物 PID<sub>59</sub> は、質問 Q2 から、自分の好みにはモチーフという観点が重要だと事前に理解していたことがわかる。さらに、質問 Q3 から、討議の前半で、抽象度と色という新しい観点に気づいたことがわかる。プロトコルから、討議の前半で、粗粒度のシナリオマップでの強い好みを表すプライマリクラスター  $c_1$  と他のクラスターとの比較についての発言が観測された。これが、 $c_1$  の作品群の表現の特徴を細かく意識し、観点の開拓につながった可能性がある。質問 Q4 から、討議の後半には、新しい気づきを得られなかったことがわかる。著者は、「PID<sub>59</sub> の気づきが、意識に上りにくい好みの開拓の方向にではなく、事前理解していた好みの緻密化の方向に向かったため」と推測している。

類似の例が、PID<sub>72</sub> でも観測された。PID<sub>72</sub> は、質問 Q2 から、自分の好みはシンプルさにあることを事前に理解していたことがわかる。さらに、質問 Q3 から、討議の前半で、シンプルさという作品の表現の特徴が、

癒されるやさしさという心理的な志向性と関連していることに気づいたことがわかる。プロトコルから、細粒度のシナリオマップでのプライマリクラスター  $c_3$  の中に、自然描写と抽象画が混在している点を指摘する発言が観測された。ここから、シンプルさがモチーフの解釈しやすさを意味するのではなく、線や面の図形としてのあっさり感を意味することを知覚した可能性がある。質問 Q4 から、討議の後半では、新しい気づきを得られなかったことがわかる。著者は、「PID<sub>59</sub> と同じく、気づきが好みの緻密化に向かったため」と推測している。今後さらに、個々の参加者がマップを読み解く際の言動について、詳しいデータを収集し分析する必要がある。

次に、他の参加者は、討議の後半が気づきを得るのに役立ったと回答した。質問 Q2, Q3 から、PID<sub>80</sub> は、討議の前半では、自然描写が好きだと感じており、事前の理解と一致していたことがわかる。質問 Q4 から、討議の後半を通して、自然描写だけだと疲れてしまう自分に気づいたことがわかる。プロトコルから、粗粒度のシナリオマップのセカンダリクラスター  $c_4$  の中の作品が、プライマリクラスター  $c_2$  の自然描写からずれているとの発言が観測された。これが、気づきのきっかけになった可能性がある。PID<sub>81</sub> の気づきは、PID<sub>80</sub> の気づきに似ている。質問 Q2, Q3 から、討議の前半で事前の理解に近い好みを感じていたことがわかる。質問 Q4 から、後半では、気分や人の影響という要因が自分の好みの一部を構成していると気づいたことがわかる。プロトコルから、細粒度のシナリオマップのセカンダリクラスター  $c_7$  の作品の好き・嫌いを述べている点が観測された。これが、気づきのきっかけになった可能性がある。このように、PID<sub>80</sub> と PID<sub>81</sub> は、自分の好みを構成する新しい要因に見出したことがわかる。これは、本実験の狙いに最も近い結果である。

他の例として、PID<sub>66</sub> は、質問 Q4 から、事前に理解していた構図(質問 Q2 の回答)だけでなく、線の動きや筆の使い方が自分の好みを左右すると気づいたことがわかる。PID<sub>71</sub> は、質問 Q4 から、自分には好き嫌いが少なく作品の差異を感じないとの事前の理解(質問 Q2 の回答)とは異なり、細粒度のシナリオマップを見ると作品ごとの好き嫌いを直感的に感じると気づいたことがわかる。

このように、気づく内容は極めて多様である。気づきは、好む作品の表現の特徴に対する分析を深める方向に進む場合もあれば、潜在性のある、意識に上りにくい好みに気づいたり、作品を選ぶ上での背景の洞察を深めたりする場合もある。いずれの場合にも、自分への理解を磨くこ

とに貢献するものと考えられる。また，個々の参加者の事前理解のレベルは多様であり，気づきにつながるシナリオマップの粒度は異なることがわかる。参加者の違いに適応するには，マップの粒度のような可変のパラメータを含む可視化が有用である点が，本実験から得られた第二の知見である。

## 7.5 まとめ

ヒューマンインタラクティブアニメーションの応用として，グループ討議によって個人の理解を深める手法を開発，検証した。アートの好みに対するアンケートを題材として，二部グラフ構造を持ったアンケートの回答結果を取り扱えるよう，アルゴリズムを改良した。

ヒューマンネットワークの形式としたシナリオマップからコミュニケーションの活性化シナリオを創出した。シナリオの品質を提案可能性（具体的な行動計画になる可能性），気づきにくさ（気づくのが容易かどうか），成長可能性（実際に実行に移される可能性）で評価すると，クラスタの全体構成をうまくつなぐシナリオを考えるよりも，個々のクラスタに着目して，細かい粒度で具体性を追求した方がよいことがわかった。

作品群クラスタの形式としたシナリオマップから潜在的な好みを表出するシナリオを創出した。ダミー事象を含むことで，新しい要因を追加したシナリオを作成することができた。さらに，グループ討議によって個人の理解を深める手法として，視覚化・言語化によって，自己の認知を再確認し，解釈のための枠組みを再構築することを支援するプロセスに改良した。気づかなかった好みの発見，心理面の記述の発生，好みの具体化といった，被験者それぞれで異なるタイプの何らかの発見があったことを確認した。

## 8 結論

はじめに、見えない構造への足がかりとして、ヒューマンインタラクティブアニメーションのプロセスの提案とシミュレーションによる基本的な特性の研究について述べる。ヒューマンインタラクティブアニメーションは、クラスタ間の隙間についての仮説を創出するプロセスを体系化したものである。隙間とは、関連性の高い事象群（クラスタ）の間の空白地帯でありながら、クラスタとの関係性が観測された領域である。参加者は、空白地帯を含むシナリオマップを描きながら、シナリオを創出する。自己の事前理解（誤解）を確認しながら、シナリオマップと仮説とを同時に作成するプロセスに特色がある。隙間を探すためのアルゴリズムを開発した。ネットワーク上のコミュニケーションから、見えないハブノードを探すシミュレーションでは、事前知識やネットワーク構造がゆらいでも、良好な性能が得られることを示した。見えないノードの確からしさを評価する様々なランキング関数の性能、コミュニケーション強度に対する依存性についても検証した。

次に、組織的な意思決定を取り巻く、見えない構造を探る問題に取り組んだ。20名の参加者で、意思決定の隠れた黒幕を探る実験を行った。組織の意思決定を目的とした電子メールによるコミュニケーションを観測データとした。黒幕は、電子メールを使用せず、観測データに残らない方法で参加者に指示を与え、意思決定に影響を与える。この見えない黒幕を探る実験において、有効な仮説を創出できた。提案するプロセスに従って、事前知識によると独立性の高そうな参加者クラスタ間に、クラスタを結びつけるダミーノードが出現した。このヒントを解釈し、事前理解と観測データへの認知とのずれから気づきを獲得し、事前知識には欠けていた仮説を創出した。ヒューマンインタラクティブアニメーションが、組織のコミュニケーションと意思決定を誘導する、見えない黒幕を探るための有効な情報を与えることを示した。

9/11 テロ組織についての文献の情報をもとに、実行犯を支援した共謀者についてのネットワークを分析した。トポロジから見たグローバル (peripherarity, partiality) ・ローカル (degree, clustering coefficient) な重要性指標を導入した。重要性比率によって、共謀者が2種類に分類できることを見出した。タイプ1の共謀者は、文字通り、クラスタ間の bridge の役割を行う。タイプ2の共謀者は、外部世界との bridge の役割を担う。このネットワークに基づくシミュレーションを行った。タイプ1の共謀者について、標準偏差を用いたアルゴリズムが効果的に機能することを



示した。一方、タイプ2の共謀者を探ることは難しく、関連性の高い上位2つのクラスタに注目すると性能が向上するものの、不完全であることがわかった。

最後に、応用として、グループ討議で好みについての行動・心理への理解を深める問題に取り組んだ。アートの好みに対するアンケートを題材として、二部グラフ構造を持ったアンケートの回答結果を取り扱えるよう、アルゴリズムを改良した。ヒューマンネットワーク形式のシナリオマップからコミュニケーションの活性化シナリオを創出した。シナリオの品質を提案可能性、気づきにくさ、成長可能性で評価すると、クラスタの全体をつなぐシナリオよりも、個々のクラスタに着目した細かい粒度のシナリオの方が、具体的でよいことがわかった。さらに、作品群クラスタ形式のシナリオマップから潜在的な好みを表出するシナリオを創出した。ダミー事象を含むことで、新しい要因を追加したシナリオを作成できた。また、グループ討議によって個人の理解を深める手法として、視覚化・言語化によって、自己の認知を再確認し、解釈の枠組みを再構築することを支援するプロセスに改良した。気づかなかった好みの発見、心理面の記述の発生、好みの具体化といった、被験者ごとに異なる何らかの発見があったことを確認した。

