

SNSにおける関係形成原理-mixiのデータ分析-
松尾豊、安田雪：How Relations are Built within a SNS World-
Social Network Analysis,人工知能学会論文誌, 22巻5号G,
pp531-541, 2007

2010/10/18 論文ゼミ#5 B4 斎藤有紗

内容

* 概要 * 論文

1.はじめに

2.分析に用いたデータ

3.友人関係の構造

3-1次数分布 L,C

3-2中心性上位のネットワーク

4.コミュニティの構造

5.友人関係とコミュニティ

5-1 友人の連鎖か、コミュニティでの出会いか

5-2コミュニティ内の友人形成

6.議論

7. 関連研究

8.まとめ

* 自分の研究との絡み

1. はじめに

* SNSの特徴

- 連鎖的にユーザーが増える・・・システム上でユーザが次々に新しいユーザを招待してユーザーが増える

- 見えない連鎖の具現化・・・現実の人間関係は見えない連鎖だが、限定的であるが具現化できる

→どのような人とどのようなコミュニケーションを行っているのか？どんなネットワークの形成原理が働いているのか？どのような関係構造が形成されるのか？

* SNSに関する研究

- 湯田、

→二者関係の承認関係の集積としてのホールネットワークの特徴の記述・・・スケールフリー性と高い凝集性、抽出したクラスタの大きさにおけるZipf則とのギャップ

•研究の概要

mixiを対象として社会ネットワーク分析を用いて、個々のユーザと全体の関係の相対的な関係についての分析・・・ネットワークの特徴、中心性の高いユーザの関係、ユーザの関心の共通性を表すコミュニティの相互作用

2. 分析に用いたデータ

データ

ユーザ数:363,819人

紐帯総数:3,813,702本(湯田の研究では約190万本の方向無しの紐帯)

一人あたりの平均友人数:10,48人

コミュニティ数:90,795個

コミュニティに一つでも参加しているユーザ数:241,423人

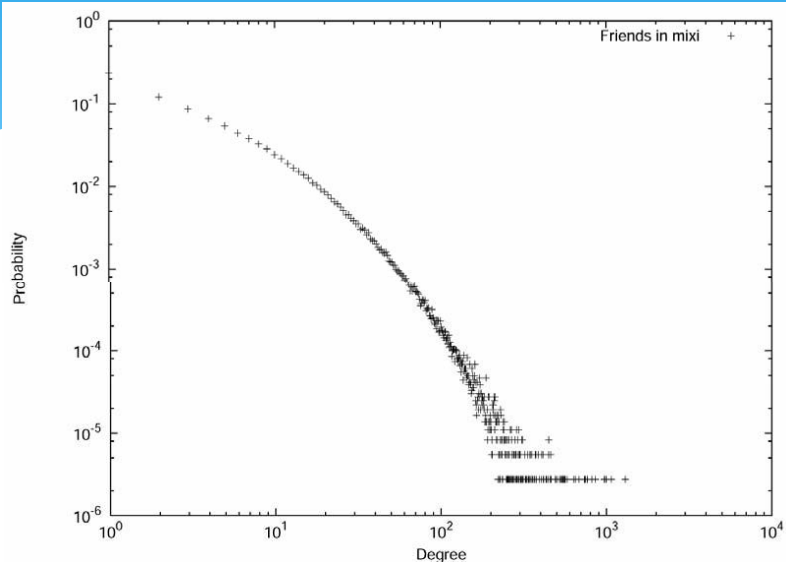
* 定義

紐帯・・・知人をmixiに招待したり、mixi上で友人を見つけることによって増加する

友人関係・・・相互承認を持って作られたマイクシィの関係を、この論文では友人関係とする

コミュニティ・・・ユーザが自由に開設でき、設定したテーマに関心のあるユーザが参加し交流する場

3-1、次数分布 L, C



x軸:対数変換した友人数

y軸:対数変換した確率

●全体の23.6%が友人数1人

●全体の35.7%が2人以下、44.3%が3人以下

次数分布

→友人関係構造がスケールフリー性を持つ

(べき乗の係数 γ は2.4程度)

