

# mixi における ソーシャルネットワーク分析

# 目次

**1. mixi ネットワークの分析**

**2. mixi の取り組み**

**3. 今後の課題**

# 1. mixiネットワークの分析

## mixiとは

- ・インターネット上で友人ネットワークを形成し、**日記**や**コミュニティ**等を介して友人同士がコミュニケーションを行うためのサービス
- ・収入源は主に**広告収入**

# ブログや掲示板との違い

## ・クローズドな世界

- ・大部分のコンテンツはyahoo 等で検索できない

## ・親しい友人とのコミュニケーション

- ・友人のコンテンツが見やすい
- ・誰がアクセスしたかがわかる(足あと)
- ・日記にコメントがあればお知らせ

# mixi 基礎データ

<b>ユーザー数</b> (2008年7月13日現在)	<b>1500万人</b>
<b>アクティブ率</b> (2008年6月現在)	<b>約55%</b>
<b>月間PV</b> (2008年6月現在)	<b>136.7億PV</b>
<b>男女比 (男:女)</b> (2008年6月30日現在)	<b>47.3 : 52.7</b>
<b>30歳未満のユーザー</b> (2008年6月30日現在)	<b>66.5%</b>

# マイミクシィとは

- **誰と誰が友人関係にあるかを示す**
  - マイミクシィ申請のメッセージを送り、  
相手が承認するとマイミクシィになる
  - 最大マイミク数は1000人
  - マイミクになった人の日記等が、  
個人のトップページに表示

# マイミクシ基礎データ

ノード数	約1500万
リンク数 <small>(マイミク関係を片方向のリンク×2とする)</small>	約3億2700万
平均マイミク数	22.02人
平均クラスタ係数	0.196

クラスタ係数:あるノードに隣接するノードの内の2個を選んだときに、  
その2個が隣接しているかどうかの割合

## 密なネットワークを形成

# スモールワールド性とは

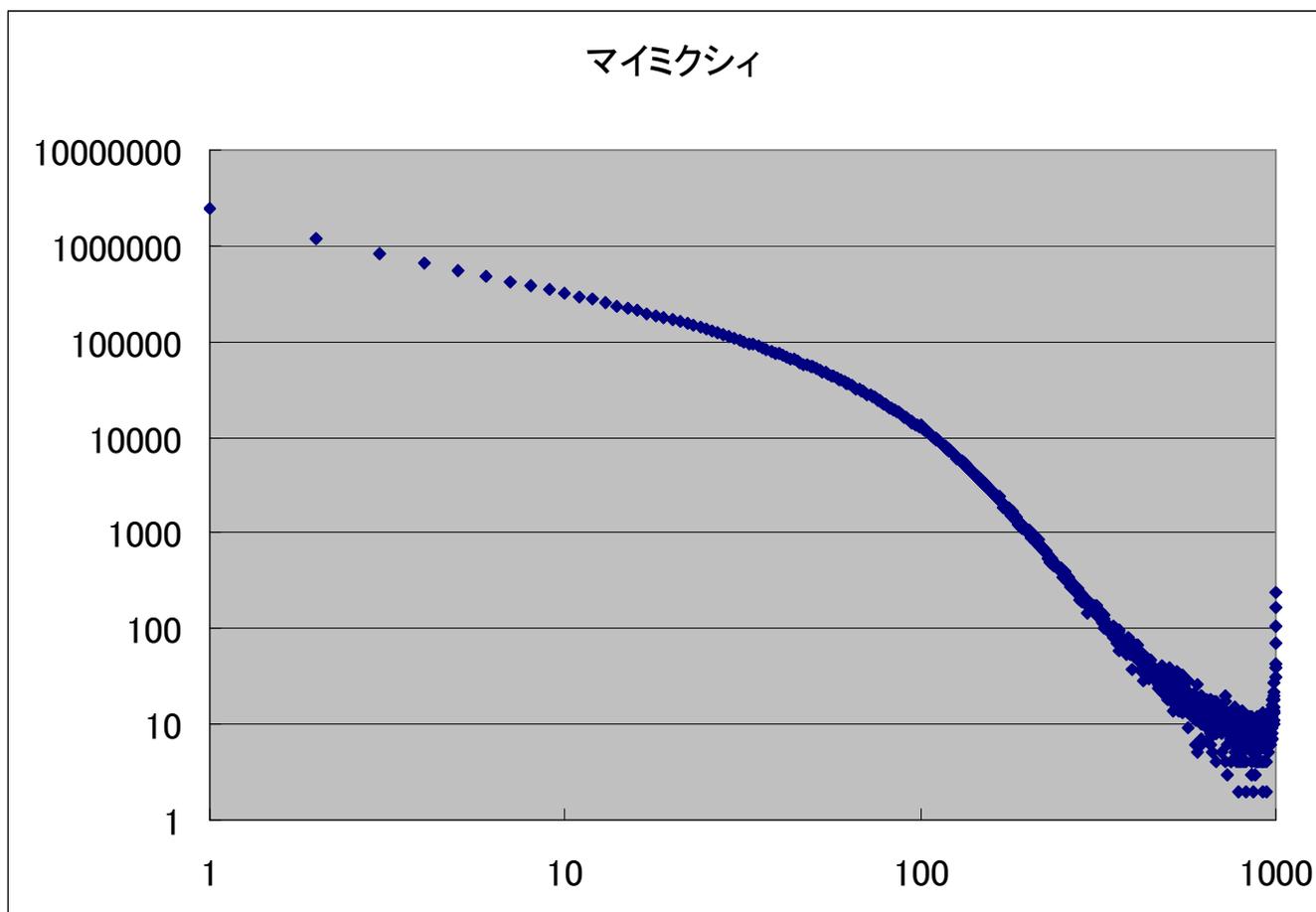
- **世間は狭い** (It's a small world!)
- **6次の隔たり**
- **スモールワールドの特徴**
  - **誰に対しても少ないホップ数で到達できる**  
(平均距離が小さい)
  - **自分の友人同士が友人であることが多い**  
(クラスタ性が高い)