残された課題を議論する。アルゴリズム面では、時系列データや数値データへの対応が挙げられる。実証面では、見えない重要人物が存在しない組織との対比が必要だろう。実験が示すように、観測データの中から、コミュニケーションや組織の機能に重要な見えない事象が強く関連する部分を取り出すことに成功した。しかし、それが、人物としての黒幕の存在を保証するとは限らない。例えば、グループの境界を越えて共有された危機意識や、意思決定にバイアスをかけるメディアや流行かもしれない。一方、テロリスト対策では、不完全な仮説に基づく過剰な警報が発信されても、人命に係わる重要事象を見過ごすよりもよいだろう。不完全な仮説の創出の抑制よりも、新たな観測や多様な人間の知見を取り込みやすくして、迅速な仮説の検証を促進する方が、有益だと考えられる。手法面では、コミュニケーションに付随する発生タイミングや発生順序といった時系列情報の活用が重要になるだろう。バッチ処理ではなく、オンライン処理も重要である。応用面では、分野に固有の手法を蓄積する必要があるだろう。仮説の創出には、問題に特有の知見が不可欠であり、企業の経営革新や社会問題の解決の現場において、経験を深める必要があるだろう。

## 謝辞

本論文をまとめるにあたり，多くの方々のご指導，ご協力を賜りました．ここに，お世話になった方々への心からの感謝の意を表します．

本論文のご指導をいただいた，大澤幸生 東京大学准教授に心から感謝いたします．椿広計 筑波大学教授（統計数理研究所教授），領家美奈 筑波大学准教授には，貴重なご指導をいただきました．元田浩 大阪大学名誉教授，新田克己 東京工業大学教授には，度々有用なご助言をいただきました．吉田健一 筑波大学教授には，学内発表会・学会において，適切なご指摘をいただきました．今日に至るまで，先生方から終始適切なご指導をいただいたことは，研究の礎となる貴重な財産となりました．

久代紀之さま，堀江健一さま，郷田慎一さま，保々佐和子さま，中村潤さまには，有用で刺激的なご議論をいただきました．筑波大学，東京大学の学生の皆様には，実験にご協力していただきました．

最後に，いつも明るくサポートしてくれた妻訓江に感謝いたします．

## 参考文献

- [Adamic 2003] L. A. Adamic, and E. Adar: Friends and neighbors on the web, *Social Networks* **25**, 211-228 (2003).
- [Adamic 2001] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. Huberman: Search in power-law networks, *Physical Review E* **64**, 046135 (2001).
- [Aggarwal 2004] C. C. Aggarwal: A human-computer interactive method for projected clustering, *IEEE Transactions on Knowledge and Data Engineering* **16**, 448-460 (2004).
- [Albert 2000] R. Albert, and A. L. Barabási: Topology of complex networks: Local events and universality, *Physical Review Letters* **85**, 5234-5237 (2000).
- [Albert 2000] R. Albert, and A. L. Barabási: Dynamics of complex systems: Scaling laws for the period of boolean networks, *Physical Review Letters* **84**, 5660-5663 (2000).
- [Albert 2000] R. Albert, and H. Jeong, and A. L. Barabási: Error and attack tolerance of complex networks, *Nature* **406**, 378-381 (2000).
- [Albert 1999] R. Albert, H. Jeong, and A. L. Barabási: Diameter of the World-Wide Web, *Nature* **401**, 130-131 (1999).
- [Allanach 2004] J. Allanach, T. Haiying, S. Singh, P. Willett, and K. Pattipati: Detecting, tracking, and counteracting terrorist networks via hidden Markov models, *Proceedings of the IEEE Aerospace Conference, Big Sky*, 3246-3257 (2004).
- [Anderson 1999] C. Anderson, S. Wasserman, and B. Crouch: A  $p^*$  primer: Logit models for social networks, *Social Networks* **21**, 37-66 (1999).
- [Arora 1998] N. Arora, G. M. Allenby, and J. L. Ginter: A hierarchical Bayes model of primary and secondary demand, *Marketing Science* **17**, 29-44 (1998).
- [Barabási 1999a] A. L. Barabási, and R. Albert: Emergence of scaling in random networks, *Science* **286** 509-512 (1999).
- [Barabási 1999b] A. L. Barabási, R. Albert, and H. Jeong: Mean-field theory for scale-free random networks, *Physica A* **272**, 173-187 (1999).

- [Batallas 2006] D. A. Batallas, and A. A. Yassine: Information leaders in product development organizational networks: Social network analysis of the design structure matrix, *IEEE Transactions on Engineering Management* **53**, 570-582 (2006).
- [Besag 1974] J. Besag: Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society B* **36**, 96-127 (1974).
- [Bianconi 2001] G. Bianconi, and A. L. Barabási: Bose-Einstein condensation in complex networks, *Physical Review Letters* **86**, 5632-5635 (2001).
- [Bishop 1998] C. M. Bishop, and M. E. Tipping: A hierarchical latent variable model for data visualization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 281-293 (1998).
- [Bollen 1989] K. A. Bollen: *Structural equations with latent variables* (Wiley series in probability and statistics). Wiley-Interscience (1989).
- [Borgatti 1997] S. P. Borgatti, and M. G. Everett: Network analysis of 2-mode data, *Social Networks* **13**, 155-168 (1997).
- [Carreira-Perpinan 1999] M. A. Carreira-Perpinan: One-to-many mappings, continuity constraints and latent variable models, *Proceedings of the IEE Colloquium on Applied Statistical Pattern Recognition*, Birmingham, 14/1-14/6 (1999).
- [Carreira-Perpinan 1998] M. A. Carreira-Perpinan, and S. Renals: Experimental evaluation of latent variable models for dimensionality reduction, *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, Cambridge, 165-173 (1998).
- [Carrington 2005] P. J. Carrington, J. Scott, and S. Wasserman eds.: *Models and methods in social network analysis*. Cambridge University Press (2005).
- [Chakrabarti 1991] A. K. Chakrabarti: Competition in high technology: analysis of patents of US, Japan, UK, France, West Germany, and Canada, *IEEE Transactions on Engineering Management* **38**, 78 - 84 (1991).
- [Chakrabarti 1993] A. K. Chakrabarti, I. Dror, and N. Eakabuse: Interorganizational transfer of knowledge: an analysis of patent citations of a defense firm, *IEEE Transactions on Engineering Management* **40**, 91-94 (1993).

- [Cheung 2004] K. W. Cheung, K. C. Tsui, and J. Liu: Extended latent class models for collaborative recommendation, *IEEE Transactions on Systems, Man & Cybernetics, Part A* **34**, 143-148 (2004).
- [Choudhury 2002] A. Choudhury, Y. S. Ong, and A. J. Keane: Extracting latent structures in numerical classification: an investigation using two factor models, *Proceedings of the International Conference on Neural Information Processing, Singapore*, 1842-1846 (2002).
- [Cusumano 2000] M. Cusumano: How Microsoft makes large teams work like small teams, *Sloan Management Review* **39**, 9-20 (2000).
- [Dempster 1977] A. Dempster: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistics Society* **B39**, 1-38 (1977).
- [Duda 2000] R. O. Duda, P. E. Hart, and D. G. Stork: *Pattern classification* (2nd edition). Wiley- Interscience (2000).
- [Everett 1998] M. G. Everett, and S. P. Borgatti: Analyzing clique overlap, *Connections* **21**, 49-61 (1998).
- [Fenner 2007] T. Fenner, M. Levene, and G. Loizou: A model for collaboration networks giving rise to a power law distribution with an exponential cutoff, *Social Networks* **29**, 70-81 (2007).
- [Figueiredo 2002] M. A. F. Figueiredo, and A. K. Jain: Unsupervised learning of finite mixture models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, 381-396 (2002).
- [Frank 1986] O. Frank, and D. Strauss: Markov graphs, *Journal of the American Statistical Association* **81**, 832-842 (1986).
- [Freeman 1979] L. C. Freeman: Centrality in networks: I. Conceptual clarification, *Social Networks* **1**, 215-239 (1979).
- [Friedman 1998] N. Friedman, K. Murphy, and S. Russell: Learning the structure of dynamic probabilistic networks, *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence, Madison*, 139-146 (1998).
- [Fruchterman 1991] T. M. J. Fruchterman and E. M. Reingold: Graph drawing by force-directed placement, *Software - Practice and Experience* **18**, 1129-1164 (1991).

