

## 知人のネットワークの概念に基づいた情報共有機構

高橋 範泰      山下 剛史

富士ゼロックス株式会社 総合研究所  
〒259-0157 神奈川県足柄上郡中井町境 430  
{noriyasu, take}@crl.fujixerox.co.jp

あらまし    個人が持つ情報内容に応じて人と人を引き合わせることで、組織やコミュニティにおける情報共有を実現することを狙いとして、ANS(Acquaintance Network System)と呼ぶ知人のネットワークの概念に基づいた情報共有機構を提案する。ANS では、知人の関係を意味と重みを持ったパイプと呼ぶデータ構造で表現し、複数のパイプの合成処理により新しい関係に対応するパイプを生成する。これにより、情報の開示範囲を知人の範囲に限定しているにもかかわらず、広範囲の情報提供者と情報消費者の引き合わせが実現できる。さらにシミュレーションにより、情報の開示範囲を限定している場合の比較で、ANS は他のシステムの 10 倍以上の引き合わせ能力を持つことを明らかにした。

キーワード    コミュニティ 情報共有 引き合わせ 関係 ワークグループ

### Information Sharing Mechanism Based on the Concept of Acquaintance Network

Noriyasu Takahashi    Takeshi Yamashita

Corporate Research Labs, Fuji Xerox Co., Ltd.  
430 Sakai, Nakai-machi, Ashigarakami-gun, Kanagawa 259-0157  
{noriyasu, take}@crl.fujixerox.co.jp

Abstract    We propose information sharing mechanism called ANS (Acquaintance Network System) based on the concept of human acquaintance network. ANS incorporates a data structure named pipe that expresses human relationship and synthetic processes between pipes that generate new pipes corresponding to new relationships. A pipe has a set of keywords and weight values as its attribute. By using ANS, we can provide global matchmaking services between information providers and information consumers using locally published information. Furthermore, by simulation analysis, we found that ANS has more than ten times higher matchmaking ability compared to the other matchmaking systems in case of personal information is published locally.

key words    community, information sharing, matchmaking, relation, work group

## 1. はじめに

組織やコミュニティの活動においては、メンバーが持つ情報を共有することで知的生産性の向上をはかる必要がある。情報共有の手段としてデータベースや Web サーバが用いられているが、全ての情報を登録し管理する事は不可能であり、個人が持つ情報を見つけ出し利用することが必要になる。このとき、何らかの手段で情報を持っている人(情報提供者)を、情報を必要とする人(情報消費者)に引き合わせる必要がある。例えば、企業の設計部門では、設計ノウハウ情報を共有するために Web ベースの情報共有システムを運用している。しかし、かなりの部分の設計ノウハウは設計者の手元で未整理のままになっており、これを見つけ出し積極的に利用することが求められている。

情報提供者を情報消費者に引き合わせるシステムとして、Matchmaker<sup>[3]</sup>では、情報提供者の能力を管理する仲介サーバを設け、情報消費者の要求にオンデマンドに引き合わせる。また Yenta<sup>[4]</sup>では、メールや個人が管理する文書などから抽出した興味ベクトルを比較し、興味の近いユーザのクラスタを自動生成し、ユーザが相互にコンタクトを取れるようにする。しかし、これらのシステムは、個人の情報を共通の場に開示する事を前提に広域的な引き合わせが成立しており、開示された情報はユーザにとって素性のわからない相手にも知られてしまう。このような状況では、ユーザは開示する情報を悪用されることを警戒し個人的な情報を開示したがいらないため、共有の対象となるのは、商品の値段や仕様など公共性の高い情報や、個人の趣味など他人に知られても差し障りのない情報に限られてしまう。

そこで我々は、ANS(Acquaintance Network System)と呼ぶ知人のネットワークの概念に基づいた情報共有の機構を提案する<sup>[4]</sup>。ANS では、知人の関係を意味と重みを持ったパイプと呼ばれるデータ構造で表現し、複数のパイプの合成処理により新しい関係に対応するパイプを生成する。この機構により、人が知人のつてをたどって情報提供者にたどり着くのと同様に、限定した範囲に開示された個人情報を用いて、広範囲の情報提供者と情報消費者の引き合わせを実現できる。