※スケールフリー性の実世界の友人関係との比較は困難

3-1、次数分布 L, C

* 友人関係のネットワーク

•1215のコンポーネント、最大のものは360,801人、2番目のもののノード数は16

* CとL

•Cは0.328、Lは5.528

* CとLの推移

•ランダムグラフやSmall Worldでは、ノード数nに対してLは $\log(n)$ のオーダーで増加する

→① $0.295\ln(n) + 1.543$ (決定係数0.963)と当てはまる

②Small Worldの性質のため人数が増え
ても、クラスタ係数は0.3を下回らない

③Lも当初より2ステップほど増えているだけ
→規模の拡張があっても

局所的凝集性と短いパス長という特徴維持

人数	C	L	最大コンポーネント
2,500	0.394	3.951	2,466 人
5,000	0.382	4.082	4,933 人
10,000	0.378	4.279	9,895 人
20,000	0.372	4.454	19,798 人
30,000	0.366	4.524	29,694 人
50,000	0.353	4.634	49,491 人
70,000	0.349	4.729	69,319 人
100,000	0.344	4.847	99,050 人
150,000	0.337	4.998	148,694 人
200,000	0.334	5.124	198,269 人
363,819	0.328	5.528	360,801 人

3-2 中心性上位のネットワーク

① mixi 内部に多数の友人関係を持つ中心的なユーザのみを抽出したネットワーク

- * 301人以上の友人を持つユーザは99人(全体の0.027%) = ネットワークのコア
- * 内部に大きく2つのクリークが存在する. 2つのクリークを連結しているのは3つのダイアド
- * 99名が形成するネットワークのCは0.382であり, Lは2.74. → 中心的なメンバーの間でもかなり密な関係が形成されている

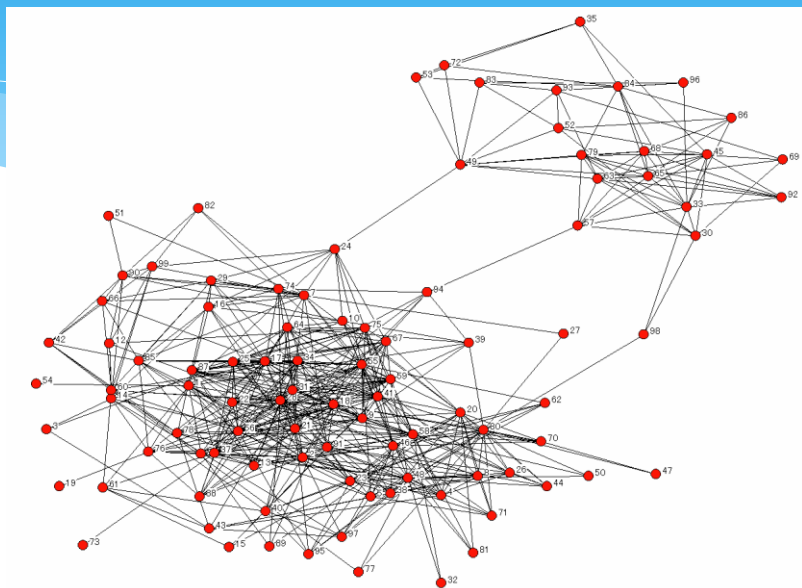
(この凝集性の高さは, 2つのクリークそれぞれの内部での凝集性, 2つのクリークの連結はいことに留意すべき)

② 201人以上の友人をもつユーザを抽出したネットワーク

- * mixi 内では365人(全体の0.100%)
Cは0.323でありLは2.877

③ 101人以上の友人を持つユーザは2564人, Cは0.233, Lは3.139となる.