- [Geiger 1994] D. Geiger, and D. Heckerman: Learning Gaussian networks, Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence, Seattle, 235-243 (1994).
- [Getoor 2005] L. Getoor, and C. P. Diehl: Link mining: a survey, ACM SIGKDD Explorations Newsletter **7**, 3-12 (2005).
- [Gill 2002] Jeff Gill: Bayesian methods - A social and behavioral sciences approach. Chapman & Hall/CRC (2002).
- [Girolami 2001] M. Girolami: The topographic organization and visualization of binary data using multivariate- Bernoulli latent variable models, IEEE Transactions on Neural Networks **12**, 1367-1374 (2001).
- [Glymour 1999] C. Glymour, and G. F. Cooper eds.: Computation, Causation, and Discovery. Amer Assn for Artificial (1999).
- [Green 1981] P. Green, J. Carroll, and S. Goldberg: A general approach to product design optimization via conjoint analysis, Journal of Marketing **43**, 17-37 (1981).
- [Greenwald 1998] A. G. Greenwald, D. E. McGhee, and J. L. K. Schwartz: Measuring individual differences in implicit cognition: the implicit association test, Journal of Personality and Social Psychology **74**, 1464-1480 (1998).
- [Haiying 2006] T. Haiying, J. Allanach, S. Singh, K. R. Pattipati, and P. Willett: Information integration via hierarchical and hybrid bayesian networks, IEEE Transactions on Systems, Man and Cybernetics, Part A **36**, 19-33 (2006).
- [Hastie 2001] T. Hastie, R. Tibshirani, and J. Friedman: The elements of statistical learning - Data mining, inference, and prediction (Springer series in statistics). Springer-Verlag (2001).
- [Heckerman 1996] D. Heckerman, and J. S. Breese: Causal independence for probability assessment and inference using Bayesian networks, IEEE Transactions on Systems, Man & Cybernetics, Part A **26**, 826-831 (1996).
- [Hinneburg 1999] A. Hinneburg, D. A. Keim, and M. Wawryniuk: HD-Eye: Visual mining of high-dimensional data, IEEE Computer Graphics and Applications, September/October, 22-31 (1999).
- [Hori 1994] Koichi Hori: A system for aiding creative concept formation, IEEE Transactions on Systems, Man, & Cybernetics **24**, 882-894 (1994).

- [Hori 2004] K. Hori, K. Nakakoji, Y. Yamamoto, and J. Ostwald: Organic perspectives of knowledge management, *Journal of Universal Computer Science* **10**, 252-261 (2004).
- [Horie 2007] K. Horie, Y. Maeno, and Y. Ohsawa: Human-interactive annealing process with pictogram for extracting new scenarios for patent technology, *Proceedings of CODATA (International Council for Science: Committee on Data for Science and Technology) International Conference, Beijing* (2007).
- [Jain 2002] S. Jain, and S Krishna: Crashes, recoveries, and "core shifts" in a model of evolving networks, *Physical Review E* **65**, 026103 (2002).
- [Johansen 2006] A. Johansen, A. Doucet, and M. Davy: Maximum likelihood parameter estimation for latent variable models using sequential Monte Carlo, *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, III/640-III/643* (2006).
- [Johnston 2000] D. A. Johnston, and J. D. Linton: Social networks and the implementation of environmental technology, *IEEE Transactions on Engineering Management* **47**, 465-477 (2000).
- [Kauffman 1996] Stuart Kauffman: *At home in the universe - The search for laws of self-organization and complexity*. Oxford University Press (1996).
- [Kecman 2001] V. Kecman: *Learning and soft computing*. The MIT Press (2001).
- [Keila 2006] P. S. Keila, and D. B. Skillicorn: Structure in the Enron email dataset, *Journal of Computational & Mathematical Organization Theory* **11**, 183-199 (2006).
- [Keim 2002] D. A. Keim: Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics* **8**, 1-8 (2002).
- [Keim 1996] D. A. Keim, and H. P. Kriegel: Visualization techniques for mining large databases: a comparison, *IEEE Transactions on Knowledge and Data Engineering* **8**, 923-938 (1996).
- [Klerks 2002] P. Klerks: The network paradigm applied to criminal organizations, *Connections* **24**, 53-65 (2002).
- [Kohonen 1990] T. Kohonen: The self-organizing map, *Proceedings of the IEEE* **78**, 1464-1480 (1990).

- [Kostoff 2001] R. N. Kostoff: Science and technology roadmaps, *IEEE Transactions on Engineering Management* **48**, 132-139 (2001).
- [Krackhardt 1987] D. Krackhardt: Cognitive social structures, *Social Networks* **9**, 109-134 (1987).
- [Krebs 2002] V. E. Krebs: Mapping networks of terrorist cells, *Connections* **24**, 43-52 (2002).
- [Kaban 2001] A. Kaban, and M. Girolami: A combined latent class and trait model for the analysis and visualization of discrete data, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 859-872 (2001).
- [Kwok 2002] R. Kwok, J. Ma, and D. Zhou: Improving group decision making: a fuzzy GSS approach, *IEEE Transactions on Systems, Man, & Cybernetics Part C* **32**, 54-63 (2002).
- [Lai 2006] P. L. Lai, and C. A. Fyfe: A latent variable implementation of canonical correlation analysis for data visualisation, *Proceeding of the International Joint Conference on Neural Networks*, Honolulu, 1143-1149 (2006).
- [Lazega 1999] E. Lazega, and P. Pattison: Multiplexity, generalized exchange and cooperation in organizations, *Social Networks* **21**, 67-90 (1999).
- [Lee 2005] J. S. Lee, M. C. Zhou, and P. L. Hsu: An application of Petri net to supervisory control for human-computer interactive systems, *IEEE Transactions on Industrial Electronics* **52**, 1220 -1226 (2005).
- [Liben-Nowell 2004] D. Liben-Nowell, and J. Kleinberg: The link prediction problem for social networks, *Proceedings of the Annual ACM International Conference on Information & Knowledge Management*, New York (2004).
- [Mase 2004] H. Mase: Two-stage patent retrieval method considering claim structure, *Proceedings of the NTCIR-4 Workshop*, Tokyo (2004).
- [Matsumura 2001] N. Matsumura, Y. Ohsawa, and M. Ishizuka: Discovery of emerging topics by co-citation graph on the web, *Lecture Notes in Computer Science* **2198**, 473-481 (2001).
- [Memon 2006] N. Memon, H. L. Larsen: Practical approaches for analysis, visualization and destabilizing terrorist networks, *Proceedings of the International Conference on Availability, Reliability and Security*, Vienna (2006).



- [Menezes 2004] M. A. de Menezes, A. L. Barabási: Separating the internal and external dynamics of complex systems: *Physical Review Letters* **93**, 068701 (2004).
- [Menezes 2004] M. A. de Menezes, and A. L. Barabási: Fluctuations in network dynamics, *Physical Review Letters* **92**, 028701 (2004).
- [Mojsilovic 2005] A. Mojsilovic: A logistic regression model for small sample classification problems with hidden variables and non-linear relationships: an application in business analytics, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, V/329 - V/332 (2005).
- [Morselli 2007] C. Morselli, C. Giguere, and K. Petit: The efficiency/security trade-off in criminal networks, *Social Networks* **29**, 143-153 (2007).
- [Mukherjea 2005] S. Mukherjea, B. Bamba, and P. Kankar: Information retrieval and knowledge discovery utilizing a biomedical patent semantic Web, *IEEE Transactions on Knowledge and Data Engineering* **17**, 1099 - 1110 (2005).
- [Murata 2003] T. Murata: Visualizing the structure of web communities based on data acquired from a search engine, *IEEE Transactions on Industrial Electronics* **50**, 860-866 (2003).
- [Nabney 2005] I. T. Nabney, Y. Sun, P. Tino, and A. Kaban: Semisupervised learning of hierarchical latent trait model for data visualization, *IEEE Transactions on Knowledge and Data Engineering* **17**, 384-400 (2005).
- [Newman 2006] M. E. J. Newman: Modularity and community structure in networks, *Proceedings of the National Academy of Science USA* **103**, 8577-8582 (2006).
- [Newman 2003a] M. E. J. Newman, and J. Park: Why social networks are different from other types of networks, *Physical Review E* **68**, 036122 (2003).
- [Newman 2003b] M. E. J. Newman: Structure and function of complex networks, *SIAM Review* **45**, 167-256 (2003).
- [Newman 2002] M. E. J. Newman: Assortative mixing in networks, *Physical Review Letter* **89**, 208701 (2002).
- [Newman 2001a] M. E. J. Newman, S. H. Strogatz, and D. J. Watts: Random graphs with arbitrary degree distributions and their applications, *Physical Review E* **64**, 026118 (2001).