我々のシステムと同様に、人の関係を用いて情報提供者を見つけるシステムに、Referral Web<sup>[2]</sup>、Schwartz らのシステム<sup>[4]</sup>、緒方らのシステム<sup>[6]</sup>、がある。これらのシステムでは、必要な情報提供者へのポインタだけを得るのに対して、我々のシステムは、引き合わせの結果得られる関係の意味に着目している点と、得られた関係はその後の処理で利用するという動的なネットワークの管理を行っている点に違いがある。

以下、2 章ではパイプによる知人のネットワークのモデル化とパイプの合成処理を説明し、3 章ではパイプの合成処理の応用サービスを説明する。4 章では ANS の実用性を検証するために行ったシミュレーションについて説明し、5 章でまとめる。

## 2. 知人のネットワークのモデル化

### 2.1 知人のネットワーク利用の概念

人は組織やコミュニティでの情報の入手において、知人のつてをたどることによって情報提供者を探し出しアクセスする。例えば、A さんが「エージェントソフトウェアの国際標準にはどんなものがあるか」を知りたいとする。このとき A さんは、その領域に知見のありそうな B さんにアクセスし、B さんから C さんの紹介を受け、C さんにアクセスする。さらに C さんから D さんの紹介を受け、最終的に D さんから必要な情報を得る。

このような知人のネットワークの利用において、知人の関係は相手と共有する話題で表現できる。また、知人のつてをたどることは、既存の知人の関係を利用して新たな知人の関係を作り出す事に相当する。我々はこれらの点に着目し、知人のネットワークのモデル化を行う。

### 2.2 パイプによるモデル化

我々の提案する機構では、知人の関係をパイプと呼ぶデータ構造で表現する。パイプは特定の二者間にはただ一つ存在し、二人が共有する話題をパイプ属性として持つ。パイプ属性は、話題中に含まれるキーワードの集合とそれぞれのキーワードの重みで表現される。図 2.1 にパイプの属性の一例を示す。パイプ属性は、例えば、二人の間でやり取りされる

メールからキーワードを抽出し、さらにキーワードの出現頻度を重みとすることで得られる。

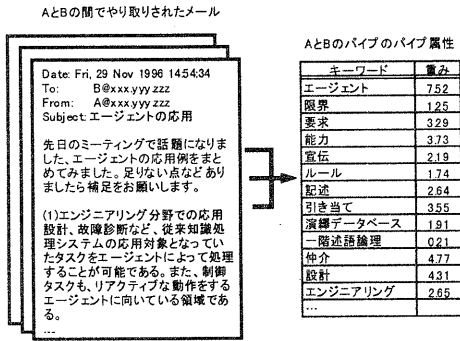


図 2.1 パイプのデータ構造

図 2.2 にパイプによって表現された知人のネットワークの例を示す。図でノードは人に対応し、ノードをつなぐリンクはパイプに相当する。リンクに付加されたラベルは、対応するパイプのパイプ属性がラベルで代表される話題を含んでいることを示す。例えば、A-B 間のパイプには、“エージェント”というラベルが付加されているが、これは A-B 間のパイプのパイプ属性がエージェントの話題(パイプ属性としては、“エージェント”、“自律”、“学習”、“秘書”などのキーワードとその重み)を含むことを示す。

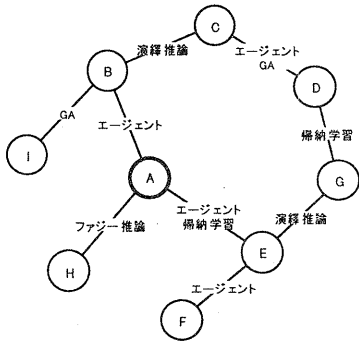


図 2.2 パイプのネットワークの例

### 2.3 パイプの基本合成処理

人が行う知人のネットワークの利用と同様の処理を実現するために、ANS では二種類のパイプの基本合成処理を導入する。以下の記述では、パイプ P

の属性を  $attr(P)$  と表記する。

#### パイプの直列合成処理

パイプの直列合成処理は、知人の仲介で新しい知人の関係を結ぶ処理である。図 2.3 にパイプの直列合成処理の概念を示す。図で、A と B の間ではエージェントに関する話題と帰納学習に関する話題を共有しており、B と C の間ではエージェントに関する話題を共有している。この場合、A と C は B を仲介してエージェントの話題を共有することができる。そこで直列合成を行って、A と B の共有話題と B と C の共有話題の共通部分をパイプ属性として持ったパイプ  $P_{AC}$  を生成する。この場合の  $P_{AC}$  の属性は次の式で現される。