# mixiのスマートフォンワールド性

距離	その距離のユーザ数	到達できるユーザ数	到達できるユーザの割合 (%)
0	1	1	0.0000
1	37	38	0.0003
2	1103	1141	0.0087
3	58042	59183	0.4519
4	1896588	1955771	14.9341
5	7524544	9480315	72.3910
6	3046429	12526744	95.6533
7	338875	12865619	98.2410
...	...	...	...
161	0	12912255	98.5971

距離が6の場合95.7%、距離が7では98.2%に到達  
 平均パス長は5.4

# マイミクシィ分布



**2割のユーザーで65%程のリンクを所有**

# コミュニティとは

- **関心のある事について話す掲示板**
  - コミュニティ作成は誰でもできる
  - 基本的に誰でも参加できるが、  
参加を承認制にすることもできる
  - 一人あたり参加数は最大1000個
  - コミュニティあたり参加数は無制限

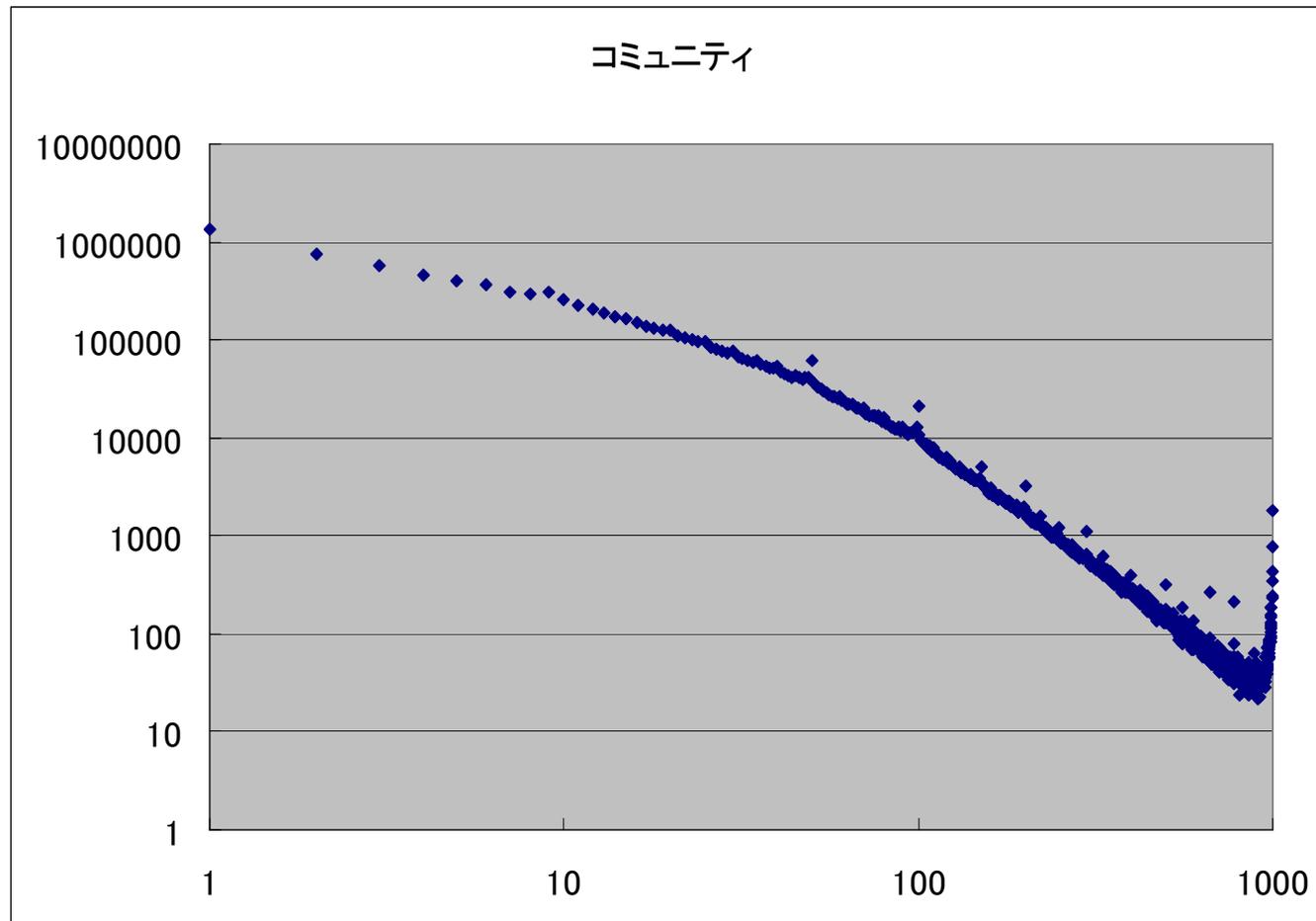
# 大きなコミュニティの例

- mixiで使える絵文字
- You Tube
- すぐ顔に出る
- なにかいいことがありますように
- よく物をなくす
- 音楽が無いと生きていけない
- 期限ギリギリまで行動できない
- 空を見る人
- 心に響いた名言集