→ 中心的ユーザが形成する大きな二つのクリーク構造が存在するという特徴



4.コミュニティの構造

コミュニティ

ユーザの関心により成長するのがコミュニティ。⇔友人関係は承認の連鎖として成長。

- * 多数のユーザの関心を集めたコミュニティは拡大, 類似した関心を持つユーザは共通のコミュニティに集まる。
- * コミュニティの共通所属は友人関係の契機になり, コミュニティ相互の関連性を形成。
- * **メンバー数の上位20位までのコミュニティ**
- * 「資料になりそうなウェブサイト」「Mac ユーザー」といった目的や対象のはっきりしたコミュニティと、「まったくわかりません」「空を見る人」などの漠然とした名前のコミュニティが上位には混在。

コミュニティ名	人数	中心性*
1 資料になりそうなウェブサイト	10,238 人	1
2 Mac ユーザー	8,826 人	3
3 まったくわかりません!	7,787 人	6
4 Photoshop	7,407 人	2
5 クリエイター・デザイナー	7,163 人	5
6 iPod User's	7,061 人	11
7 面白ネタで笑おう!	7,022 人	4
8 美術館・博物館展示情報	6,824 人	13
9 空を見る人	6,581 人	7
10 料理作るのが好き	6,467 人	15
11 笑える画像	5,828 人	12
12 フォント	5,516 人	8
13 にゃんこ組	5,342 人	38
14 O 型	5,337 人	27
15 Illustrator	5,309 人	9
16 ヴィレッジヴァンガード	5,107 人	18
17 名前覚えられません	4,892 人	16
18 めんどくさい	4,850 人	10
19 カレー大好き	4,849 人 ²²	
20 水曜どうでしょう	4,831 人 ¹¹⁸	