- [Newman 2001b] M. E. J. Newman: The structure of scientific collaboration networks, *Proceedings of the National Academy of Science USA* **98**, 404-409 (2001).
- [Ohsawa 2006] Y. Ohsawa eds.: *Chance discovery in real world decision making*. Springer-Verlag (2006).
- [Ohsawa 2005] Y. Ohsawa: Data crystallization: chance discovery extended for dealing with unobservable events, *New Mathematics and Natural Computation* **1**, 373-392 (2005).
- [Ohsawa 2003a] Y. Ohsawa, and P. McBurney eds.: *Chance discovery (Advanced information processing)*. Springer-Verlag (2003).
- [Oja 1991] E. Oja, and K. Kiviluoto: Data analysis, visualization, and hidden factor discovery by unsupervised learning, *Proceedings of the International Joint Conference on Neural Networks*, Washington, D.C., 3871-3876 (1999).
- [Ozidal 2004] M. M. Ozidal, and C. Cevdet: Hypergraph models and algorithms for data-pattern-based clustering, *Data Mining and Knowledge Discovery* **9**, 29-57 (2004).
- [Pan 2007] S. M. Pan, and K. S. Cheng: Evolution-based tabu search approach to automatic clustering, *IEEE Transactions on Systems, Man & Cybernetics* **37**, 828-838 (2007).
- [Pattison 2002] P. Pattison, and G. L. Robins: Neighborhood-based models for social networks, *Sociological Methodology* **32**, 301-337 (2002).
- [Pattison 1999] P. Pattison, and S. Wasserman: Logit models and logistic regressions for social networks, *British Journal of Mathematical and Statistical Psychology* **52**, 169-194 (1999).
- [Popp 2006] R. L. Popp, and J. Yen: *Emergent information technologies and enabling policies for counter- terrorism*. IEEE Press (2006).
- [Potlapalli 1996] H. Potlapalli, and R. C. Luo: Projection learning for self-organizing neural networks, *IEEE Transactions on Industrial Electronics* **43**, 485-491 (1996).
- [Roberts 2001] S. Roberts, C. Holmes, and D. Denison: Minimum-entropy data partitioning using reversible jump Markov chain Monte Carlo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 909-914

- [Robins 2001] G. L. Robins, P. E. Pattison, and P. Elliott: Network models for social influence processes, *Psychometrika* **66**, 161-190 (2001).
- [Robins 1999] G. L. Robins, P. E. Pattison, and S. Wasserman: Logit models and logistic regressions for social networks, *Psychometrika* **64**, 371-394 (1999).
- [Rush 1997] R. Rush: Elicitation of knowledge from multiple experts using network Interface, *IEEE Transactions on Knowledge and Data Engineering* **9**, 688-696 (1997).
- [Sageman 2004] M. Sageman: Understanding terror networks, University of Pennsylvania Press (2004).
- [Schön 2006] D. A. Schön: The reflective practitioner: How professionals think in action. Basic Books (2006).
- [Schuetze 1998] H. Schuetze: Automatic word sense discrimination, *Computational Linguistics* **24**, 97-123 (1998).
- [Shafti 2005] L. S. Shafti, and E. P. Perez: Constructive induction and genetic algorithms for learning concepts with complex interaction, *Proceedings of the Conference on Genetic and Evolutionary Computation*, Washington DC, 1811-1818 (2005).
- [Shen 2006] Z. Shen, K. L. Ma, and T. Eliassi-Rad: Visual analysis of large heterogeneous social networks by semantic and structural abstraction, *IEEE Transactions on Visualization and Computer Graphics* **12**, 1427-1439 (2006).
- [Silva 2006] R. Silva, R. Scheines, C. Glymour, and P. Spirtes: Learning the structure of linear latent variable models, *J. Machine Learning Res.* **7**, 191-246 (2006).
- [Silva 2003] R. Silva, R. Scheines, C. Glymour, and P. Spirtes: Learning measurement models for unobserved variables, *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence*, Acapulco, 543-55 (2003).
- [Singh 2004] S. Singh, J. Allanach, T. Haiying, K. Pattipati, and P. Willett: Stochastic modeling of a terrorist event via the ASAM system, *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, Hague, 6/5673-6/5678 (2004).
- [Stauffer 2003] D. Stauffer: Sociophysics simulations, *IEEE Computing in Science & Engineering* May/June, 71-75 (2003).

- [Sugiyama 2002] K. Sugiyama: Graph drawing and applications for software and knowledge engineers (Series on software engineering and knowledge engineering 11). World Scientific Publishing (2002).
- [Suzuki 2005] E. Suzuki, and J. M. Zytkow: Unified algorithm for undirected discovery of exception rules, *International Journal of Intelligent Systems* **20**, 673-691 (2005).
- [Taskar 2003] B. Taskar, M. F. Wong, P. Abbeel and D. Koller: Link prediction in relational data, *Proceedings of the Neural Information Processing Systems Conference, Vancouver* (2003).
- [Taskar 2002] B. Taskar, P. Abbeel and D. Koller: Discriminative probabilistic models for relational data, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence, Edmonton* (2002).
- [Thurlow 2004] C. Thurlow, L. Lengel, and A. Tomic: Computer mediated communication. Sage Publications Ltd. (2004).
- [Tian 2001] F. Tian, H. Zhang, Y. Lu, and C. Shi: Incremental learning of Bayesian networks with hidden variables, *Proceedings of the IEEE International Conference on Data Mining, San Jose*, 651-652 (2001).
- [Tino 2002] P. Tino, and I. Nabney: Hierarchical GTM: constructing localized nonlinear projection manifolds in a principled way, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, 639-656 (2002).
- [Tory 2004] M. Tory, and T. Moeller: Human factors in visualization research, *IEEE Transactions on Visualization and Computer Graphics* **10**, 72-84 (2004).
- [Tsuruoka 2005] Y. Tsuruoka, and J. Tsujii: Bidirectional inference with the easiest-first strategy for tagging sequence data, *Proceedings of the Human Language Technology Conference, and the Conference on Empirical Methods in Natural Language Processing, Vancouver*, 467-474 (2005).
- [Uchida 2004] H. Uchida: Patent map generation using concept-based vector space model, *Proceedings of the NTCIR-4 Workshop, Tokyo* (2004).
- [Watts 1998] D. J. Watts, and S. H. Strogatz: Collective dynamics of small-world networks, *Nature* **398**, 440-442 (1998).
- [Weiss 1998] G. M. Weiss, and H. Hirsh: Learning to predict rare events in event sequences, *Proceedings of the International Conference on Knowledge Discovery and Data Mining, New York* (1998).

- [Xie 2003] L. Xie, Q. L. Zhang, M. Guo, and S. Q. Wang: Linear pruning techniques for neural networks- based on projection latent structure, IEEE International Conference on Systems, Man and Cybernetics, Washington, D.C., 1304-1309 (2003).
- [Yook 2001] S. H. Yook, H. Jeong, A. L. Barabási, and Y. Tu: Weighted evolving networks, Physical Review Letters **86**, 5835-5838 (2001).
- [Yoon 2007] B. Yoon, and Y. Park: Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information, IEEE Transactions on Engineering Management **54**, 588-599 (2007).
- [Yuan 2005] C. Yuan, C. Neubauer, Z. Cataltepe, and H. G. Brummel: Support vector methods and use of hidden variables for power plant monitoring, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, V/693-V/696 (2005).
- [Zaltman 2003] G. Zaltman: How customers think: Essential insights into the mind of the market. Harvard Business School Press (2003).
- [Zhang 2004] N. L. Zhang, and T. Kocka: Efficient learning of hierarchical latent class models, Proceedings of the IEEE International Conference on Tools with Artificial Intelligence, Boca Raton, 585-593 (2004).
- [Zivkovic 2004] Z. Zivkovic, and F. van der Heijden: Recursive unsupervised learning of finite mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence **26**, 651-656 (2004).
- [Zoeter 2003] O. Zoeter, and T. Heskes: Hierarchical visualization of time-series data using switching linear dynamical systems, IEEE Transactions on Pattern Analysis and Machine Intelligence **25**, 1202-1214 (2003).
- [浅田 2005] 浅田 洋平, 松尾 豊, 石塚 満: Web からの研究者ネットワーク抽出の大規模化, 人工知能学会論文誌 **20**, 370-378 (2005).
- [石井 2004] 石井 成郎, 三輪 和久: プロセスの自己省察を軸とした創造性教育, 人工知能学会論文誌 **19**, 126 -135 (2004).
- [稲田 2002] 稲田 政則, 寺野 隆雄: 肝機能検査データからの因果モデルの構築, 人工知能学会論文誌 **17**, 708 -715 (2002).
- [井山 2004] 井山 晃洋: 多角的な話題の収集を目的とした話題の独自性に基づく Web ページの分類システム, 人工知能 学会論文誌 **19**, 561-568 (2004).