# コミュニティ基礎データ

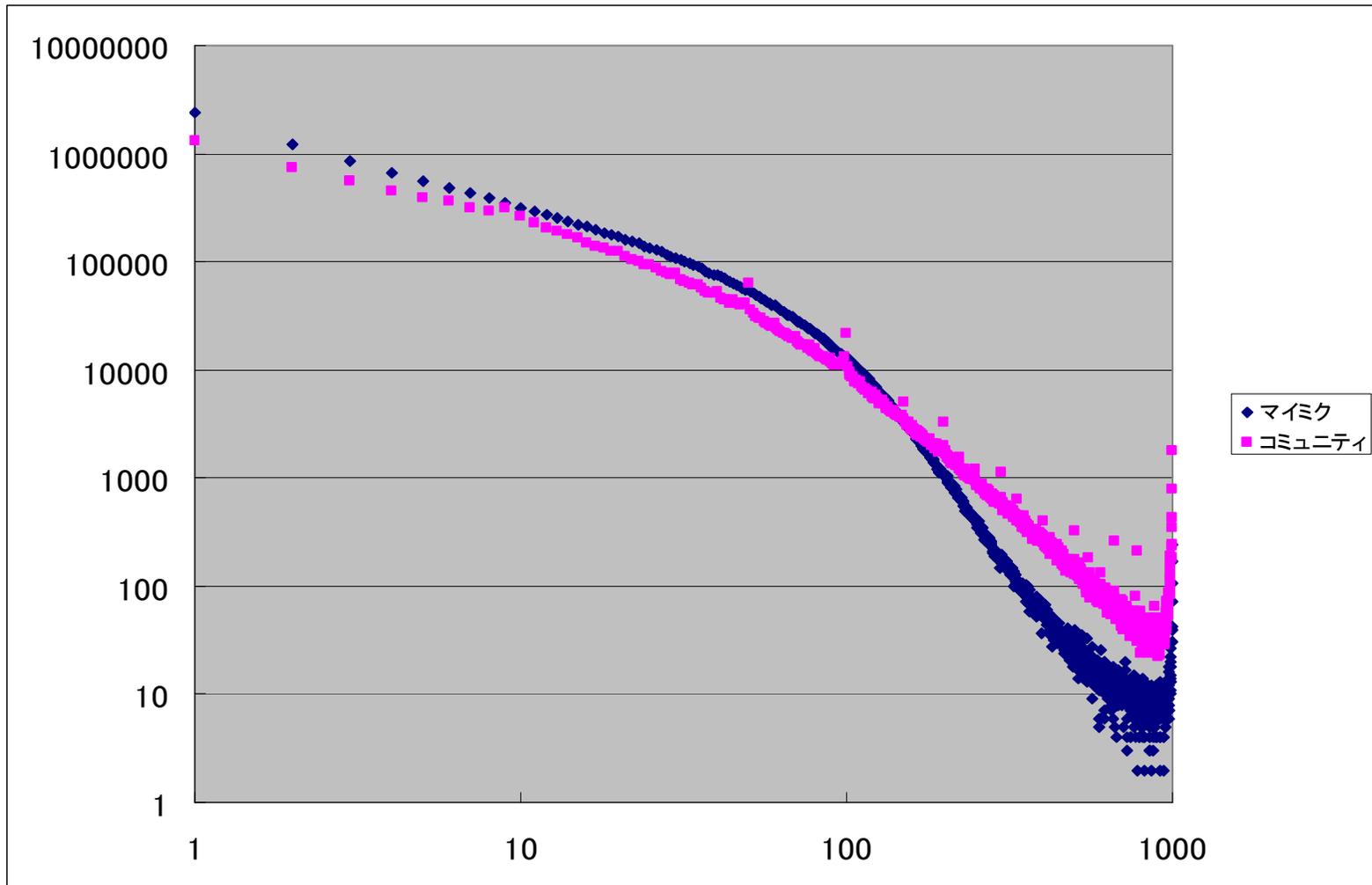
ノード数	約200万
リンク数	約3億200万
平均コミュニティ参加数	20.31個
最大のコミュニティ	約43万人

# コミュニティ分布



**2割のユーザーで80%のリンクを所有**

# 比較



## 2. mixiの取り組み

# mixi の取り組み

- 1. おすすめマイミクシィ**  
(本サービス提供中)
- 2. おすすめコミュニティ**  
(試験サービス提供中)
- 3. コミュニティブラウザ**  
(本サービス提供中)