$$attr(P_{AC}) = attr(P_{AB}) \cap attr(P_{BC})$$

$P_{AC}$  が既に存在する場合は、 $P_{AC}$  の属性にエージェントの話題が加えられる。

$$attr(P_{AC})' = attr(P_{AC}) \cup (attr(P_{AB}) \cap attr(P_{BC}))$$

これらの式で、“ $\cap$ ” はパイプ属性の積演算と呼ばれ、二つのパイプ属性のキーワード集合の積集合をとり、重みについては最小値をとる演算である。また、“ $\cup$ ” はパイプ属性の和演算と呼ばれ、二つのパイプ属性のキーワード集合の和集合をとり、重みについては最大値をとる演算である。

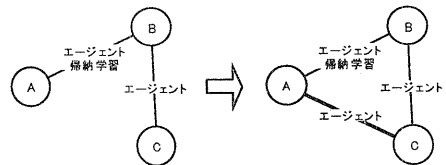


図 2.3 パイプの直列合成処理

#### パイプのたみこみ合成処理

パイプのたみこみ合成処理は、知人との関係に潜在的に共有可能である話題を見つけ、新たな共有話題として追加する処理である。図 2.4 にたみこみ合成処理の概念を示す。図で、B と C の間では、演繹推論に関する話題を共有している。一方、B は A と、C は D とそれぞれエージェントに関する話題を共有しており、B と C は潜在的にはエージェントに関する話題を共有できる。そこで、たみこみ合

成を行って、A と B の共有話題と C と D の共有話題の共通部分を、B と C の共有話題に加えることにより、B と C との潜在的な共有話題を顕在化する。この場合の  $P_{BC}$  の属性は次の式で表される。

$$\text{attr}(P_{BC}') = \text{attr}(P_{BC}) \cup (\text{attr}(P_{AB}) \cap \text{attr}(P_{CD}))$$

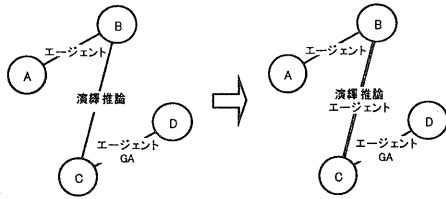


図 2.4 パイプのたたみこみ合成処理

### 3. 応用サービス

#### 3.1 ワークグループの形成

直列合成とたたみこみ合成を組み合わせることで、さまざまな応用サービスを実現することができる。ここでは、もっとも典型的なサービスである「ワークグループの形成」について説明する。他の応用サービスには、リクエストした話題に関する情報提供者を発見する「情報提供者の発見」、および、任意の相手と共有できる話題を見出す「二者間の関係の算出」がある。

ワークグループの形成は、要求された話題に関して議論できる人を集めてワークグループを作るサービスである。集められた人には、情報提供者も情報消費者も含まれる。図 3.1 にワークグループ形成の例を示す。図は、A から「エージェントに関する話題を共有するワークグループを形成したい」というリクエストが出された場合の、パイプのネットワークが変化の様子を示している。この例でワークグループが形成されるステップを以下に示す。

1. A がリクエストを発行する。
2. リクエストがエージェントの話題を共有する B と E に伝達される。
3. E に伝達されたリクエストから 4 を実行する。
  4.  $P_{AE}$  と  $P_{EF}$  を直列合成し、エージェントの話題を含むパイプ  $P_{AF}$  を生成する。
5. B に伝達されたリクエストから 6,7,8 を実行する。
  6.  $P_{AB}$ 、 $P_{BC}$ 、 $P_{CD}$  をたたみこみ合成し、 $P_{BC}$  の属

性にエージェントの話題を加える。

7.  $P_{AB}$  と  $P_{BC}$  を直列合成し、エージェントの話題を含むパイプ  $P_{AC}$  を生成する。
8.  $P_{AC}$  と  $P_{CD}$  を直列合成し、エージェントの話題を含むパイプ  $P_{AD}$  を生成する。
9. 最終的にワークグループを形成する相手とのパイプ、 $P_{AB}$ 、 $P_{AC}$ 、 $P_{AD}$ 、 $P_{AE}$ 、 $P_{AF}$ 、が得られる。

