DAI-Labor
TU Berlin

PIA+COMM

An Intelligent Search Engine Supporting Scientific Communities

Dipl. soc. tec. Michael Hahne
MA Corinna Jung
Dipl.-Inform. Jérôme Kunegis
Dipl.-Inform. Andreas Lommatzsch
Dipl.-Inform. André Paus

INFORMATIK 2006

Technische Universität Berlin
Zentrum Technik und Gesellschaft

DAI-Labor
TU Berlin

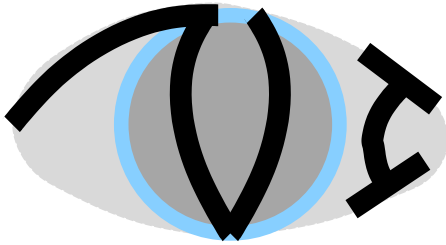
PIA+COMM – INFORMATIK 2006
Outline

- ⇨ **Introduction**
- ⇨ **Community Model**
- ⇨ **Community Supporting Features**
- ⇨ **Conclusion and Outlook**

Introduction

Introduction
Project Data

- ⇨ **Interdisciplinary research project including information scientists and sociologists**
- ⇨ **Project goal: an intelligent search engine with integrated support for scientific communities**
- ⇨ **Iterative system development process**
- ⇨ **Formative evaluation: continuous user questioning and user feedback integration**

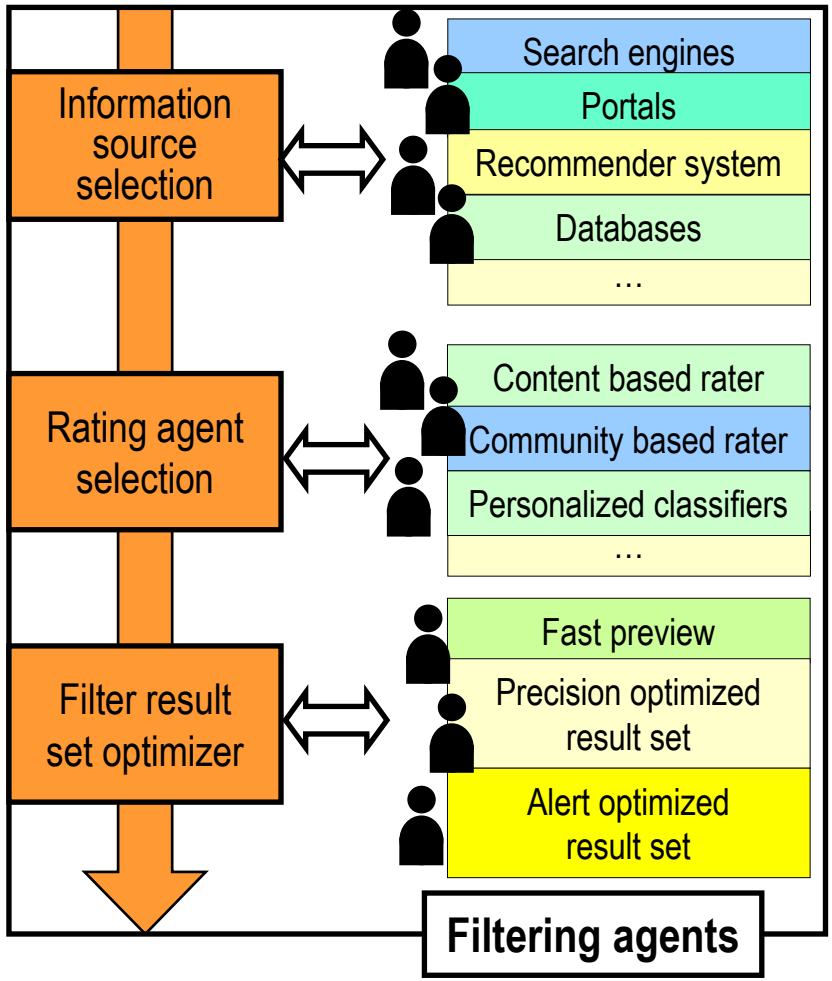


DAI-Labor
TU Berlin

PIA+COMM supports all aspects of information needs of scientific communities!