- [大澤 1999] 大澤 幸生, ネルス E. ベンソン, 谷内田 正彦: KeyGraph : 語の共起グラフの分割・統合によるキーワード抽出, 電子情報通信学会論文誌 18, 333-344 (2003).
- [尾崎 2004] 尾崎 圭司, 大倉 隆司, 野口 達毅, 大川 人資, 矢野 米雄: 観点の共有と可視化が可能な仮想美術館システム, 人工知能学会論文誌 19, 561-568 (2004).
- [鹿島 2007] 鹿島 久嗣, 安倍 直樹: ネットワーク構造の確率的な時変モデルに基づく教師ありリンク学習, 人工知能学会論文誌 22, 209-217 (2007).
- [川面 2003] 川面 恵司: 創造的工学設計の方法. 養賢堂 (2003).
- [河原 2006] 河原 吉伸, 矢入 健久, 町田 和雄: Dynamic Bayesian Networks を用いた宇宙機異常診断法, 人工知能学会論文誌 21, 45-54 (2006).
- [キッスイリクン 1992] ブンサーム・キッスイリクン, 沼尾正行, 志村正道: 弁別に基づく構成的帰納学習, 人工知能学会誌 7, 1027-1037 (1992).
- [金 2007] 金 英子, 松尾 豊, 石塚 満: Web 上の情報を用いた企業間関係の抽出, 人工知能学会論文誌 22, 48-57 (2007).
- [産業能率大学 2003] 産業能率大学 CPM・TRIZ 研究会: TRIZ の理論とその展開. 産業能率大学出版部 (2003).
- [砂山 1999] 砂山 渡, 大澤 幸生, 谷内田 正彦: 事象毎の生起確率から未知事象発見を支援する手法とそのアンケート調査への適用, 人工知能学会誌 14, 349-358 (1999).
- [諏訪 2004] 諏訪正樹: 「創造」の研究: 現象を生む実践の学, 人工知能学会誌 19, 205-213 (2004).
- [諏訪 2005] 諏訪正樹: シナリオ創発の認知科学的裏づけ, 人工知能学会誌 20, 34-39 (2005).
- [田中 2006] 田中 貴紘, 片上 大輔, 新田 克己: 調停トレーニングにおけるアドバイザーエージェント, 人工知能学会論文誌 21, 319-329 (2006).
- [西原 2004] 西原 陽子: 有効な組合せの発見による創造活動支援, 電子情報通信学会論文誌 J87D, 939-946 (2004).
- [長谷川 2007] 長谷川 禎彦, 伊庭 斉志: ベイジアンネットワーク推定による確率モデル遺伝的プログラミング, 人工知能学会論文誌 22, 37-47 (2007).
- [松尾 2006] 松尾 豊, 篠田 孝祐, 中島 秀之: 中心性に着目した合理エージェントのネットワーク形成, 人工知能学会論文誌 21, 122-132 (2006).

- [松尾 2005] 松尾 豊, 友部 博教, 橋田 浩一, 中島 秀之, 石塚 満: Web 上の情報から人間関係ネットワークの抽出, 人工知能学会論文誌 20, 46-56 (2005).
- [松村 2003] 松村 真宏, 大澤 幸生, 石塚 満: 影響の普及モデルに基づくオンラインコミュニティ参加者のプロファイリング, 人工知能学会論文誌 18, 165-172 (2003).
- [松村 2002] 松村 真宏, 大澤 幸生, 石塚 満: 語の活性度に基づくキーワード抽出法, 人工知能学会論文誌 17, 398-406 (2002).
- [マン 2004] ダレル・マン: TRIZ 実践と効用 - 体系的技術革新. 創造開発イニシアチブ (2004).
- [三浦 2003] 三浦佳世: 感性印象の知覚的・認知的基盤, VISION 15, 143-149 (2003).
- [宮川 1997] 宮川 雅巳: グラフィカルモデリング. 朝倉書店 (1997).
- [本村 2002] 本村 陽一: ベイジアンネットソフトウェア, 人工知能学会誌 17, 559-565 (2002).

## A 最尤法による解法

1章で提起したノート発見問題に対して、3章では、ヒューリスティックなアルゴリズムを用いたヒューマンインタラクティブアニーリングの手法を提案した。複雑ネットワークにおけるノート発見問題に対して、最尤法 [Hastie 2001], [Duda 2000] を応用して、より厳密なアプローチを採用することもできる。最尤法を応用した手法では、以下の2つの手順を用いる。

1. 観測データから、尤度関数を用いたネットワーク上のすべてのリンク構造の推定
2. 推定したリンク構造から、ランキング関数によるノードの発見

### A.1 観測モデル

観測データ全体を  $D$  と表す。個々の観測データ ( $i$  番目) を  $d_i$  と表す。観測データは、式 (46) で定義されたノード  $n_j$  の集合である。個々の観測データは、独立であると仮定する。観測データの数を  $|d|$  とする。

$$D = \{d_i\}, d_i = \{n_j\} \quad (0 \leq i \leq |d| - 1). \quad (46)$$

観測  $d_i$  は、集合の形式だけでなく、ベクトルの形式でも表現できる。観測  $d_i$  にノード  $n_j$  が含まれるかどうかを式 (47) で表す。ノードの種類数を  $|n|$  とする。

$$d_{ij} = \begin{cases} 1 & \text{if } n_j \in d_i \\ 0 & \text{otherwise} \end{cases} \quad (0 \leq j \leq |n| - 1). \quad (47)$$

観測データを生成するネットワーク上のメカニズムとして、単純な刺激 - 反応モデル (single stimulator and multiple responders model) を仮定する。確率  $f_j$  は、ノード  $n_j$  が刺激を発生される確率である。

$$0 \leq f_j \leq 1. \quad (48)$$

$$\sum_{j=0}^{|n|-1} f_j = 1. \quad (49)$$



ここでは，単純に，確率  $f_j$  はノードに依存せず，どのノードも同じ確率で刺激を発生させるものとする．

$$f_j = \frac{1}{|n|} \quad (0 \leq j \leq |n| - 1). \quad (50)$$

ノード  $n_i$  から responder ノード  $n_j$  へ刺激が伝播する確率を  $r_{ij}$  とする．刺激は，複数の responder に向かって平行に，独立に伝播する．

$$0 \leq r_{ij} \leq 1. \quad (51)$$

$$r_{ii} = 1. \quad (52)$$

以降の計算を簡単化するために，刺激の伝播は，式 (53) のような双方向同一確率であると仮定する．

$$r_{ij} = r_{ji}. \quad (53)$$

確率  $r_{ij}$  がネットワークのリンク構造を指定する変数である．これらの変数をすべてまとめて， $\mathbf{r}$  と表記する．

$$\mathbf{r} = \{r_{ij}\} \quad (0 \leq i, j \leq |n| - 1). \quad (54)$$

## A.2 尤度とランキング

変数  $\mathbf{r}$  に対して，観測データ  $D$  が得られる確率から，対数尤度関数は式 (55) で与えられる．

$$L(\mathbf{r}) = \log p(D|\mathbf{r}) = \log \prod_{i=0}^{|d|-1} p(d_i|\mathbf{r}) = \sum_{i=0}^{|d|-1} \log p(d_i|\mathbf{r}). \quad (55)$$

Stimulator ノード  $n_j$  への responder ノードとして  $n_k$  が含まれるかどうか，観測データ  $d_i$  と合致する確率を  $f_{jk}$  とする．

$$f_{jk} = \begin{cases} r_{jk} & \text{if } d_{ik} = 1 \text{ for given } j \text{ and } i \\ 1 - r_{jk} & \text{otherwise} \end{cases} \quad (56)$$

$$f_{jk} = d_{ik}r_{jk} + (1 - d_{ik})(1 - r_{jk}). \quad (57)$$

式 (55) に含まれる条件付確率  $p(d_i|\mathbf{r})$  は，式 (58) で与えられる．

$$p(d_i|\mathbf{r}) = \sum_{j=0}^{|n|-1} [d_{ij} \times f_j \prod_{(0 \leq k \leq |n|-1) \wedge (k \neq j)} \{d_{ik}r_{jk} + (1 - d_{ik})(1 - r_{jk})\}]. \quad (58)$$

$$p(d_i|\mathbf{r}) = \sum_{j=0}^{|n|-1} [d_{ij} \times f_j \prod_{k=0}^{|n|-1} \{1 - d_{ik} + (2d_{ik} - 1)r_{jk}\}]. \quad (59)$$