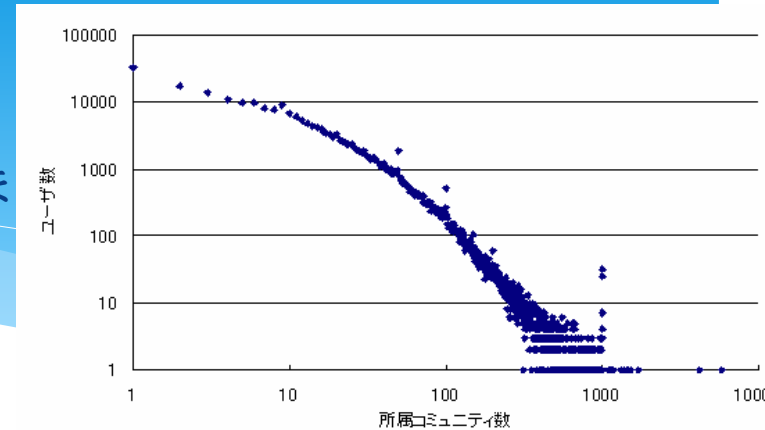
4. コミュニティの構造

コミュニティへのユーザの参加

各ユーザの所属コミュニティ数の平均は24.97 個

所属コミュニティの数が増えるにしたがって人数は減少. 所属コミュニティが数十を超えるあたりからは両対数グラフではほぼ直線状の分布。

- * 所属コミュニティ数が50, 100, 200 などのプロットや1000 近辺でのずれ→所属コミュニティの上限を1000 としているため



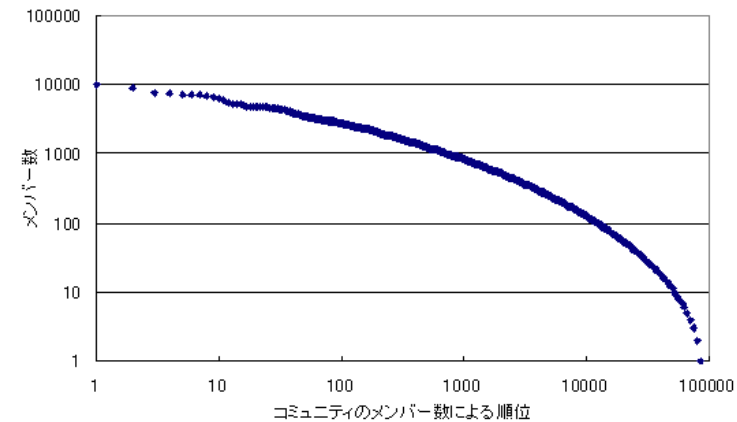
コミュニティのメンバー数の分布

- * 各コミュニティのメンバー数の平均は66.40 人

- * 所属人数が最大のコミュニティは10238 人, 所属者1 人だけのコミュニティは5395 個, メンバー数が1000 を超えるものは全体の7.9%

- * 両対数をとったときにこの分布は直線にならない→Zipf 則ではない。

- * 10 人以下のコミュニティはZipf 則を仮定した場合よりも少なくすなわちロングテール部分がない. →コミュニティという性質上, ある程度の規模までは情報交換しやすく発展しやすいからか?



4.コミュニティの構造

コミュニティの時系列の順序関係

→ (i)古いコミュニティほどメンバー数が増える傾向があり、その上限は時間によっておおよそ制約されている、

(ii)同じ時期のコミュニティでも、メンバー数が非常に多いものが存在する。

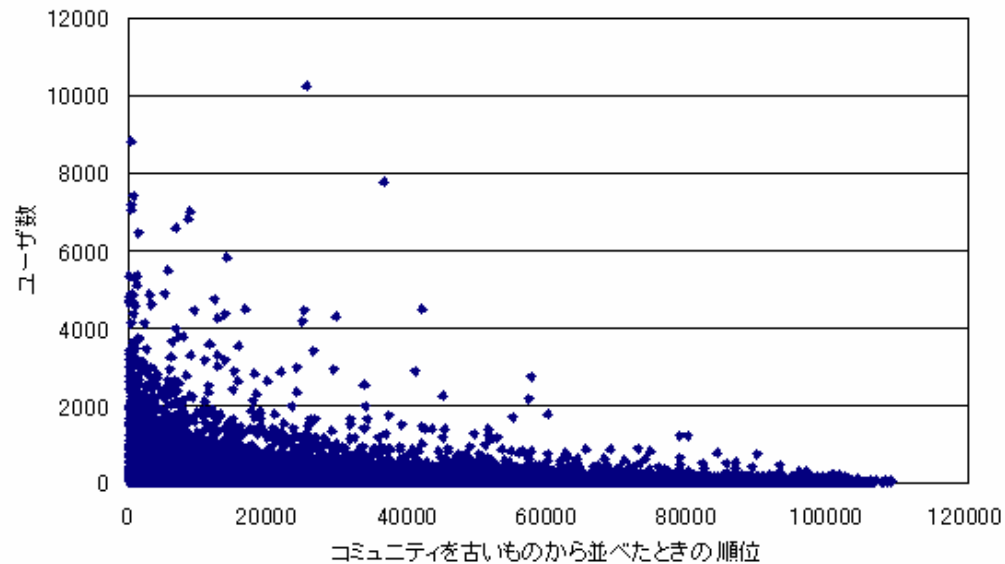
コミュニティ間の関連

- * メンバー数の上位200位までのコミュニティで、メンバー数は最大10238人から最小で2016人(mixi全体のコミュニティの延べ人数の10.6%)
- * ユーザとコミュニティの所属関係は二分グラフを形成するが、これをコミュニティ間の関係性に縮退させ、ネットワークを抽出する。

⇒ある二つのコミュニティに共通に所属するユーザが多いほど、その2つのコミュニティの関連は強い(コミュニティ連関ネットワークと呼ぶ。)