- ⇨ Supply of large, scientific information sources: databases, portals, multimedia content, …
- ⇨ Advanced information search and management
- ⇨ Advanced data analysis, feature extraction, meta data calculation for classification and indexing
- ⇨ Integration of community building and functionalities supporting communities

Introduction
System Components



Web – GUI

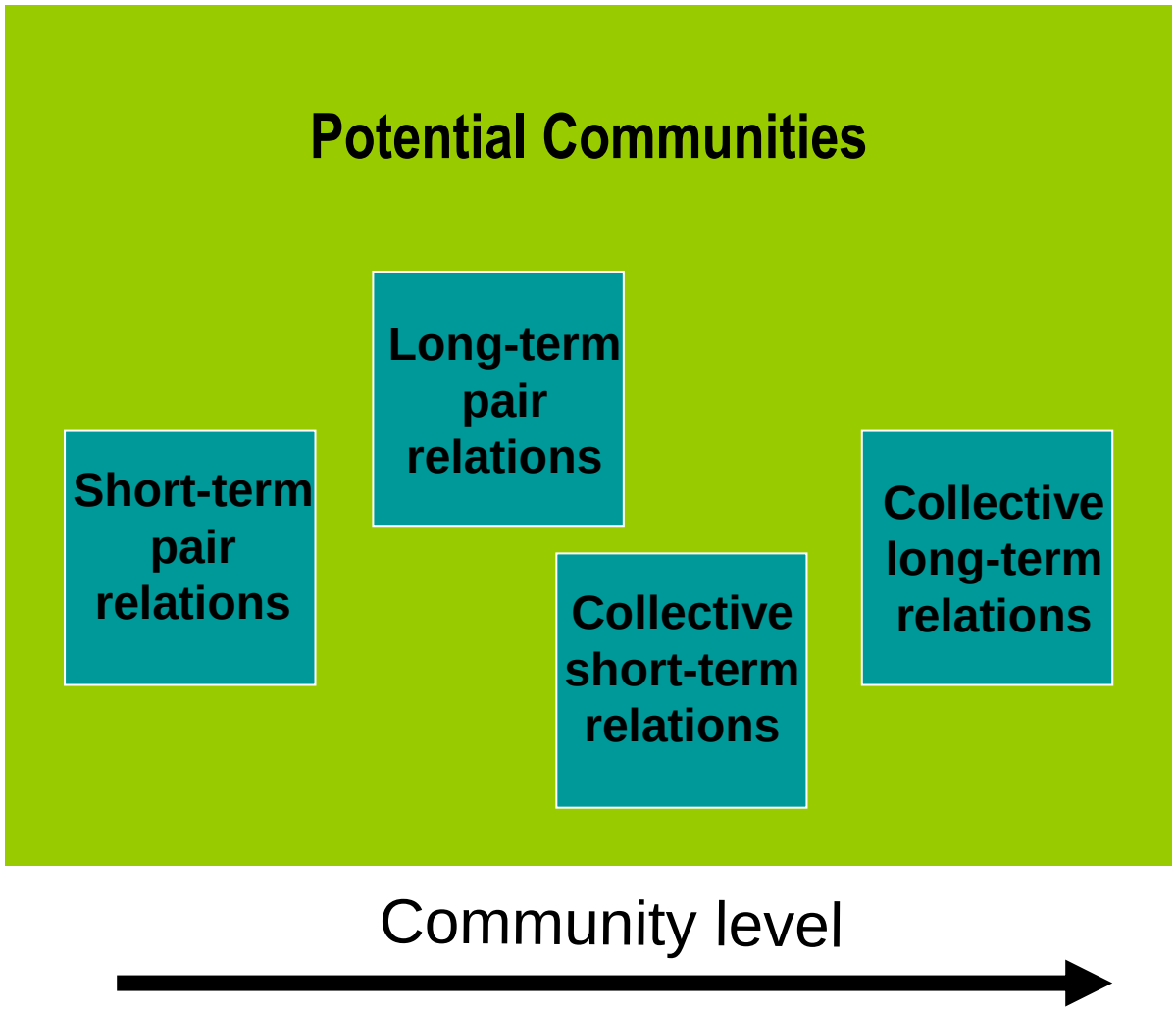
Features:

- **Keyword suggestion**
- **Document clustering**
- **Alerts**

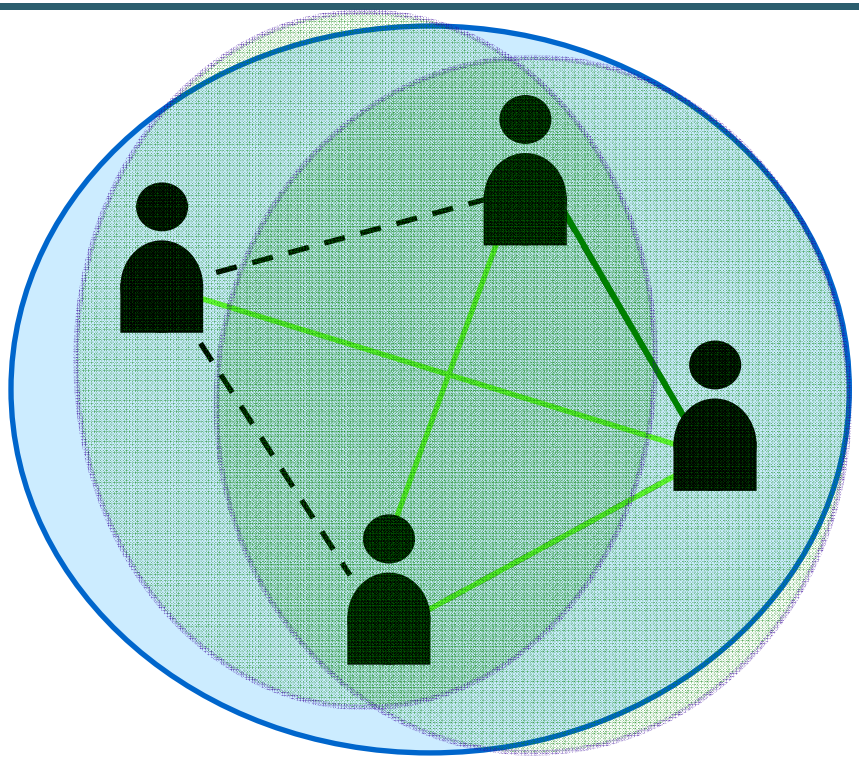
Community Model

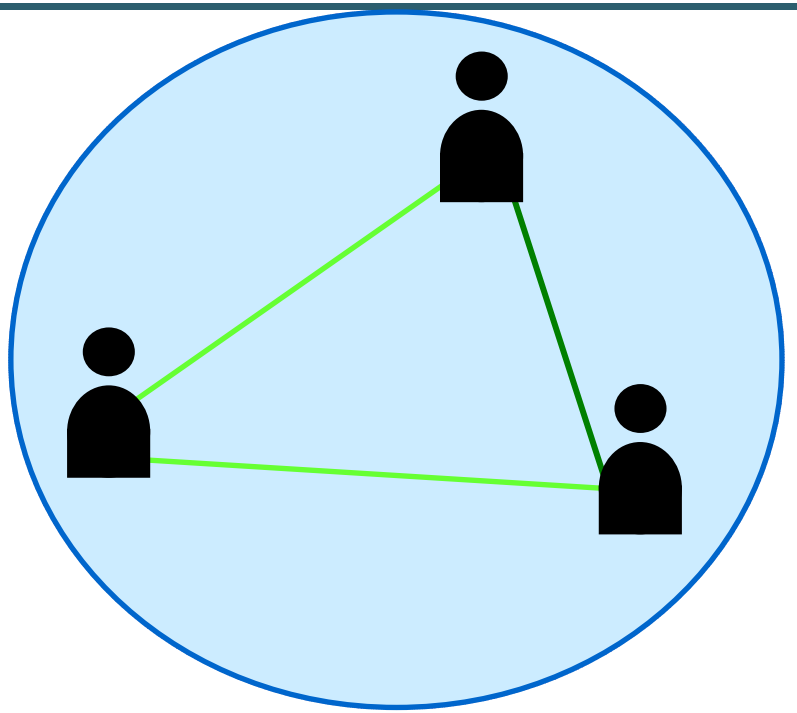
Community Model

The Gradual Community Model I



Community Model
The Gradual Community Model II