式 (55) は，式 (60) となる．

$$L(\mathbf{r}) = \sum_{i=0}^{|d|-1} \log \left( \sum_{j=0}^{|n|-1} [d_{ij} \times f_j \prod_{k=0}^{|n|-1} \{1 - d_{ik} + (2d_{ik} - 1)r_{jk}\}] \right). \quad (60)$$

対数尤度関数を最大化する変数  $\hat{\mathbf{r}}$  は，式 (61) から得られる．

$$\hat{\mathbf{r}} = \arg \max_{\mathbf{r}} L(\mathbf{r}). \quad (61)$$

式 (61) を解くのに，単純な山登り法を用いる．山を登る方向  $\Delta r_{nm}$  は，式 (62) で与えられる尤度関数の微分係数の方向 (gradient) と一致した，最も急な上り勾配の方向でなければならない．

$$\Delta L(\mathbf{r}) = \sum_{0 \leq n, m \leq |n|-1} \frac{\partial L(\mathbf{r})}{\partial r_{nm}} \Delta r_{nm}. \quad (62)$$

式 (52) と式 (53) を用いて，式 (62) は式 (63) となる．

$$\Delta L(\mathbf{r}) = \sum_{(0 \leq n \leq |n|-1) \wedge (m > n)} \left( \frac{\partial L(\mathbf{r})}{\partial r_{nm}} + \frac{\partial L(\mathbf{r})}{\partial r_{mn}} \right) \Delta r_{nm}. \quad (63)$$

ここの微分係数は式 (64) で計算することができる．

$$\frac{\partial L(\mathbf{r})}{\partial r_{nm}} = \sum_{i=0}^{|d|-1} \frac{f_n d_{in} (2d_{im} - 1) \prod_{k \neq m} \{1 - d_{ik} + (2d_{ik} - 1)r_{nk}\}}{\sum_{j=0}^{|n|-1} [d_{ij} \times f_j \prod_{k=0}^{|n|-1} \{1 - d_{ik} + (2d_{ik} - 1)r_{jk}\}]}]. \quad (64)$$

このようにして推定したリンク構造  $\hat{\mathbf{r}}$  を入力として，ノード発見を行うランキング関数を導入する．見えないノードを含む可能性が高いデータ，見えないノードの近くにある可能性が高いリンク，見えないノードに接続している可能性が高いノードを取り出すための評価関数がランキン

グ関数である．見えないノードを含む可能性が高いデータに高い得点を出力するためのランキング関数は，式 (65) である．関数  $g(x)$  は， $x$  の単調減少関数である．推定したリンク構造  $\hat{r}$  から発生しにくい観測データが，見えないノードを含む可能性が高い．

$$s_d(d_i) \propto g(p(d_i|\hat{r})). \quad (65)$$

同様に，見えないノードの近くにある可能性が高いリンクに高い得点を出力するためのランキング関数は，式 (66) である．推定したリンク構造  $\hat{r}$  が，存在  $r_{ij} = 1$  も不存在  $r_{ij} = 0$  も示さないリンクが，見えないノードの近くにある可能性が高い．

$$s_r(r_{ij}) \propto g(|\hat{r}_{ij} - 0.5|). \quad (66)$$

見えないノードに接続している可能性が高いノードに高い得点を出力するためのランキング関数は，式 (67) である．たくさんの，見えないノードの近くにある可能性が高いリンクに接続したノードが，見えないノードに接続している可能性が高い．

$$s_n(n_i) \propto \sum_{(0 \leq j \leq |n|-1) \wedge (j \neq i)} s_r(r_{ij}). \quad (67)$$

### A.3 単純な例と応用展開

A.2 の手法を単純なネットワークに適用した例を示す．

図 41 は，見えないノードを含まない場合 (左側 case 1) と含む場合 (右側 case 2 の  $n_5$  が見えないノード) との例を示す．Case 1 では，リンク構造を正確に推定できた．ランキング関数を適用すれば，データが見えないノードを含む可能性，リンクが見えないノードの近くにある可能性，いずれも低いことがわかる．Case 2 では，リンク構造を推定した後，見えないノードの近くにある可能性が最も高いリンクは， $r_{41}$  であることがわかる．実際に，見えないノード  $n_5$  は， $n_1$  と  $n_4$  との間であり，この推測は正しい．

ここで述べたネットワーク上の見えないノード発見問題とその解法は，より一般的な問題へ拡張できる可能性がある．ノード発見問題には，以下の 3 つのパターンがある．今後，尤度関数とランキング関数を組み合

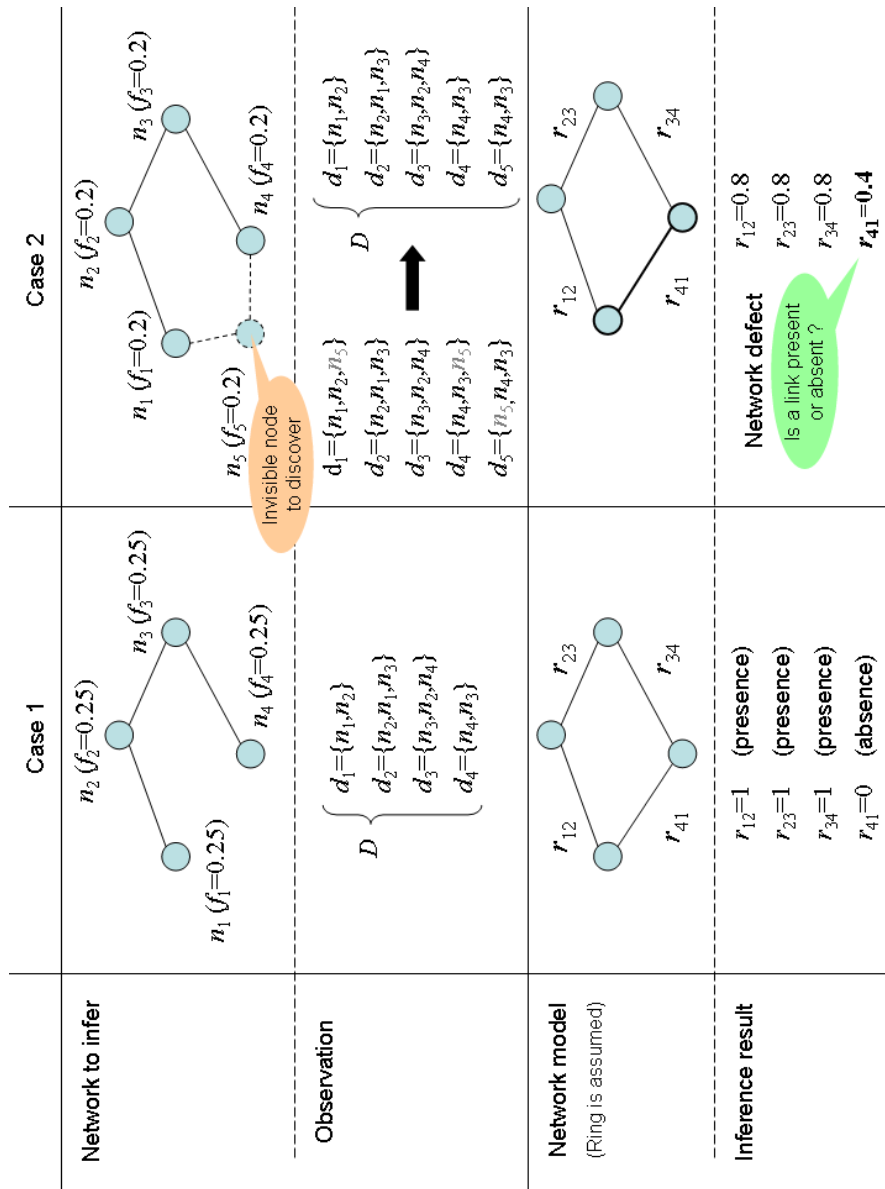


図 41: 見えないノードを含まない場合 (左側 case 1) と含む場合 (右側 case 2 の  $n_5$  が見えないノード) とにおける, ネットワーク上でのノード発見の簡単な例. Case 2 では, リンク構造を推定した後, 見えないノードの近くにある可能性が最も高いリンクは,  $r_{41}$  であることがわかる.