- * コミュニティ間の関連度はJaccard係数を用いて計算

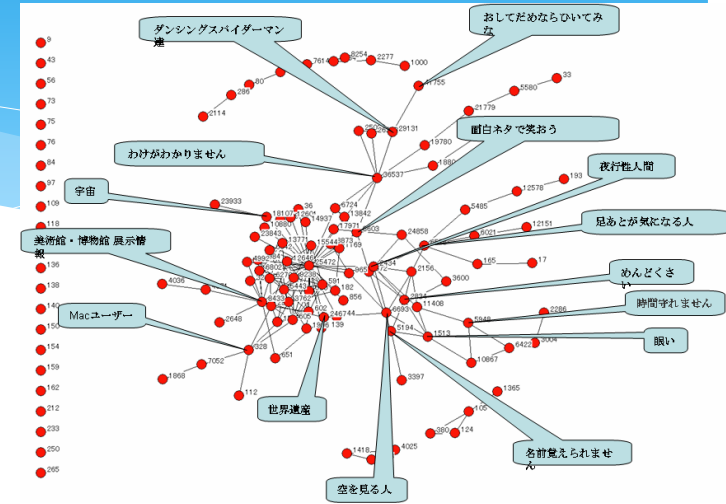
コミュニティCとコミュニティDに共通して所属するユーザ数を $n(C \cap D)$ 、CとDのいずれかに所属するユーザ数を $n(C \cup D)$ とすると、コミュニティCとDの関連度は $Jaccard(C,D) = n(C \cap D) / n(C \cup D)$ と計算される。



4.コミュニティの構造

コミュニティの連関ネットワーク (Jaccard 係数が 0.2 以上の紐帯のみを表示)

- * 21 個のコミュニティが孤立点となるが、その他の179 個のコミュニティは連結したコンポネントを形成
- * 一般的な話題から次第にテーマが深堀されていく構造ネットワークでの中心性 (固有ベクトル中心性)
- * メンバー数の多いコミュニティは中心性も高い傾向。
- * 中心性が低くてもメンバー数が上位20 位までに入っているものの存在→連関ネットワークの中心からやや外れたところで、新たにユーザを引き付けつつあるコミュニティ。

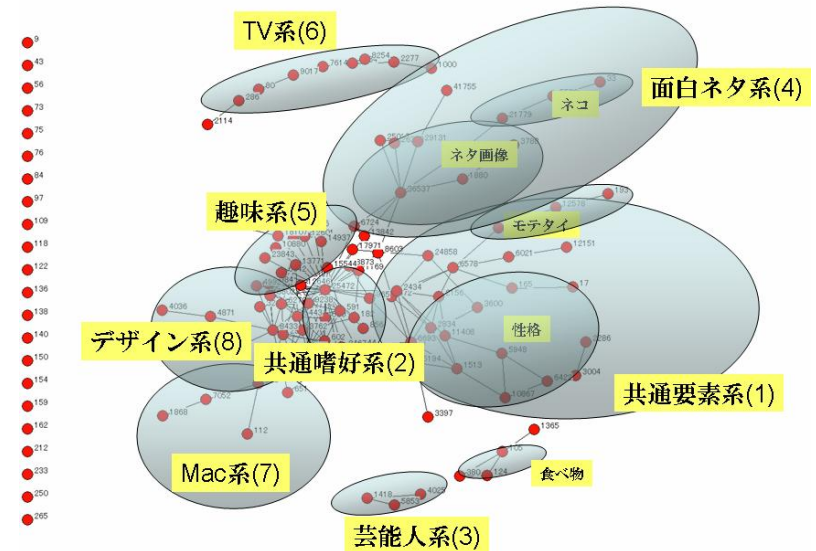
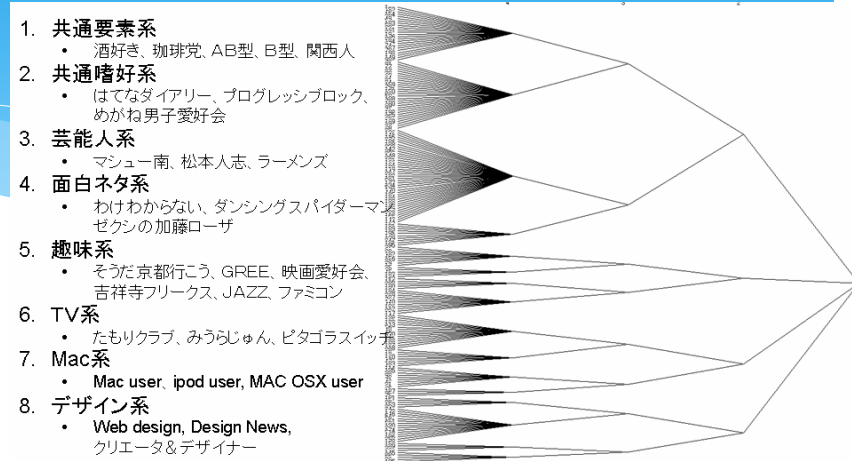


