

集団のオンラインテキストコミュニケーションにおける 発言関係性の自動構造化による生産性の評価

谷森 一貴¹ 齊藤 裕樹¹

概要：集団作業において、コミュニケーションは必要不可欠な要素である。また、コミュニケーションツールの多様化により、コミュニケーションの形態も変化してきており、チャットツールを使ったインフォーマルなオンラインテキストコミュニケーションも、アイデア創出の場としてチームの生産性に大きな影響をもたらすことが期待される。これまでのコミュニケーション研究において、組織のコミュニケーションの対象者間の関係性に基づき、個人間や組織間を評価することは多く行われてきたが、コミュニケーションそのものの概念の曖昧さから、コミュニケーションを測るものさしは確立されておらず、また、コミュニケーションの良否を決める明確な指標も解明されていない。そこで本論文では、発言間の関係性に着目して、チャットのコミュニケーションログからどの発言がどの発言に遷移したという発言間の構造を自動構築し、発言間構造に対して、複雑ネットワーク科学のアプローチを用いることで、チームの生産性に対する影響を評価する手法を提案する。構造的特徴解明のための実験を行った結果、クラスタ間の構造的な特徴に有意差が確認された。

1. はじめに

複数の人が一緒に1つの作業をする際に、作業進捗の報告やお互いの意見交換など様々な場面でコミュニケーションを図る。特にビジネスシーンでは、チームを取りまとめるリーダーが数人から数百人の集団を指揮して活動する形態が主流であり、関係性を築くためにコミュニケーションが重要な役割を担っている。これまでも組織のコミュニケーションは優れたチームワークを発揮する上で、重要な行動的役割を担うものであり、コミュニケーションのあり方がチームのポテンシャルを引き出すことが報告されている [1][2]。

また、近年ではコミュニケーションツールの普及に伴い、集団作業の中で LINE や Slack のようなチャットツールを使ったインフォーマルなテキストコミュニケーションの機会も増えている。集団組織の中でチャットがコミュニケーションツールとしての機能を果たした例もあり [3]、今後、ビジネスシーンなどでの活躍が見込まれる。その際、チームを運営するリーダーは、このようなインフォーマルに行われるコミュニケーションの良否を基にチームパフォーマンスを評価することが期待される。しかし、コミュニケーションそのものの概念の曖昧さから、コミュニケーションを測るものさしは確立されておらず、また、コミュニケー

ションの良否を決める明確な指標も解明されていない。

本研究の目的は、チームコミュニケーションにおける拡散や集約、冗長といった構造的な特徴を解明し、それがチームの生産性に影響をもたらすことを定量的に評価することである。そこで、第一に発言間の関係性に着目して、チャットのコミュニケーションログからどの発言がどの発言に遷移したという発言間の構造を自動構築する手法を提案する。第二に、発言ユーザを考慮した発言間構造に対して、複雑ネットワーク科学のアプローチを用いることで、チームの生産性に対する影響を評価する。構造的な特徴を明らかにする実験の結果、コミュニケーションの構造的な特徴とチームの生産性に統計的な因果関係は確認されなかったが、クラスタ間の構造的な特徴に有意差が確認された。

本論文の構成は以下のとおりである。本章に引き続き、2章では、組織のプロジェクトにおけるコミュニケーションや社会ネットワーク分析の先行研究を取り上げる。次に3章では、本研究の対象となる生産的なチームやコミュニケーションについて具体的に説明する。4章では、発言間構造を構築する手法とそれを分析する手法について説明する。5章では、ネットワークの構造的な特徴とチームの生産性との因果関係を調べるために行った回帰分析と、チームごとの構造的な特徴を調べるために行ったクラスタ分析についての結果と考察を述べる。最後に6章では本論文の内容をまとめ、今後の課題を述べる。