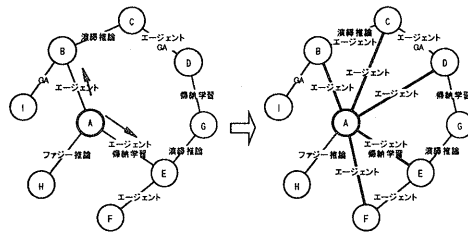


図 3.1 ワークグループの形成

#### 3.2 リクエスト

ワークグループの形成は、リクエストの発行によって起動される。リクエストの例を図 3.2 に示す。このように自然言語で記述されたリクエストからキーワードを抽出し、出現頻度をキーワードの重みとして、リクエスト属性が得られる。リクエスト属性は、リクエストを伝達する際に伝達先のノードを絞り込むために用いる。具体的には、リクエスト属性とパイプ属性の内積値が設定された閾値を超えている場合のみ、パイプの先のノードをリクエストの伝達先とする。

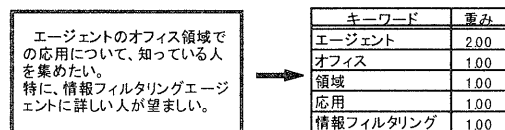


図 3.2 リクエスト

#### 3.3 システム構成

応用サービスを実現する ANS のシステム構成を図 3.3 に示す。システムは電子メールツールに接続して稼動する事を想定している。各個人に対応したパーソナルマネージャが存在し、電子メールからのパイプの生成と、リクエストに対するサービスを行

う。ユーザがシステムにリクエストを入力すると、形成されたワークグループのメンバーのリストが表示されるとともに、ワークグループのメンバーとの間に既存のコミュニケーションツール(電子メール、ビデオ会議、チャットシステムなど)を用いてコミュニケーションチャンネルが形成される。その後は、これらのツールを用いてユーザ主導のコミュニケーションを行う。

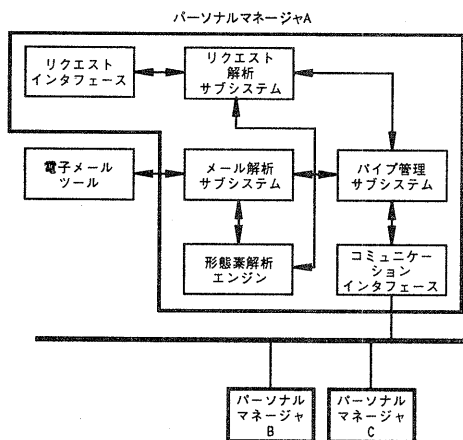


図 3.3 ANS のシステム構成

システムは Java を用いて構築中で、現在までに形態素解析エンジン、メール解析サブシステム、および、パイプ管理サブシステムを有する、単一システム上で動作する集中型のシステムが完成しており、動作検証を完了している。

### 3.4 システムの動作例

我々は、実際にやり取りされたメールを用いて、システムの動作の確認を行った。用いたメールは 7 人の協力者から収集した 257 通のメールである。メールの内容は特に限定しておらず、業務、趣味、新製品情報など、様々なものを含む。協力者の知人の知人までの範囲で計 186 のノードと、282 本のパイプが生成された。このデータを用いて、協力者に提供してもらったリクエストを処理したところ、ワークグループの形成が確認できた。図 3.4 に収集したデータを用いたシステムの動作例を示す。図 3.4 の例は、「F 1 日本グランプリの宿の手配」に関するリクエストを入力したときのシステムの処理結果を