コミュニティ名	人数	中心性*
1 資料になりそうなウェブサイト	10,238 人	1
2 Mac ユーザー	8,826 人	3
3 まったくわけがわかりません!	7,787 人	6
4 Photoshop	7,407 人	2
5 クリエイター・デザイナー	7,163 人	5
6 iPod User's	7,061 人	11
7 面白ネタで笑おう!	7,022 人	4
8 美術館・博物館展示情報	6,824 人	13
9 空を見る人	6,581 人	7
10 料理作るのが好き	6,467 人	15
11 笑える画像	5,828 人	12
12 フォント	5,516 人	8
13 にゃんこ組	5,342 人	38
14 O 型	5,337 人	27
15 Illustrator	5,309 人	9
16 ヴアレージヴァンガード 5,107 人	18	
17 名前覚えられませぬ 4,892 人	16	
18 めんどくさい 4,850 人	10	
19 カレー大好き 4,849 人	22	
20 水曜どうでしょう 4,831 人	118	

4.コミュニティの構造

ブロックモデルという手法によりクラスタリングし、ラベル付けしておおまかに分類⇒8分割の時点で、共通要素系、共通嗜好系、芸能人系、面白ネタ系、趣味系、TV系、Mac系、デザイン系の8つに分けられる。

* それぞれがネットワーク図のどのあたりの領域であるか⇒コミュニティは、このような領域をカバーしながら相互に関連の構造



5-1 友人の連鎖か、コミュニティでの出会いか

友人関係とコミュニティ

- * 友人登録とコミュニティへの参加が相互発展的なので、友人関係と共通のコミュニティがあるかどうかはある程度相関
- * 共通の友人がいるのは、82% 共通のコミュニティがあるのは54% ⇒ 友人を介しての新たな友人関係の方が多い
- * 友人関係のある・なしと、共通のコミュニティのある・なしの分割表

⇒ 友人関係があるー共通のコミュニティがあるという正の相関が明確(あまりに全体の数が多いので、友人関係がないー共通のコミュニティがないという負の相関は読み取りにくい)