¹ 明治大学大学院先端数理科学研究科先端メディアサイエンス専攻

2. 関連研究

本章では、チームのコミュニケーション評価手法やネットワーク分析に関する先行研究を紹介する。

ソフトウェア開発プロジェクトにおけるインフォーマルコミュニケーションの評価手法 [4] では、コミュニケーションを「交わり」「伝達」の2側面にとらえ、共同作業に必要なコミュニケーションを人間関係維持型、個別対応型、徹底討論型、問題対応型、情報伝達重視型の5つの枠組みに分類し、「各メンバによる主観的評価→集計→結果の視覚化→プロジェクト診断」という一連の手続きを行うことで、プロジェクトの対面コミュニケーションを個人間のミクロな視点とプロジェクト全体のマクロな視点で評価している。この研究ではコミュニケーションを図る人間間の関係性に注目しているが、本研究ではコミュニケーションを構成する会話の発言間構造を分析する。

Twitter の情報伝播ネットワークの分析 [5] では、リプライやリツイートを手がかりに情報拡散構造、情報集約構造、情報転送構造の3種類の構造から情報伝播ネットワークを構成し、そのネットワークを分析することで、ユーザの興味を引くツイートの発見を試みている。本研究の発言間構造もこの研究の情報拡散構造、情報集約構造、情報転送構造を基にしているが、本研究では、生産的なチームのコミュニケーションに適用し、構造を分析することでチームの生産性を評価する。

3. 生産的集団のオンラインテキストコミュニケーション

本研究の対象は、プロジェクトや課題解決のために組織されたチームである。このような集団で何らかの目的達成やアウトプットの要求により結成された組織を生産的集団と定義する。このような生産的集団において、コミュニケーションの活発度が生産性を上げるという報告が数多くされている [6][7]。

一方、コミュニケーションの手法に着目すると、従来は対面会議のような組織のメンバが一同に会してお互いの意見を述べる場が多く設けられてきた。しかし、インターネットの普及に伴い、コミュニケーションの形態も多様化しており、企業等においても社内での連絡手段としてメールやビジネスチャットのようなコミュニケーションツールの導入を試みている。その中でもチャットツールを用いたコミュニケーションの場合は、コミュニケーションツールの中でも同期性が高い [8]。また、対面コミュニケーションで起こりがちな同時に複数メンバからの発言時の割り込みの困難さによる発言機会の損失によって、発言のタイミングを阻害されることによって成果ロスが引き起こされることが指摘されているが [9]、チャットツールではパラレルコミュニケーションが可能であるため、同時に発言されて

もシステムサーバがメンバの発言を独立に処理してディスプレイに表示することができる。以上の特徴から、チャットツールはアイディア創出といったいわゆる、ブレインストーミングのような創造性課題に向いているとされている [10]。

ブレインストーミングでは一般的に以下の4原則を守ることとされている [11]。

- (1) 判断・結論を出さない (結論厳禁)
- (2) 粗野な考えを歓迎する (自由奔放)
- (3) 量を重視する (質より量)
- (4) アイディアを結合し発展させる (結合改善)

アイディアをたくさん出し合うことによって、相互の連鎖反応や発想の誘発を期待できる。つまり、アイディアの拡散や集約、パスといった発言の働きが課題解決を成功させ、生産性を向上させることに繋がる。組織のチャットツールで行われるコミュニケーションもブレインストーミングの形態を取っていることが考えられ、本研究ではこのようなコミュニケーションの構造的特徴を分析の対象とする。

4. 提案手法

本章では、チャットのログからコミュニケーション構造を分析する手法を提案する。4.1 節では、コミュニケーションを定量的に評価するための発言間構造の構築方法について述べる。4.2 節では、構築した発言間構造を評価するためのネットワーク指標について述べる。

4.1 発言間構造の自動構造化手法

本節では、発言と発言の繋がりを特定して、自動的に発言間構造を構築する手法について述べる。

4.1.1 発言間の繋がりのルール

本研究で提案する原理は、組織のチャットなどで行われるテキストコミュニケーションのログから発言と発言の繋がりを基に発言間構造を構築することである。具体的には、チャットのような単純な時系列情報から複数のノードとリンクで構成される有向グラフに変換することで、定量的な評価を可能にする。ここで、ノードは発言、リンクは発言の関係性に対応する。発言と発言の繋がりは以下のように定義する。図1 発言の繋がりを特定して、グラフを構