# おすすめマイミクシィ

## • 個人ごとにマイミクシィをおすすめ

- common neighbors

共通のマイミク数

- jacarrd's coefficient

共通マイミク数  $(A \cap B)$  / ユニークマイミク数  $(A \cup B)$

- preferencial attachment

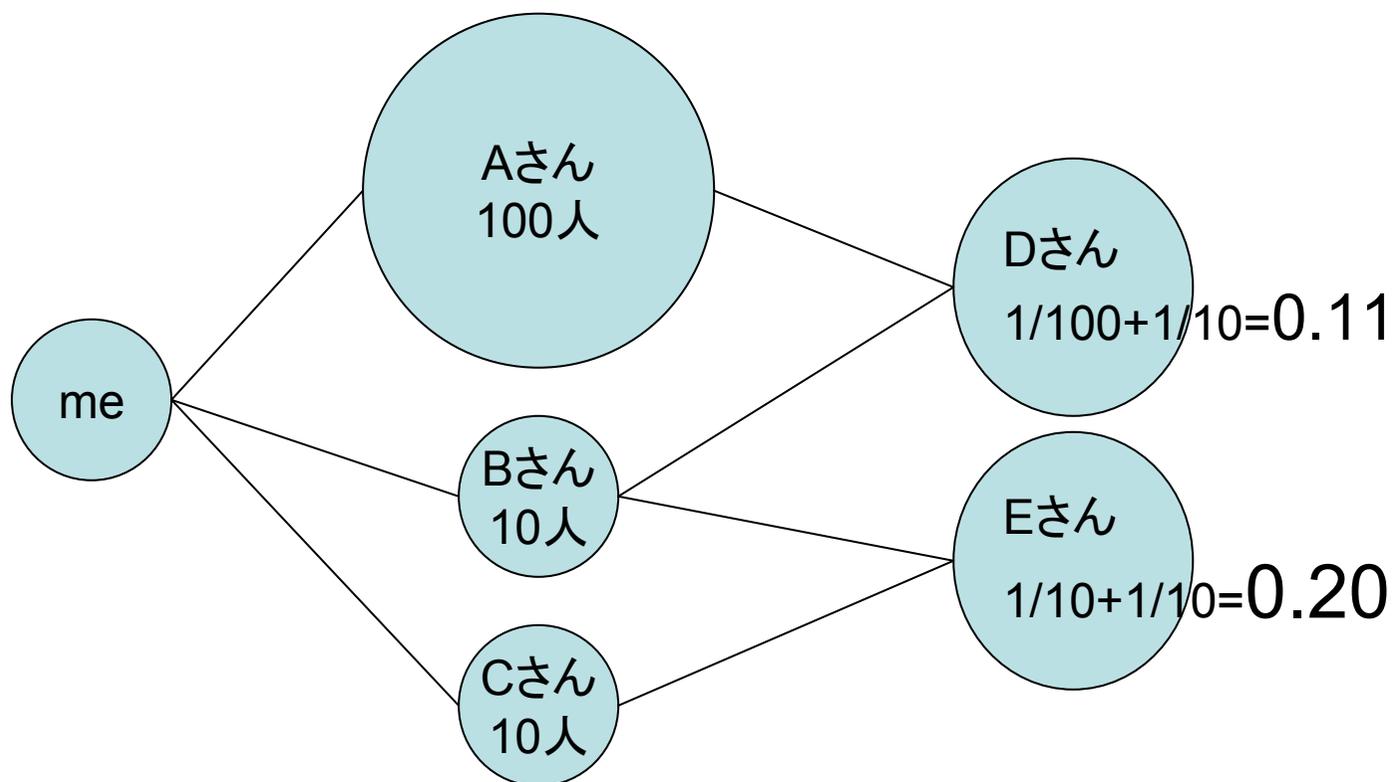
マイミクが多い人同士ほどマイミクになりやすい

- Adamic / Adar