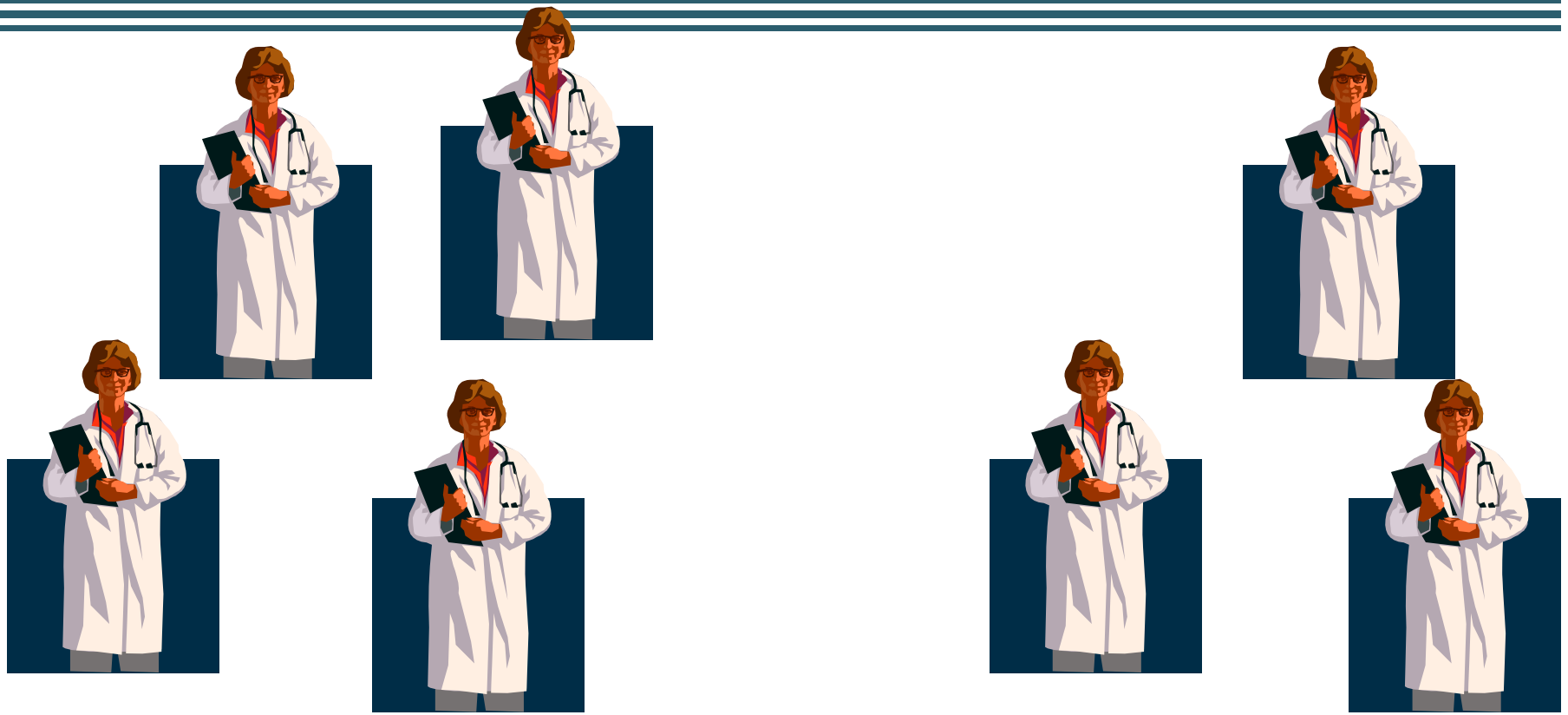




Collective short-term relations	**Criterion 1 + C2 + C3 + C4 + C5 + C6 + C7 + + community stereotype (C8) + role differential (C9)**	**C6: coordination, information** **C1: finding similar users**