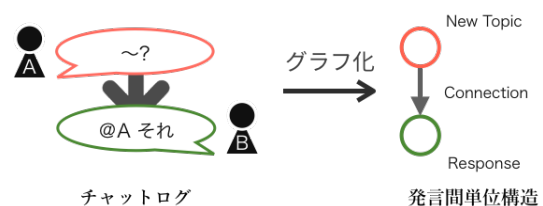


図1 発言の繋がりを特定して、グラフを構築する図

築する例を示す。

- (1) 引用：対象発言内に、それ以前の発言の一部を引用した箇所がある場合、その引用発言から対象発言にリンクを張る。
- (2) メンション：対象発言内に、それ以前の発言の発言者名を含む場合、この発言者の直近の発言から対象発言にリンクを張る。
- (3) 連続：対象発言が(1)、(2)に該当しない場合、対象発言の直前の発言から対象発言にリンクを張る。

4.1.2 発言間の構造化

本節では、発言間構造を自動で構築する手法を説明する。はじめに、組織のチャットなどで行われるテキストコミュニケーションのログを収集する。その中で時系列順で話題の一番最初にあたる発言を、初期ノードとする。初期ノードに対して、4.1.1 節のルールに基づいて、最初の発言のノードを上位ノード、新たな発言のノードを下位ノードとして上位ノードから下位ノードの向きにリンクを張り、ネットワークを形成する。この手順を繰り返すことで組織のコミュニケーションから発言間構造を構築する。組織のコミュニケーション全体としては、複数の有向部分グラフが形成される。ノードは発言するごとに時系列順に生成されていくため、新しいノードを上位ノード、古いノードを下位ノードとするリンクが形成されることはない。文脈から関係のある話題にノードとリンクを形成していくため、1つの部分グラフは1つの話題を表すことになる。組織のコミュニケーションのグラフ構造を可視化した例を図2に示す。

4.2 ネットワーク指標

本節では、4.1 章を基に構築した発言間構造からコミュニケーションを評価するためのネットワーク指標について述べる。

発言間構造はノードが発言、リンクが発言の関係性を表す有向グラフになる。図3に有向グラフを構成する発言間の最小の構造パターンを示す。図3(a)は発言の拡散構造を表し、例えば、発言に対してたくさんのリアクションが

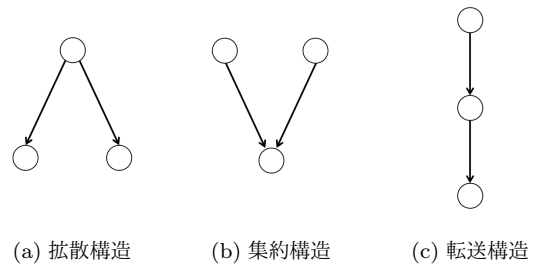


図3 有向グラフを構成する発言間の構造パターン

ある場合など、特定の上位ノードから複数のリンクが存在するときに構成される。図3(b)は発言の集約構造を表し、話題に対しての複数のリアクションの要点をまとめる発言や複数の話題を受けて新たな話題をスタートさせるリアクションの発言など、特定のノードへ複数のリンクが存在するときに構成される。図3(c)は発言の転送構造を表し、質問に対しての応答や発言への質問などにおいて、ある単一の発言からある単一の発言へのパスのような構造を構成することになる。

以下に、図3の構造パターンを、複雑ネットワーク科学のアプローチに基づいて測定するための指標の説明をする。

平均出次数

次数のうち特定のノードから別のノードに出ていくリンクの本数を出次数といい、平均出次数はネットワークにおける出自数の平均である。出次数はアイデアや発言の発散を表し、ネットワークはチームの話題ごとに分割されている部分グラフになるため、平均出次数はその話題の拡散度を表すことになる。ネットワーク構成としては、図3(a)のような構造を対象とする。ノードから出ていくリンクを持つノード数を n_{out} 、ノード i における出ていくリンクの本数を $od(i)$ とすると、平均出次数 *Diffusion* は以下の式(1)で表される。

$$Diffusion = \frac{\sum_{i=1}^{n_{out}} od(i)}{n_{out}} \quad (1)$$

平均入次数

次数のうち別のノードから特定のノードに入ってくるリンクの本数を入次数といい、平均入次数はネットワークにおける入次数の平均である。入次数はアイデアや発言のまとめを表し、ネットワークはチームの話題ごとに分割されている部分グラフになるため、平均入次数はその話題の集約度を表すことになる。ネットワーク構成としては、図3(b)のような構造を対象とする。ノードに入ってくるリンクを持つノード数を n_{in} 、ノード i における入ってくるリンクの本数を $id(i)$ とすると、平均入次数 *Intensively* は以下の式(2)で表される。