$\sum 1/n$  (nはマイミク数)

# 仕組み

Adamic/Adar:  $\Sigma 1/n$  (nはマイミク数)



**マイミクが少ないほど付合いが深い**

# おすすめコミュニティ

- 個人ごとにコミュニティをおすすめ

- common neighbors

- 共通のマイミク数

- Adamic / Adar

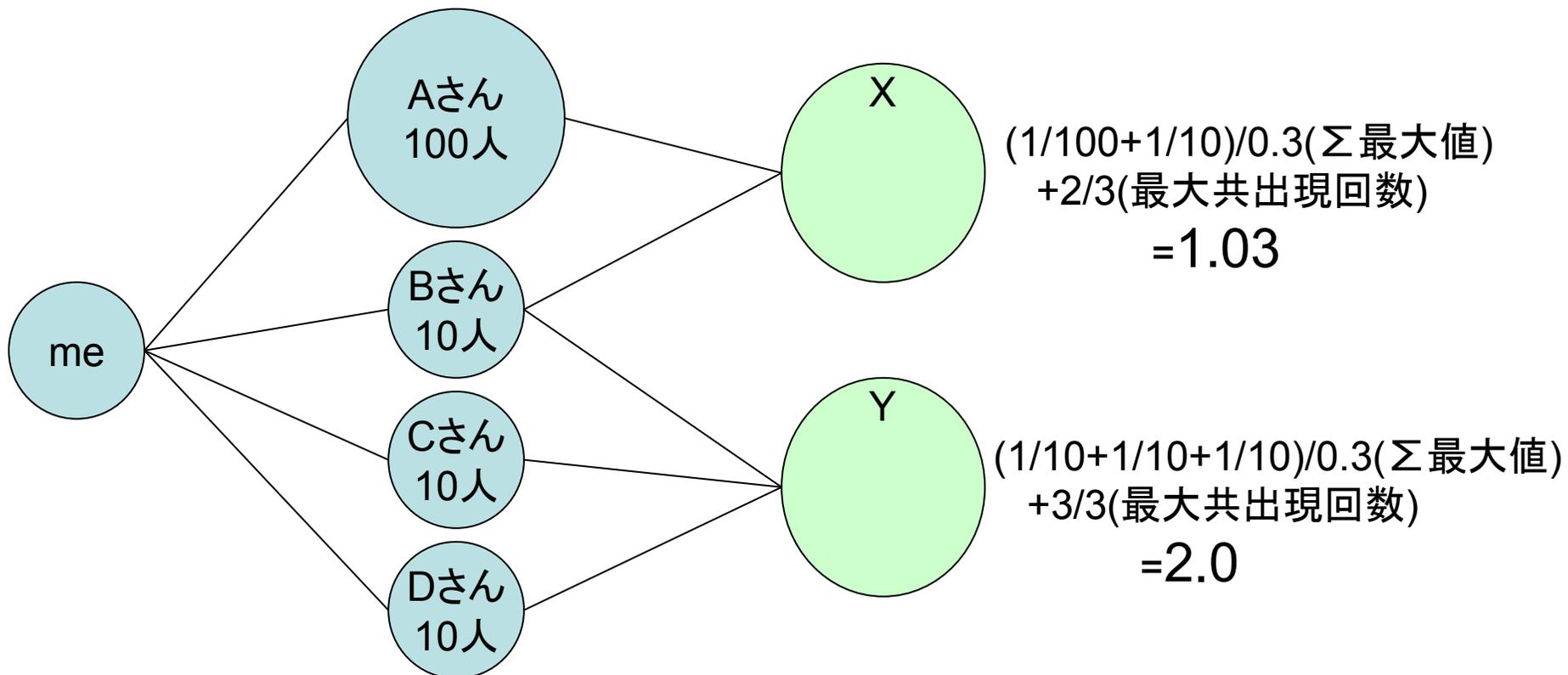
- $\Sigma 1/n$  (nはマイミク数)