わせて、観測と構造の不整合を特定することで、偽装ノードやなりすましノードの発見に活用できると期待している。

1. 見えないノード発見：ネットワークには含まれるが、観測データには含まれない、見えないノードを発見する問題
2. 偽装ノード発見：観測データには含まれるが、ネットワークには存在しない、偽装ノードを発見する問題
3. なりすましノード発見：ネットワークに含まれるが、異なるノードとして観測データに含まれている、なりすましノードを発見する問題

## B テキスト解析・発想支援への応用

3章で述べたヒューマンインタラクティブアニメーションは、ヒューマンネットワークだけでなく、テキストにも応用することができる。代表的な応用例として、特許テキストを解析して技術開発アイデアの発想支援 [Horie 2007] に役立てる例について述べる。

特許は、特定の工学的デザイン問題を解決するための技術的な手段を表している。既知の公開された特許文献を出発点として、未知の重要な技術要素についての着想を得たり、ライバル企業が意図的に隠蔽している技術要素（サブマリン特許など）を探ったり、異業種企業、ニッチ企業などが潜在的に保有している技術を掘り起こしたりするための手がかりを得ることを目指す。これらの技術要素を他の企業が先に手に入れることは、研究開発における大きな脅威である。しかし、これらのリスクに気づき、技術開発を先行すれば、脅威を機会に転じることが期待できる。

### B.1 特許からの発想支援

グループ討議などで発想支援を行う場合を想定する。既存の特許からシナリオマップを作成し、ダミー事象が表す新しい技術要素に着目する。そこから、より大きな技術体系に広がる可能性がないか、シナリオの創出を検討することに役立てられる。

知識獲得についての公開済みの23件の米国特許<sup>1</sup>を用いて、新しい技術要素に関するシナリオの創出を試みた。特許をUS Classに基づいて分類し、それぞれをバスケットとした。サブジェクト = US class : { 事象 = 特許番号 } となる。US class は、木構造の技術カテゴリーを表すコード体系で、すべての米国特許に1つ以上のUS class が付与されている。作成したシナリオマップを図42に示す。5つのクラスタの間に、5つのダミー事象が表出した。複数のダミー事象が表出したクラスタ2と3に注目する。ダミー事象DE<sub>14</sub>は、US class では、706/50: knowledge processing system/having specific management of a knowledge に対応する。ダミー事象DE<sub>18</sub>は、US class では、706/59: knowledge processing system/creation or modification に対応する。

---

<sup>1</sup>アメリカ特許商標庁 United States Patent and Trademark Office (USPTO), Patent Electronic Business Center, Patent Full-Text and Full-Page Image Databases <http://www.uspto.gov/patft/index.html>

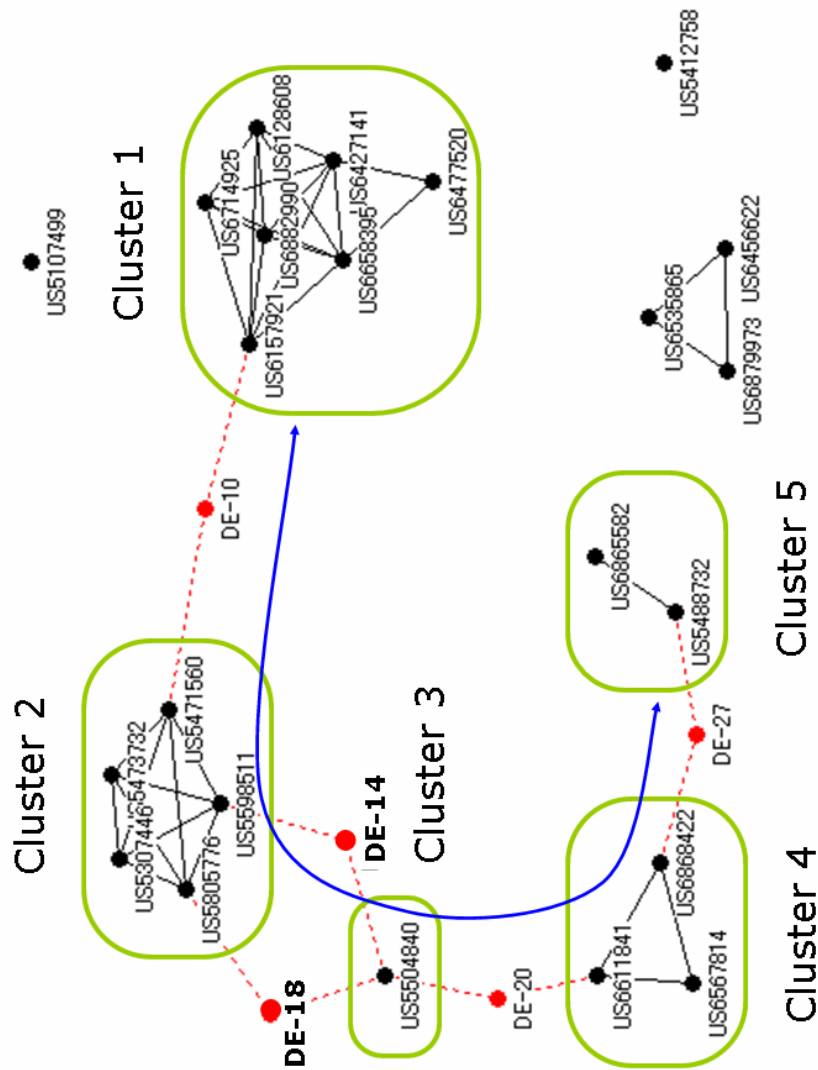


図 42: 知識獲得についての公開済みの 23 件の米国特許を用いて, US Class に基づいた分類から生成したシナリオマップ. ダミー事象  $DE_{14}$  は, US class では, 706/50: knowledge processing system/having specific management of a knowledge に対応する. ダミー事象  $DE_{18}$  は, US class では, 706/59: knowledge processing system/creation or modification に対応する.

シナリオの創出を検討した結果，ダミー事象  $DE_{14}$ ,  $DE_{18}$  でつながった，クラスタ  $c_2$  の特許 5805776 号 (名称 Device for automatic generation of a knowledge base for a diagnostic expert system (1998))，5598511 号 (名称 Method and apparatus for interpreting data and accessing on-line documentation in a computer system (1997)) とクラスタ  $c_3$  の特許 5504840 号 (名称 Knowledge acquisition support system and method in an expert system (1996)) とに着目したと仮定して，B.2 の特許テキストの解析を進める．

## B.2 特許テキストの解析

解析するテキストの例として，特許のテキストに含まれる請求項 (claim) を採り上げる．請求項とは，その特許の権利範囲を表し，最も直接的に，問題を解決するためのユニークな手段を記述している．特許 5805776，5598511 と特許 5504840 の請求項のテキストを請求項ごとに形態素解析して，ヒューマンインタラクティブアニーリングを行い，シナリオマップを作成した．例えば，特許 5805776 の第一の請求項の原文は，以下のようになっている．

*A device for building the knowledge base of an expert system, which diagnoses a technical system comprised of modules, the device comprising: computer (Kb-builder) having reading access to a first memory (KM) and to a second memory (Conf) as well as having at least intermittent writing access to a third memory (Kb), and information about the technical system, about its malfunctions and about its diagnosis, the information being stored in the first memory (KM), the information comprising: for the type to which the technical system belongs, a knowledge module, which contains all necessary information about the internal design of a technical system of this type as well as information, when needed, about the malfunctions, remedies, and tests in a technical system of this type, and for each type of module, which occurs at least once in the technical system, a respective knowledge module, which contains all necessary information about the internal design of a module of this type as well as information, when needed, about the malfunctions, remedies, tests, and results in a module of this type, wherein in the knowledge module for the type to which the technical system belongs, information is contained as to which roles, that*



is, which subfunctions must be filled by an assignment of a module, that is, which subfunctions must received an assignment of a module, so that the technical system can fulfill its standard function, wherein for each type of composite modules that occur in a technical system of the type, information is contained in the knowledge module for this module type as to which roles must be filled in a module of this module type so that the module can fulfill its standard function, wherein for each role in the technical system and each role in each composite module in the technical system, information is available as to which module fills the role in the technical system or in the composite module, wherein the configuration of the technical system is stored in the second memory (Conf), and wherein in the third memory (Kb), a knowledge module for the technical system is created as a copy of the knowledge module for the type to which the technical system belongs, for each module in the technical system, a separate knowledge module is contained as a copy of the knowledge module for the type to which the module belongs, and wherein each thus-formed knowledge module for a module corresponding to the role that the module plays in the composite module in the technical system is combined with the thus-formed knowledge module for the composite module or for the technical system.