$$Intensively = \frac{\sum_{i=1}^{n_{in}} id(i)}{n_{in}} \quad (2)$$

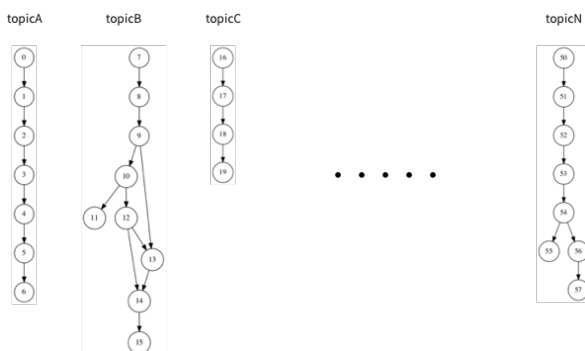


図2 組織コミュニケーションの可視化

転送数

特定のノードの出次数と入次数がどちらも1つのノードの前後において転送構造を構成する。本指標はこの転送構造を数え上げることでその話題における転送数とする。ネットワーク構成としては、図3(c)のような構造を対象とする。

平均経路長

任意の2つのノードを結ぶリンクの最小本数をパス長と呼び、ネットワークにおけるすべてのノード間のパス長の平均が平均経路長である。部分グラフの総ノード数を n 、ノード i とノード j 間の最短距離を d_{ij} とすると、平均経路長 $Path$ は以下の式(3)で表される。

$$Path = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j}^n d_{ij} \quad (3)$$

これはネットワークのおおよその大きさを表す指標である。発言間構造においては話題ごとにグラフが形成されるため、平均経路長は話題の冗長性を表す。

5. 実験

本章では、4章で提案したネットワーク指標に基づいたコミュニケーションの分析結果と、コミュニケーションの構造的特徴について述べる。

5.1 実験方法

本実験では、Github^{*1}で行われたコミュニケーションを対象として行った。Githubはソフトウェア開発のプラットフォームクラウドサービスである。Githubでは、コミュニティの中でリポジトリというプロジェクト単位にソースコードを分けてホスティングしており、問題や今後の展開などをリポジトリごとに紐づくIssues上で議論が行われる。このことから、Githubのリポジトリに所属するメンバで構成される集団を生産的集団、Issuesを組織のインフォーマルなオンラインテキストコミュニケーションの場とみなして、そこで行われるコミュニケーションログを基に実験を行う。Issuesにおける1つのIssueが1つの話題に相当する。

データセットは、2018年5月1日から2019年5月1日までに更新があったリポジトリのIssues上で行われた発言ログをGithub API^{*2}を用いて収集を行い、1008リポジトリ、66283の話題を収集した。得られた話題を4章の提案手法に基づいてネットワークを構築し、リポジトリごとのネットワーク指標を算出した。4.1.1を基に構築した発言の繋がりの割合を表1に示す。1つのリポジトリにおいて話題の数だけ部分グラフが構築されるため、話題の数だけネットワーク指標が算出されることになり、それら

表1 構造化したリンクの関係性の割合

	引用	メンション	連続
リンク数	27465	131975	398372
割合	5%	24%	71%

のIssuesにおける平均と分散を分析の際の指標として用いる。例えば、平均出次数の平均はそのチームの1つの話題における拡散度を表し、分散はそのチームの拡散度の偏りを表す。

本実験の目的はチームの構造的特徴を明らかにすることである。そのために、まず、ネットワークの構造的特徴とチームの生産性との因果関係を調べるために回帰分析を行った。次に、チームごとの構造に特徴が見られたため、ネットワーク指標によるクラスタ分析を行った。ただし、Issuesで全く議論が行われておらず、発言が0だったものは除外するものとした。