を、影響が等価になるように、

それぞれの最大値で割って足す

# 仕組み

$\Sigma 1/n$ と共出現回数を比率1:1で集計

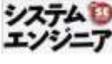


**マイミクに共通する + 内輪のコミュニティ**

# コミュニティブラウザ

コミュニティブラウザ インディーズ掲載

コミュニティブラウザ > システムエンジニアの部屋



**システム  
エンジニア**

**システムエンジニアの部屋**

カテゴリ : PG、インターネット

メンバー数 : 8364人

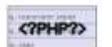
日本にしか無いらしい謎な職業システムエンジニア。みんなから期待されている仕事の範囲は半端じゃない！客と折衝したり〜。仕様書書いたり〜。プログラム組んだり〜。デバッグしたり〜。新技術を追い求めたり〜。なぜか社内でのパソコンのメンテナンスまでしたり〜。そんなシステムエンジニア

[もっと見る](#)

**オススメコミュニティ**

管理人:

選択したコミュニティに関連するオススメのコミュニティが自動で表示されます。関連性が高いほど、星印が表示されます。

 Java(9644) ★★★★★	 ネットワークエンジニア(625) ★★★★★	 ネットワーク&セキュリティ(10682) ★★★★★	 情報処理技術者試験(8112) ★★★★★	 迷えるSEの仲間たち(2170) ★★★★★
 [dir] プログラミング(7447) ★★★★★	 C言語とC++言語(9180) ★★★★★	 Eclipse(8494) ★★★★★	 Linux(7220) ★★★★★	 C#(3149) ★★★★★
 Let's PHP(7308) ★★★★★	 Java質問 & 情報提供サイト(3209) ★★★★★	 SI・IT業界人脈ネットワーク(3645) ★★★★★	 DBならOracleでしょ♪(2441) ★★★★★	 MySQL(8768) ★★★★★
 Perl(4781) ★★★★★	 Webプログラミングについて語る(3623) ★★★★★	 Visual Basic User's Club(3390) ★★★★★	 [dir] UNIX系OS(3264) ★★★★★	 Ajax - Async JavaScript w/ XML(5177) ★★★★★