示している。

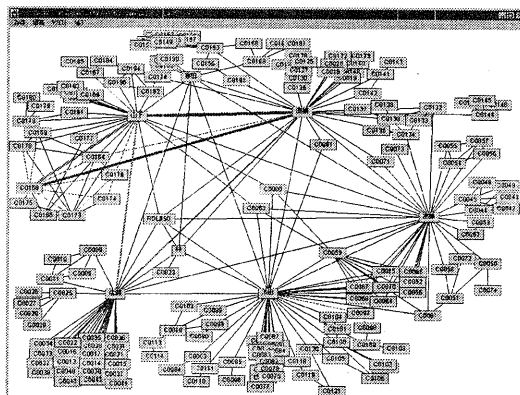


図 3.4 実メールを用いたシステムの動作例

## 4. シミュレーション

### 4.1 シミュレーションのねらい

我々は、ANS の引き合わせ能力を検証するためにシミュレーションを行った。シミュレーションのねらいは以下の二つである。

(1)情報の開示範囲と引き合わせ能力の関係を明らかにする。引き合わせのための情報を共通の場を開示する方式の従来の引き合わせシステムでは、開示範囲を限定すると引き合わせ対象も限定されるため、引き合わせ能力は減少する。これに比較して ANS では、情報の開示範囲を超えて具体的にどの程度の引き合わせ能力を発揮するかを定量的に明らかにする。なお、広域的な引き合わせの目安として、我々は、従来の引き合わせシステムより一桁高い引き合わせ能力を設定した。

(2)動的なネットワーク管理に関する特徴を明らかにする。ANS が持つアルゴリズム上の特徴の一つは、引き合わせの結果得られた知人の関係によって知人のネットワークが更新されるという、動的なネットワークの管理を行っている点である。このことは、知人のネットワークが強化されその後の引き合わせ能力が向上するという利点を生む。そこで、知人のネットワーク強化による引き合わせ能力向上の度合いを定量的に捉える。

## 4.2 シミュレーション方法

シミュレーションは、コミュニティをあらゆるネットワークをランダムに生成し、ワークグループ形成処理を適用して引き合わせ結果を検証する事で行った。シミュレーションにおいて設定したパラメータにはネットワークの形状に関するパラメータと、リクエストに関するパラメータがある。

ネットワークの形状に関するパラメータには、ネットワークのノード数(N: コミュニティに属するユーザ数)、平均リンク数(L: 知人の数の平均)、情報提供者の数(S: 情報提供者の数)、関連する話題を有するリンク数(M: 関連する話題を知っているユーザ数)がある。実際のコミュニティでの情報収集では、リクエストの直接の答えは持っていないが関連する話題に関与している人が、仲介役を果たす。関連する話題を有するリンク数(M)はこの仲介役の人数に対応している。

リクエストに関するパラメータには、閾値(TH: リクエストに適合する話題を絞りこむ度合い)、たたみこみ戦略(問いかけて仲介を行う度合い)がある。閾値はパイプがリクエストに合致するかどうかの判定に用いており、リクエストを伝える相手を選ぶ際にどの程度相手を厳選するかに対応している。たたみこみ戦略については次節で説明する。

コミュニティ内でやり取りされる情報は、「目的情報に全く関係ないもの」、「目的情報に関係ないが関連する話題に関係するもの」、および、「目的情報に直接関係するもの」、の三種類であるとし、それぞれに、 $K_l = \{l_1, l_2, \dots, l_{100}\}$ 、 $K_m = \{m_1, m_2, \dots, m_{10}\}$ 、 $K_h = \{h_1, h_2, h_3\}$ というキーワード集合で表現する。その上で、全てのパイプに  $K_l$  からランダムに選択した 5 個のキーワードを、M 個のパイプに  $K_m$  からランダムに選択した 5 個のキーワードを、S/2 個(S は常に偶数)のパイプに  $K_h$  の 3 個のキーワードを設定する。また、キーワードの重みはすべて 1 とした。

データは、同一パラメータに対してランダムに生成した 500 個のネットワークデータに対してシミュレーションを行い、その結果の平均を取った。

## 4.3 たたみこみ戦略

たたみこみ戦略は、どのような条件でたたみこみ

合成を実施するかを示すものである。ここでは、“None”(たたみこみ合成をしない)、“<1”(リクエスト伝達候補が 0 のときたたみこみ合成を実施する)、“<2”(リクエスト伝達候補が 2 未満のときたたみこみ合成を実施する)、“<3”(リクエスト伝達候補が 3 未満のときたたみこみ合成を実施する)、“Always”(常にたたみこみ合成を実施する)、“Human[0.7, 0.7]”(人の行動に対応)を用意した。

“Human[0.7, 0.7]”は人の行動に対応したたたみこみ戦略である。これは直列合成を  $0.7^k$  の率で、たたみこみ合成を  $0.7^k$  の率で実行する事を意味する。ここで  $k$  はリクエスト発行者からの距離である。アンケート調査の結果、実際の人の仲介行動における直接の仲介行動(直列合成に相当)と、問いかけての仲介行動(たたみこみ合成に相当)の率は、それぞれ高々  $0.7^k$  であることが判明している。