回帰分析

話題から構築したネットワークの平均出次数、平均入次数、転送数、平均経路長それぞれの平均と分散を説明変数とした。目的変数には、リポジトリのスター数と更新頻度をそれぞれ正規化した和をチームの生産性と定義して用いて、最小2乗法によってモデリングを行った。モデルの評価は、各指標の8割を訓練用データ、2割をテストデータとして交差検証法によって行った。

クラスタ分析

話題から構築したネットワークの平均出次数、平均入次数、転送数、平均経路長それぞれの平均と分散を指標として、K-means法によってクラスタリングを行った。なお、セントロイドの初期値選択は、異なる乱数のシードでランダムに重心を選ぶ処理を10回実行することで決定した。

5.2 実験結果

本節では、ネットワーク指標を用いた回帰分析とクラスタ分析の結果と考察について述べる。

5.2.1 回帰分析

説明変数を基に線形モデルを構築し、交差検証による精度を評価した結果、予測精度は1.3%と低い結果になった。このことから説明変数と目的変数の間には線形的な因果関係はないと考えられる。

このような結果になった要因と考えられるのは、説明変数の不足、不適合および、チームの生産性要因の不足である。説明変数と目的変数の相関関係はすべて0.4以下で相関がない可能性が高い。また、本実験では、正規化したリポジトリのスター数と更新頻度の和を目的変数としたが、チームの生産性を決定づける別の指標が存在することが考えられ、それらの指標に置き換える、もしくは複合するこ

^{*1} <http://github.com>

^{*2} <https://developer.github.com>

表 2 クラスタごとの各指標の平均

	クラスタ 1	クラスタ 2	全体
サンプル数	800	168	968
総発言数平均	673	355	618
平均出次数 (平均)	1.080	1.191	1.099
平均出次数 (分散)	0.023	0.145	0.044
平均入次数 (平均)	1.023	1.074	1.032
平均入次数 (分散)	0.005	0.024	0.009
転送数 (平均)	0.745	0.660	0.730
転送数 (分散)	0.022	0.046	0.026
平均経路長 (平均)	1.332	1.230	1.315
平均経路長 (分散)	0.697	0.600	0.680

とで最適なモデルを構築できる可能性がある。

5.2.2 クラスタ分析

クラスタについては、シルエット係数の値から 2 つのクラスタに分類した。シルエット係数は、クラスタ内のサンプルがどの程度密になっていて (凝集度)、かつ、クラスタ間がどれだけ離れているか (乖離度) を表す値であり、各サンプルの平均クラスタ内距離を凝集度を a 、各サンプルの平均最近接点距離を乖離度を b とすると、シルエット係数 S は以下の式 (4) で表される。

$$S = \frac{|b - a|}{\max(b, a)} \quad (4)$$

シルエット係数は -1 から 1 までの範囲を取り、1 に近いほどクラスタ間が遠く離れていることを表す。本実験のクラスタ分析におけるシルエット係数は 0.44 であり、まずまずのクラスタリング精度であった。

表 2 に分類されたクラスタごとのサンプル数、総発言数の平均、サンプルごとの各指標の平均を示す。

平均出次数と平均入次数に関しては、クラスタ 1 は、平均・分散の値が共に全体の平均に比べて低い値を取っている。すなわち、1 つの話題の発言の拡散度・集約度が共に低く、また、各話題の拡散度・集約度の偏りが小さいため、拡散度・集約度が同程度の話題で構成されていることがわかる。それに比べてクラスタ 2 は、平均・分散の値が共に全体の平均に比べて高い値を取っている。すなわち、1 つの話題の発言の拡散度・集約度が共に高く、また、各話題の拡散度・集約度の偏りが大きいため、拡散度・集約度の高い話題と低い話題が混在していることがわかる。図 4 にクラスタ 2 に分類されたチームのネットワーク構成例を示す。総発言数平均に着目すると、クラスタ 1 の方がクラスタ 2 よりも高いことから、クラスタ 2 のチームの発言数が大きいために平均や分散が大きくなっているわけではないことがわかり、クラスタにおける固有の構造的特徴であると言える。