- * ⇒ コミュニティによっては友人関係が多いものとそうでないものがあるのだろうか？

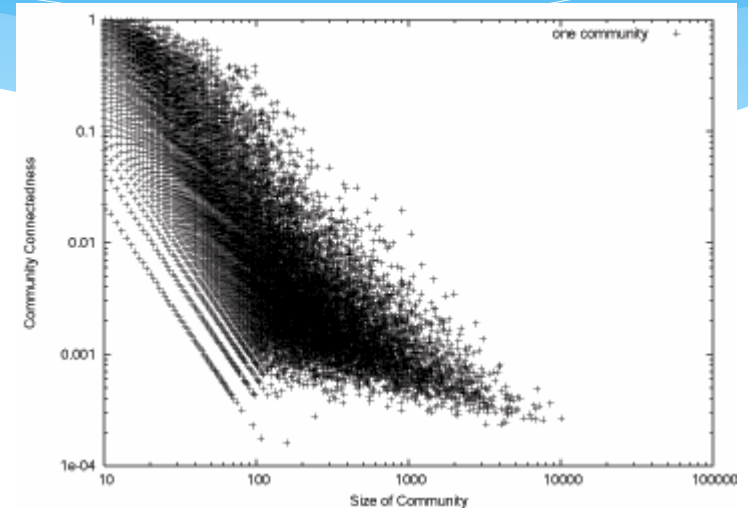
→ n 人から成るコミュニティであれば、最大で $n(n-1)/2$ 組の友人関係が存在し得る。このうち、実際に友人関係が存在する割合を、本論文では**コミュニティの結合性**(CC: community connectedness)と呼ぶ(コミュニティに所属するメンバー内でのネットワークの密度である。結合性が1であれば、コミュニティ内の全てのメンバーは友人関係であり、0であればどの2人をとっても友人関係でない。)

	共通コミュあり	なし	計
友人関係あり	$2.04 \times 10_7$	$1.77 \times 10_7$	$3.81 \times 10_7$
なし	$2.80 \times 10_{10}$	$1.30 \times 10_{12}$	$1.32 \times 10_{12}$
計	$2.80 \times 10_{10}$	$1.30 \times 10_{12}$	$1.32 \times 10_{12}$

5-2コミュニティ内の友人形成

コミュニティのサイズと結合性

- * コミュニティのサイズが大きくなると一般的に結合性は弱くなる。→各ユーザの平均友人数が約10人であることから考えると当然。
- * 結合性はおよそ n に比例して小さくなる。同じサイズのコミュニティでも、結合性が大きなものもあれば小さなものもある。→右上に近いものがサイズの割に結合性が高い。



5-2コミュニティ内の友人形成

結合性の強いコミュニティ

- * 結合性CCをnで割った $f=CC/n$ を指標とし、その値の上位5位までのコミュニティ。
- * リアルな世界でのインタラクションがあるコミュニティ。つまり結合度が高いのは、リアルな空間での関係とmixi上でのコミュニケーションが融合し、うまくサイクルとなって回っているコミュニティであった。

結合性の弱いコミュニティ

- * 結合性は上位5位と比べて1000倍以上の大きな差。
- * ユーザの趣味や興味の対象であり、かといって誰も共感できるものではなく、友人関係の契機となるのが難しいコミュニティである。

⇒mixiのコミュニティ機能は、ユーザの多様なニーズに応じてうまく機能しており、結果的に独特の生態系を形作っている

コミュニティの概要	人数	結合性
1 Keio SFC のダンス系のコミュニティ	97	0.377
2 鍋と焼肉の集まり	97	0.334
3 割烹のお店の集まり	56	0.539
4 ホームパーティのコミュニティ	84	0.356
5 ある人の家でときどきご飯を食べる会	45	0.656

コミュニティの概要	人数	結合性
1 常盤貴子	107	0.000176
2 バイオハザード4(ゲーム)	94	0.000229
3 カレーうどん 80	80	0.000316
4 マカロニほうれん荘(漫画)	158	0.000161
5 鷺沢萌(作家)	74	0.000370

8.まとめ

- 判明したこと:
1. L が5.5と低いのに、 C は0.33と高い
 2. 中心人物同士のネットワークは2つのク
リークに分かれていて、クリークは疎である
 3. コミュニティへの加入のネットワークは個性的なものへ
とのちに導く、入りやすいものが存在する

5. 自分の研究との絡みについて

用語

- * **L:平均距離**
- * **C:クラスター係数**
- * **ジップ則:出現頻度が k 番目に大きい要素が全体に占める割合が $1/k$ に比例するという経験則**