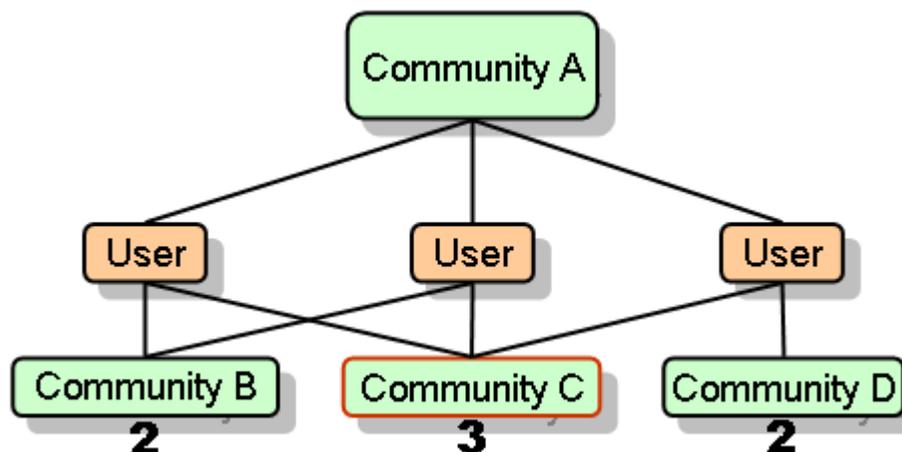
# コミュニティブラウザ

- **関連するコミュニティをおすすめ**

- **協調フィルタリング**

コミュニティAのユーザが参加している

コミュニティBはコミュニティAに関連性が高い



# 参考論文

## Link Mining: A Survey

<http://www.sigkdd.org/explorations/issues/7-2-2005-12/1-Getoor.pdf>

## The Link-Prediction Problem for Social Networks

<http://www.cs.carleton.edu/faculty/dlibenno/papers/link-prediction/link.pdf>

## An Algorithmic Approach to Social Networks

<http://erikdemaine.org/theses/dliben-nowell.pdf>

## Friends and Neighbors on the Web

<http://www.hpl.hp.com/research/idl/papers/web10/fnn2.pdf>

# mixiならではの問題 1

- **データ量がとにかく多い**

予めデータを作っておくとして、

1人あたりの処理に

0.5秒かかったとしても・・・

**1500万人 × 0.5秒 = 750万秒 = 87日**

## mixiならではの問題 2

- **アクセス数がとにかく多い**

**136億7千万PV／月** (2008年6月現在)

**=4億6千万PV／日**

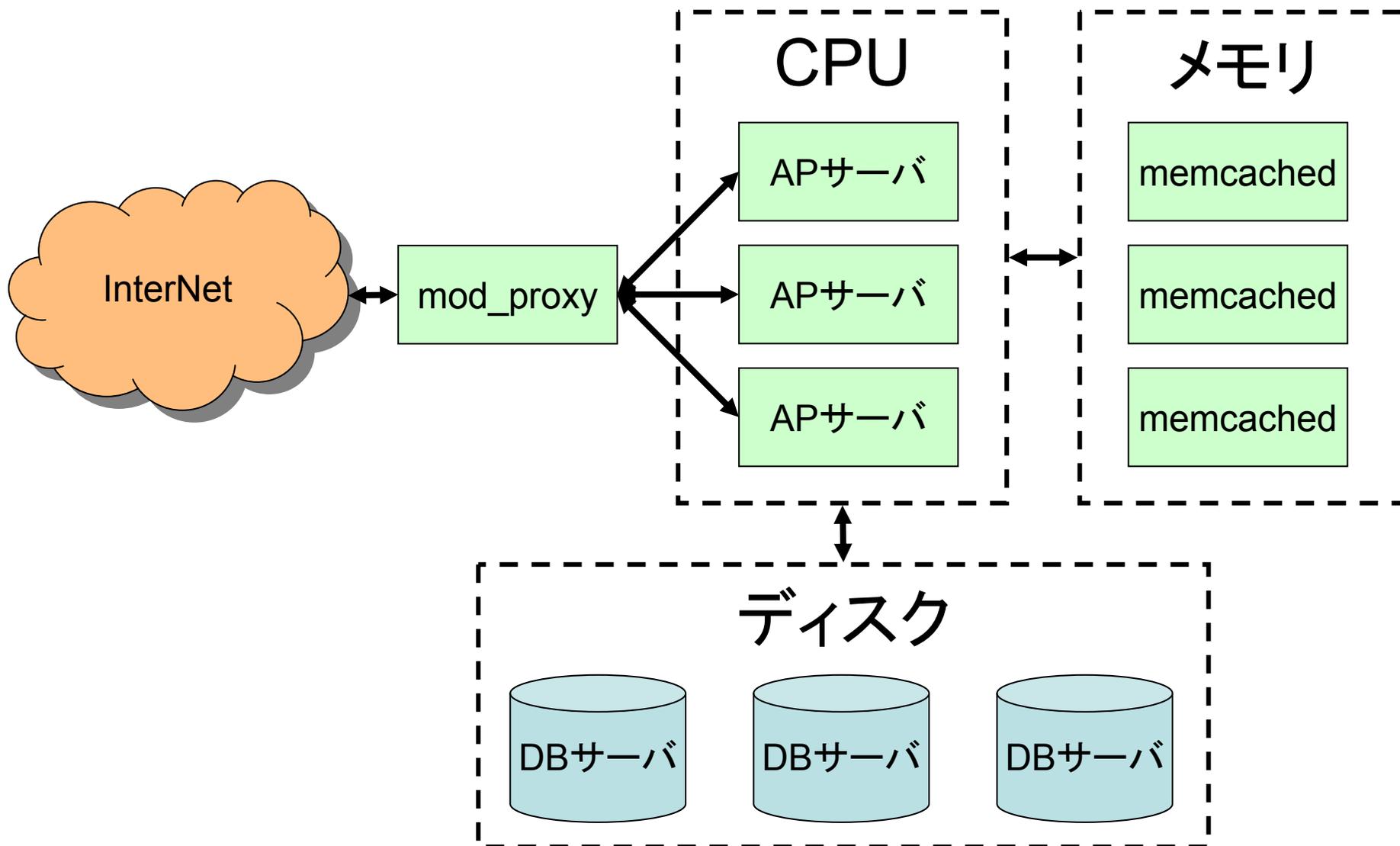
**=1900万PV／時**

**=31万PV／分**

**=5300PV／秒** (実際のピークはもっと多い)

**国内4位 世界70位** (2008/10/9 alexa)

# mixiのシステム構成



# mixiのシステム構成

## ほぼオープンソース

- Apache mod\_perl (APサーバ)
- memcached (メモリ)
- MySQL (DB)
- Squid (画像配信用キャッシュ)
- Nagios/RRD (監視系)
- Fedora (OS)

負荷分散装置など一部は商用を使用

# DBのスケールアウト

- **Master-Slave**

データをレプリケーションさせる

書き込みはMasterへ読み込みはSlaveへ