## 4.4 シミュレーション結果

我々が想定している人を介した情報共有の形態では、情報の開示範囲は知人の数、すなわち平均パイプ数  $L$  に相当する。例えば、ノード数  $N$  が 500 のネットワークで平均パイプ数  $L$  が 24 であれば、情報の開示範囲は 4.8%(24/500)である。そこで、平均パイプ数と引き合わせ率(引き合わせることできたノード数のネットワーク全体に存在する情報提供者ノード数に対する割合。検索システムにおける再現率に相当する)の関係をシミュレーションする事で、情報の開示範囲と引き合わせ能力の関係を明らかにした。図 4.1 に、平均パイプ数と引き合わせ率の関係を示す。たたみこみ戦略“Always”では、平均パイプ数が増えるにしたがって引き合わせ率は上昇し、平均パイプ数が 30 に達したあたりから上昇の度合いが鈍り 100%に漸近する。平均パイプ数 24 での開示範囲は 4.8%であることから、従来の引き合わせシステムでの引き合わせ率は 4.8%(情報提供者数ノード数 6 個のうち 0.23 個)を超える事はないのに対して、ANS では 66.2%と従来の引き合わせシステムの 10 倍を超える引き合わせ率を示している。従来の引き合わせシステムにおける限界値の 10 倍の引き合わせ率を、図 4.1 の二点鎖線で示した。引き合わせ戦略“Always”では、平均リンク数 18 から 42 の領域でこの値を超えており、平均リンク

数のほぼ全域で従来の引き合わせシステムに比べて一桁高い引き合わせ率を示している。

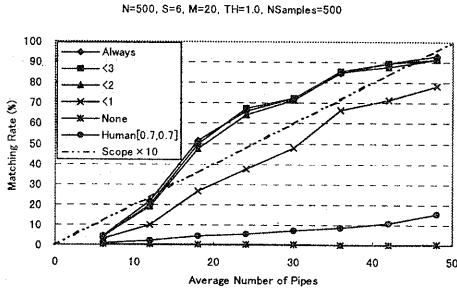


図 4.1 ANS の引き合わせ能力

たたみこみ戦略の比較では、すべての閾値で“Always”が最大の引き合わせ能力を示した。“<3”および“<2”は、“Always”と同等の引き合わせ能力を発揮している。一方、“<1”は、“Always”のほぼ60%の引き合わせ能力である。“Human[0.7,0.7]”と“None”は、ほとんど引き合わせが成功しない。

図 4.2 に、同じリクエストを継続的に発行した場合の引き合わせ能力の変化を示す。リクエストを発行するノードは毎回変化させた。引き合わせ能力が回毎に強化され、一回目では 36%であった引き合わせ率が、10 回の繰り返しの結果、ほぼ倍の 69%にまで上昇する。

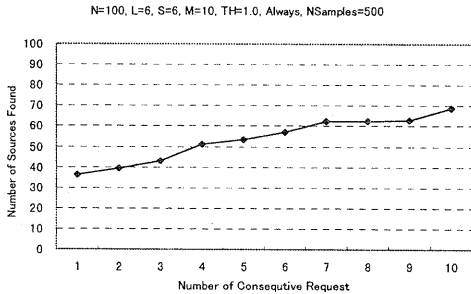


図 4.2 継続的引き合わせによるネットワーク強化

#### 4.4 シミュレーション結果の考察

##### より大きなコミュニティでの引き合わせ能力

引き合わせ能力検証のためのシミュレーション結果から、ANS はメンバーの数が 500 人規模までのコミュニティであれば、情報の開示範囲を限定した

状態で、従来の引き合わせシステムに比べて 10 倍以上の引き合わせ能力が得られる事が明らかになった。しかし、シミュレーション結果をより規模の大きなコミュニティに適用した場合、ANS の引き合わせ能力はノード数の 1.5 乗に反比例して減少するため、一般的にコミュニティの規模が大きい場合は知人の数も増えることを考慮しても、ある程度以上の規模では十分な引き合わせ能力は得られない。例えば、5000 人のコミュニティで知人の数が 72 であった場合、以下の式で引き合わせ率の推定ができる。