転送数に関しては、クラスタ 1 は、平均が全体の平均と比べて高く、分散は全体の平均と比べて低い値を取っている。すなわち、1 つの話題における転送数が多く、また、

各話題の転送数の偏りが小さいため、どの話題も同じくらい多くの転送数が存在することがわかる。転送数が連続すると拡散も集約もしない冗長的な構造になるため、コミュニケーションの構造としてはふさわしくない構造になると考えられる。それに比べてクラスタ 2 は、平均が全体の平均と比べて低く、分散は全体の平均と比べて高い値を取っている。すなわち、1 つの話題における転送数が少なく、また、各話題の転送数の偏りが大きいため、転送数が多い話題と少ない話題が混在していることがわかる。しかし、表 1 より連続リンクの割合が高いため、これが要因で本来、拡散や集約といった構造になるものが転送構造になっている可能性も考えられる。

平均経路長に関しては、クラスタ 1 は、平均・分散の値が共に全体の平均に比べて高い値を取っている。すなわち、チームの話題が冗長であり、また、各話題の平均経路長の偏りが大きいため、冗長な話題と簡潔な話題が混在していることがわかる。図 5 にクラスタ 1 に分類されたチームのネットワーク構成例を示す。平均経路長は連続した転送構造において大きくなる傾向があるため、転送数の指標と合わせても、クラスタ 1 のチームは転送構造が連続した冗長的な構造を取っている可能性が高い。それに比べてクラスタ 2 は、平均・分散の値が共に全体の平均に比べて低い値を取っている。すなわち、チームの話題が簡潔であり、また、各話題の平均経路長の偏りが小さいため、どの話題も簡潔な話題で構成されていることがわかる。しかし、総発言数平均がクラスタ 2 よりクラスタ 1 の方が高いことによって平均・分散が高くなった可能性も考えられる。

以上の結果をまとめると、クラスタ 1 のチームは、すべての話題における発言の拡散度・集約度が同程度に共に低く、また、すべての話題における転送数が同程度に高く、さらに、1 つの話題における話題の大きさが大きく、その中でも話題の大きい話題と小さい話題が混在する。クラスタ 2 のチームは、1 つの話題における発言の拡散度・集約度が共に高く、その中でも拡散度・集約度が共に高い話題と低い話題が混在しており、また、1 つの話題における転送数が低く、その中でも転送数が高い話題と低い話題が混在しており、さらに、すべての話題における話題の大きさが同程度に小さい。

本実験では、2 つのクラスタに分類したが、データの数を増やしたときに新たなクラスタが形成される可能性がある。クラスタ 1 とクラスタ 2 に内在される構造的特徴を持ったチームが発見されることが考えられる。

6. おわりに

本研究では、チームコミュニケーションの構造的特徴を解明し、それがチームの生産性に影響をもたらすことを定量的に評価することを目的とし、チャットの時系列情報から複数のノードとリンクで構成される有向グラフに変換