- **機能別の分散**

日記DB・コミュニティDBなど

- **ID別の分散**

メンバーID別・コミュニティID別など

# その他のポイント

- **画像の保存と配信**

ファイル数とアクセス数の特性が違う  
数が少なくアクセスが多いもの  
数が多くアクセスが偏っているもの

- **携帯向けサービス**

動画変換・画像変換など

- **検索**

ユーザー検索・コミュニティ検索など

# レコメンドでは

- **バッチ処理**

あらかじめデータを作っておく

- **並列処理**

同時にできる処理は複数サーバで

- **データを小さく**

できるだけディスクにアクセスしない

データをある程度小さくする

# サービス提供上の諸問題

## • レコメンド

- おすすめしてほしくないものもあるので  
一覧から消す機能なども必要になる
- 知り合いでもマイミクになるのをためらう文化

## • ユーザビリティ

- 他のサービスも多数ある中での表示方法

## 3. 今後の課題

# mixi における課題 1

- **コミュニケーション対象の選択肢を増やす**
  - **マイミクシィを増やす**  
ユーザ検索／レコメンド
  - **参加コミュニティを増やす**  
コミュニティ検索／レコメンド
- **コミュニケーション活性化**  
ニュース／ミュージック／レビュー等

→各ユーザが処理する情報量を増大させる

## mixi における課題 2

### • 読むべき情報の整理

- 「面白い」情報とそうでないものを分ける
  - 一般的な面白さを定量化
  - 各ユーザの興味への適合を定量化
- 面白い情報の発信源を特定する
  - ユーザ毎、コミュニティ毎に集計
  - ネットワーク分析でオーソリティを推定

→ 面白い情報を選択してリアルタイムに提示

# 研究開発グループの取り組み 1

- 「読むべき情報の整理」にフォーカス
  - 主に検索とデータマイニング
- 検索エンジンの精度向上
  - かんたん友人検索 / コミュニティ検索
- ランキング系 = 面白さの定量化
  - キーワードランキング
  - ユーザランキング
  - コミュニティランキング

## 研究開発グループの取り組み 2

- **レコメンド** = 各ユーザの興味の推定
  - マイミクシィレコメンド
  - コミュニティレコメンド
  - 日記レコメンド
- **ミドルウェア開発**
  - 高効率DB  
(Tokyoシリーズ、memcached、drizzle)
  - 言語解析プラットフォーム  
(形態素解析、特徴語抽出)