ヒューマンインタラクティブアニメーションへの入力データは、 $b_1 = \{\text{device, build, knowledge, base, expert, system, diagnose, ...}\}$  といったデータとなる。得られたシナリオマップを図 43 に示す。

10 個のクラスター（単語群）と 33 個のダミー事象が可視化された。この中で、クラスター  $c_5$  は、このクラスターを中心としてクラスター  $c_1, c_2, c_3, c_4, c_6, c_8$  をつなぐ重要な役割を担っているようだ。クラスター  $c_9, c_{10}$  ように孤立したクラスターもある。まず、クラスターの表す技術要素を解釈した。技術要素を表すキーワードをピックアップし、表 16 にまとめた。

多数のダミー事象が表出している領域（クラスター間の隙間）に注目しながら、技術要素に関するシナリオを創出する。クラスター  $c_5$  を軸に 2 つのシナリオを創出した。シナリオ 1 は、クラスター  $c_1, c_2, c_3, c_5, c_8$  を結びつけ、ダミー事象群の領域 1 (dense dark events 1) の実体として、「専門家の知見の導入すること」を想起した。シナリオ 2 は、クラスター  $c_4, c_5, c_6, c_7$  を結びつけ、ダミー事象群の領域 2 (dense dark events 2) の実体として、「知識を複数の観点から比較検証すること」を想起した。このよう



Cluster	Interpretation (keywords)
$c_2$	correct, check, logic, query
$c_3$	interpret, execute, apply, implement, rule, user
$c_4$	support, store, significance, magnitude
$c_5$	replace, compare, select, produce, repeat, fact
$c_6$	learn, infer, identify, reason, text, document, multimedia, graphics
$c_7$	diagnose, extract, frequency, occurrence, rate, cost, knowledge base

表 16: シナリオマップに現れたクラスター  $c_2, c_3, c_4, c_5, c_6, c_7$  の解釈で得られた、クラスターを表すキーワード。

に、ダミー事象の実体を想起しながら、全体を接続するシナリオを探ることで、新しい発想が出てくることがわかる。特許の目的や分類から獲得した関心をもとに、テキストや他のデータを活用して、発想支援を行う可能性を示した。

1. Enhance learning by replacing facts with the aid of interpretation of queries on *expert knowledge* (= dense dark events 1).
2. Assign significance to knowledge inferred from multiple information (texts, documents, multimedia, graphics, etc. based on diagnosis by *cross-validating knowledge* (= dense dark events 2).

この事例は、ごく簡単なものである。しかし、特許に基づいて、技術開発の発想支援を行い、未知の技術要素に起因する脅威を機会に転じる方法の確立のきっかけになるだろう。複数の観測データを統合し、関心に応じて、視点を変更することの有効性も示唆している。TRIZ [マン 2004]、価値工学、Taguchi method などは、既存の成功事例をもとに、コストや品質の制約下で技術要素の組み合わせを最適化 [産業能率大学 2003] するものと言える。ヒューマンインタラクティブアニーリングを応用した手法では、ダミー事象に対応する未知の事象を探りながら、シナリオを練っていく点で、既存の方法とは異なる。交渉ログ [田中 2006] といったさまざまなテキストの解析、マーケティングなどのアイデア創出支援の面においても、本研究が取り組んだ見えない構造を発見する手法が応用されていくことが期待できる。

## C 公表済みの研究成果

### C.1 論文投稿

1. Y. Maeno, and Y. Ohsawa: Human-computer interactive annealing for discovering invisible dark events, *IEEE Transactions on Industrial Electronics* **54**, 1184-1192 (2007).
2. 前野 義晴, 大澤 幸生: コミュニケーションから探る組織の見える黒幕, *人工知能学会論文誌* **22**, 389-396 (2007).
3. Y. Maeno, and Y. Ohsawa: Node discovery problem for a social network, *Connections* 投稿 (*e-print* <http://arxiv.org/abs/0710.4975>).
4. 前野 義晴, 大澤 幸生: アートの好みを視覚化・言語化して磨く自分への理解, *情報処理学会論文誌* 投稿.
5. Y. Maeno, and Y. Ohsawa: Predicting relevant empty spots in social interaction, *Journal of Systems Science and complexity* 投稿 (*e-print* <http://arxiv.org/abs/0711.1466>).
6. Y. Maeno, and Y. Ohsawa: Analyzing covert social network foundation behind terrorism disaster, *International Journal of Services Sciences* 投稿 (*e-print* <http://arxiv.org/abs/0710.4231>).

### C.2 出版

1. Y. Maeno, and Y. Ohsawa: Understanding of dark events for harnessing risk, *Chance Discoveries in Real World Decision Making (Studies in Computational Intelligence Series 30)*, 373-392, Springer-Verlag (2006).
2. Y. Maeno, K. Ito, and Y. Ohsawa: Detecting invisible relevant persons in a homogeneous social network, *Lecture Notes in Computer Science* **4490**, 74-81, Springer-Verlag (2007).

3. Y. Maeno, Y. Ohsawa, and T. Ito: Catalyst personality for fostering communication among groups with opposing preference, Lecture Notes in Artificial Intelligence **4570**, 806-812, Springer-Verlag (2007).
4. Y. Maeno, Y. Ohsawa, and T. Ito: Trigger to switch individual's interest toward unconscious preference, Lecture Notes in Artificial Intelligence **4693**, 970-977, Springer-Verlag (2007).

### C.3 学会投稿

1. 前野 義晴, 丹生 有紀, 伊藤 貴一, 大澤 幸生: Data Crystallization による隠れたノードの探知, 第 33 回日本行動計量学会, 長岡, 2005 年 8 月.
2. 前野 義晴, 大澤 幸生: データ結晶化の原理と応用, 第 5 回シナリオ創発ワークショップ, 東京, 2006 年 1 月.
3. Y. Maeno, and Y. Ohsawa: Crystallization highlighting hidden leaders, Information Processing and Management of Uncertainty in Knowledge-Based Systems, International Conference (IPMU), Paris, June 2006.
4. Y. Maeno, and Y. Ohsawa: Detection of threat activities hidden behind complicated relationship, JSAI International Workshop on Risk Mining, Tokyo, June 2006.
5. Y. Maeno, and Y. Ohsawa: Interactive annealing for discovering invisible dark events, International Workshop on Emergent Synthesis, Kashiwa, August 2006.
6. Y. Maeno, K. Horie, K. Ito, and Y. Ohsawa: Human-interactive annealing for multilateral observation, International Council for Science: Committee on Data for Science and Technology, International Conference (CODATA), Beijing, September 2006.
7. Y. Maeno, and Y. Ohsawa: Stable deterministic crystallization for discovering hidden hubs, IEEE International Conference on Systems, Man, and Cybernetics (SMC), Taipei, October 2006.

8. Y. Maeno, and Y. Ohsawa: Hidden structure visualization adaptive to human's prior understanding, Joint Conference on Information Science (JCIS), Kaohsiung, October 2006.
9. 前野 義晴, 堀江 健一, 伊藤 貴一, 大澤 幸生: 特許データに対する結晶化アルゴリズムによる技術イノベーション, 電子情報通信学会, 人工知能と知識処理研究会, 東京, 2006年12月.
10. Y. Maeno, K. Ito, K. Horie, and Y. Ohsawa: Human-interactive annealing for turning threat to opportunity in technology development, IEEE/WIC/ACM International Conference on Data Mining (ICDM), Workshop on Risk Mining, Hong Kong, December 2006.
11. 前野 義晴, 伊藤 貴一, 大澤 幸生: 無意識的な選好構造の可視化手法, 第6回シナリオ創発ワークショップ, 東京, 2007年1月.
12. Yoshiharu Maeno, Takaichi Ito, and Yukio Ohsawa: Detecting invisible relevant persons in a homogeneous social network, International Workshop on Meta-synthesis and Complex Systems (MCS), Beijing, May 2007.
13. Yoshiharu Maeno, Yukio Ohsawa, and Takaichi Ito: A node discovery problem for a social network, JSAI International Workshop on Risk Informatics, Miyazaki, June 2007.
14. Yoshiharu Maeno, Yukio Ohsawa, and Takaichi Ito: Catalyst personality for fostering communication among groups with opposing preference, International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE), Kyoto, June 2007.
15. Yoshiharu Maeno Yukio Ohsawa, and Takaichi Ito: Restoring missing network nodes featuring in relevance-to-visibility ratio, Joint Conferences on Information Science (JCIS), Salt Lake City, July 2007.
16. Yoshiharu Maeno, Yukio Ohsawa, and Takaichi Ito: Trigger to switch individual's interest toward unconscious preference, International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), Vietri sul Mare, September 2007.

17. 前野 義晴, 大澤 幸生: 複雑ネットワークにおけるノード発見問題, 第2回横断型基幹科学技術研究団体連合(横幹連合)コンファレンス, 京都, 2007年11月.
18. Yoshiharu Maeno and Yukio Ohsawa: Can hidden leaders be detected? - Data crystallization for simulated criminal groups, International Conference on Enhancement and Promotion of Computational Methods in Engineering and Science, Kyoto, December 2007.