(引き合わせ率) =

$$(86 / (5000 / 500)^{1.5}) \times (72 / 36) = 5.4(\%)$$

ここで、86 は図 4.1 に示した、L=36 におけるたたみこみ戦略“Always”の引き合わせ率である。また、上式では、リンク数の増加に対しては係数 1 で比例して引き合わせ率は増加すると仮定している。

しかし実際の企業内の組織や、地域や趣味のコミュニティを想定した場合は今回のシミュレーションで想定したように全体が一様なネットワークではなく、図 4.3 に示すように、サブコミュニティが存在し、それらが橋渡し役の人を通じてつながっていると考えられる。このような場合、ANS のたたみこみ合成の機能でひとたびサブコミュニティを超えた引き合わせが実現されれば、つながったコミュニティ内の情報提供者への距離は、同一のコミュニティに属する情報提供者への距離と同等になる。したがって、橋渡し役の人が存在するサブコミュニティに対しては今回のシミュレーション結果と同等の高い引き合わせ能力を示すと考えられる。これについては、今後、サブコミュニティを想定したネットワークでのシミュレーションを行うなどして、定量的な検証を行う必要がある。

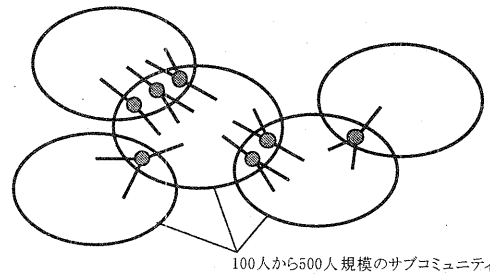


図 4.3 大規模なネットワークでの ANS の動作

## ネットワーク強化能力の実際的な意味

ANS のアルゴリズム上の特徴である、引き合わせの結果得られたパイプでネットワークを強化する事で、同一種類のリクエストが継続されて出される場合には、引き合わせ能力が向上することが定量的に明確になった。しかし、ANS が実際に運用される場合には、このシミュレーションで想定したように同一種類のリクエストが継続して出されるだけでなく、さまざまな種類のリクエストがさまざまなノードから出される。この場合、ネットワークの強化が急速に進み、パイプ数およびパイプの情報が増大するために、ネットワークがどんどん密な状態になり、リクエストの伝達が爆発的に増大する可能性がある。この問題に対しては、パイプ属性の重みを時間の経過とともに減少させる仕組みを導入することが考えられる。このような仕組みの導入により、頻繁にリクエストが出される情報については ANS の特徴でネットワークが強化され、それ以外の情報については時間の経過とともにネットワークが疎なものになっていくと考えられる

## 5. おわりに

組織やコミュニティにおける情報共有の手段として知人の関係をパイプと呼ぶデータ構造で表現し、パイプ間の合成処理により人の引き合わせを実現する機構を提案した。さらにこの機構を用いてワークグループの形成をはじめとする応用サービスが実現できる事を示した。また、シミュレーションにより、我々の提案する機構が、情報の開示範囲を限定している場合における比較で他のシステムの 10 倍以上の引き合わせ能力を持つことと、ネットワークの強化能力を持つ事を明らかにした。

現在、大規模な実際のメールデータを用いたシステムの検証を行うとともに、分散環境で動作するシステムの構築を進めている。今後は、より大規模なネットワークや一様でないネットワークでの実用性をさらに検証し、実用システムの構築につなげていく予定である。

## 参考文献

- [1] L. Foner, "A Multi-Agent Referral System for Matchmaking," Proceedings of the First

International Conference on Practical Application of the Intelligent Agents and Multi-Agent Technology (PAAM '96), April 1996.

- [2] H. Kautz, B. Selman, and A. Milewski, "Agent Amplified Communication," Proceedings of the Thirteents National Conference on Artificial Intelligence (AAAI'96), pp. 3-9, Menlo Park, USA, 1996, AAAI Press.
- [3] D. Kuokka and L. Harada, "Integrating Information via Matchmaking," Jurnal of Information Systems, O, pp. 101-121, 1996.
- [4] M. Schwartz and D. Wood, "Discovering Shared Interests Using Graph Analysis," Communications of The ACM, Vol. 36, No. 8, pp. 78-89, August 1993.
- [5] 緒方、相曾、矢野、"分散型人脈支援システムにおける人脈データベースの構築"、平成 8 年度科学研究費重点領域研究、メディアおよび環境統合のための高機能データベースシステムの研究開発、pp. 109-114, 松江, 1996.