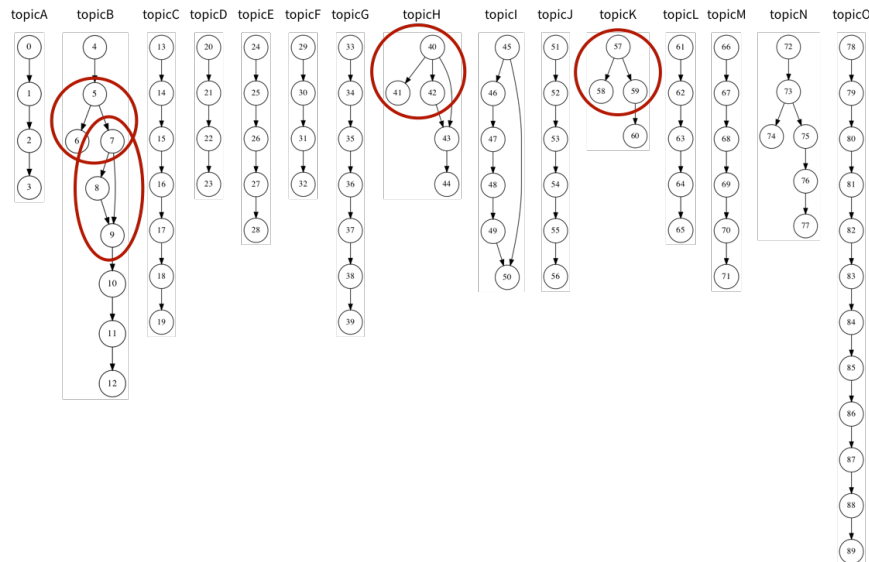


図 4 クラスタ 2 のネットワーク構成例と特徴

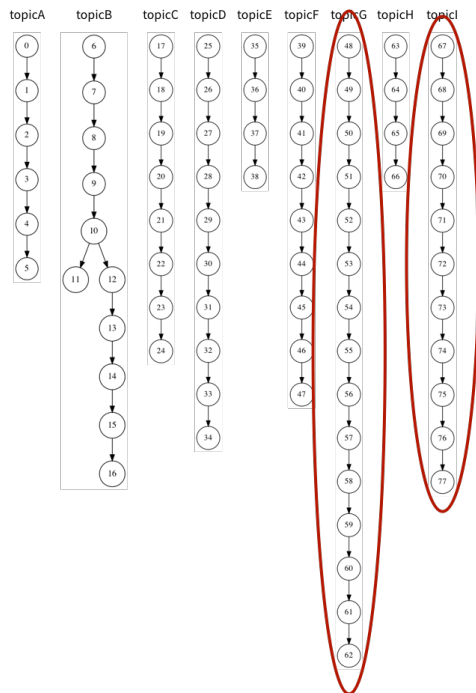


図 5 クラスタ 1 のネットワーク構成例と特徴

し、ネットワーク指標に基づいてコミュニケーションの構造的特徴とチームの生産性を評価した。実験から、コミュニケーションの構造的特徴とチームの生産性に統計的な因果関係は確認されなかったが、クラスタ間の構造的特徴については有意差が確認された。今後は、チームの生産性の指標を再検討し、コミュニケーションの構造的特徴とチームの生産性の因果関係の解明、および、検討された指標を基に話題の盛り上がりを支援する会話シミュレーションシステムの設計をする予定である。

参考文献

- [1] Salas, E., Stagl, K. C., Burke, C. S., Goodwin, G. F.: Fostering team effectiveness in organizations: Toward an integrative theoretical framework, Nebraska symposium on motivation, Vol. 52, p. 185 (2007).
- [2] Manser, T.: Teamwork and patient safety in dynamic domains of healthcare: a review of the literature, Acta Anaesthesiologica Scandinavica, Vol. 53, No. 2, pp. 143-151 (2009).
- [3] Jun Rekimoto, Yuji Ayatsuka, Hirotaka Uoi, Toshifumi Arai: Adding Another Communication Channel to Reality: An Experience with a Chat-Augmented Conference, CHI '98 conference summary on Human factors in computing systems, Vol. 18, No. 23, pp. 271-272 (1998).
- [4] 仲谷美江, 原島博, 西田正吾: ソフトウェア開発プロジェクトにおけるインフォーマルコミュニケーションの評価手法, 電子情報通信学会論文誌 D, Vol. 77, No. 4, pp. 823-837 (1994).
- [5] 風間一洋, 今田美幸, 柏木啓一郎: Twitter の情報伝播ネットワークの分析, 第 24 回人工知能学会全国大会 (2010).
- [6] 渡邊純一郎, 藤田真理奈, 矢野和男, 金坂秀雄, 長谷川知之: コールセンタにおける職場の活発度が生産性に与える影響の定量評価, 情報処理学会論文誌, Vol.54, No. 4, pp. 1470-1479 (2013).
- [7] 縄田健悟, 山口裕幸, 波多野徹, 青島未佳: 企業組織において高業績を導くチーム・プロセスの解明, 心理科学研究, Vol. 85, No.6, pp. 529-539 (2015).
- [8] 山本修一郎: CMC で変わる組織コミュニケーション, NTT 出版 (2010).
- [9] Coskun, H., Paulus, P., Brown, V., Sherwood, J. J.: Cognitive stimulation and problem presentation in idea-generating groups, Group Dynamics: Theory, Research, and Practice, Vol. 4, No. 4, pp. 307-329 (2000).
- [10] 古川久敬: 電子コミュニケーションとチーム活動, 組織科学, Vol. 29, No. 1, pp. 18-28 (1995).
- [11] Osborn, A. F.: Applied imagination, New York: Scribner's (1963).