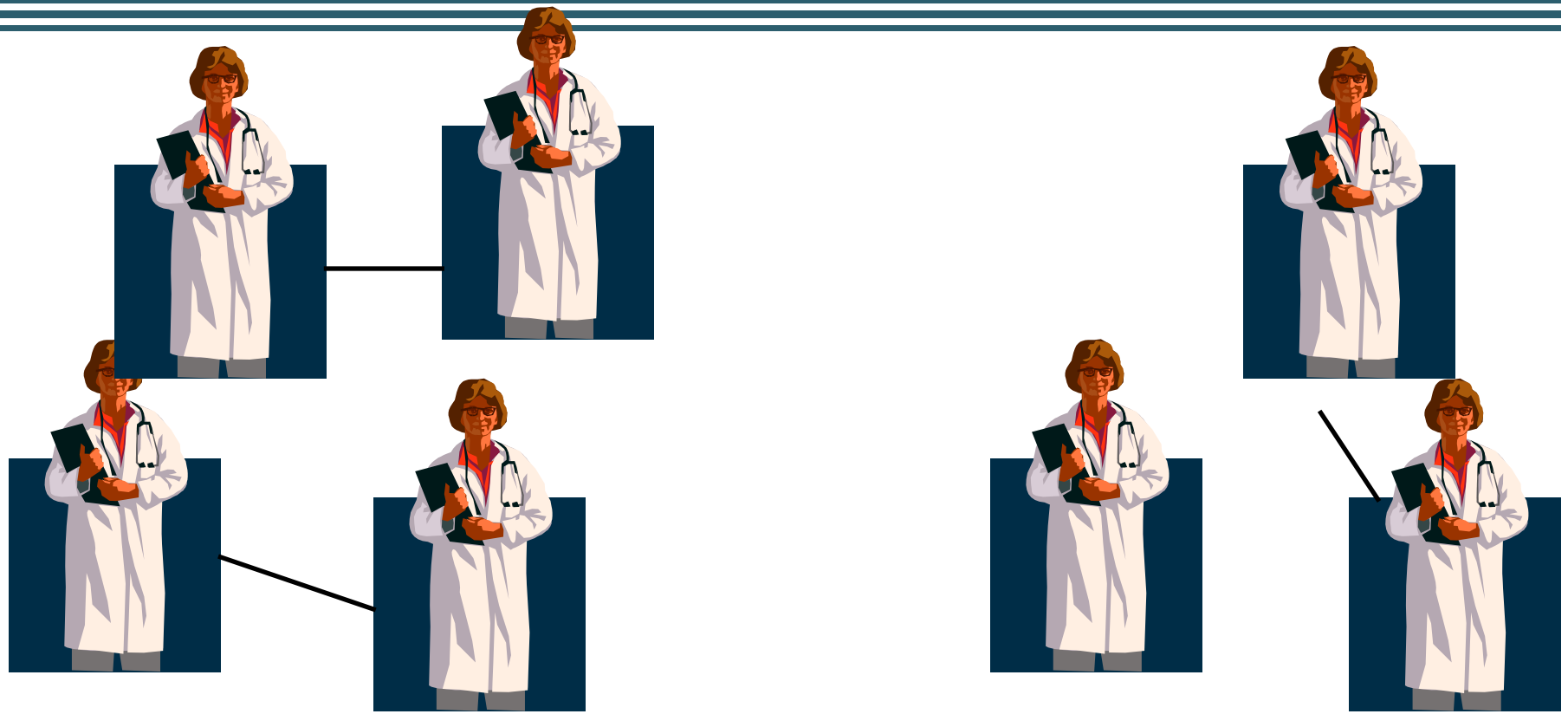
Community Model
The Gradual Community Model II

type	minimal community criteria	features
potential communities	pool of actors with something in common (C1)	C1: Finding Similar Users
short-term pair relations	Criterion 1 + mutual awareness (C2) + character / issue of interaction (C3)	C2: User-radar, User-page C3: offer several interaction channels
long-term pair relations	Criterion 1 + C2 + C3 + continuous interaction (C4) + external input (K 5)	C4: repository C5: Community Based Document Recommendation
collective short-term relations	Criterion 1 + C2 + C3 + community interaction (C6) + positive community feeling (C7)	C6: coordination, information control, distribution C7: conflict management, comparison
collective long-term relations	Criterion 1 + C2 + C3 + C4 + C5 + C6 + C7 + + community stereotype (C8) + role differential (C9)	C8: external representation C9: internal representation



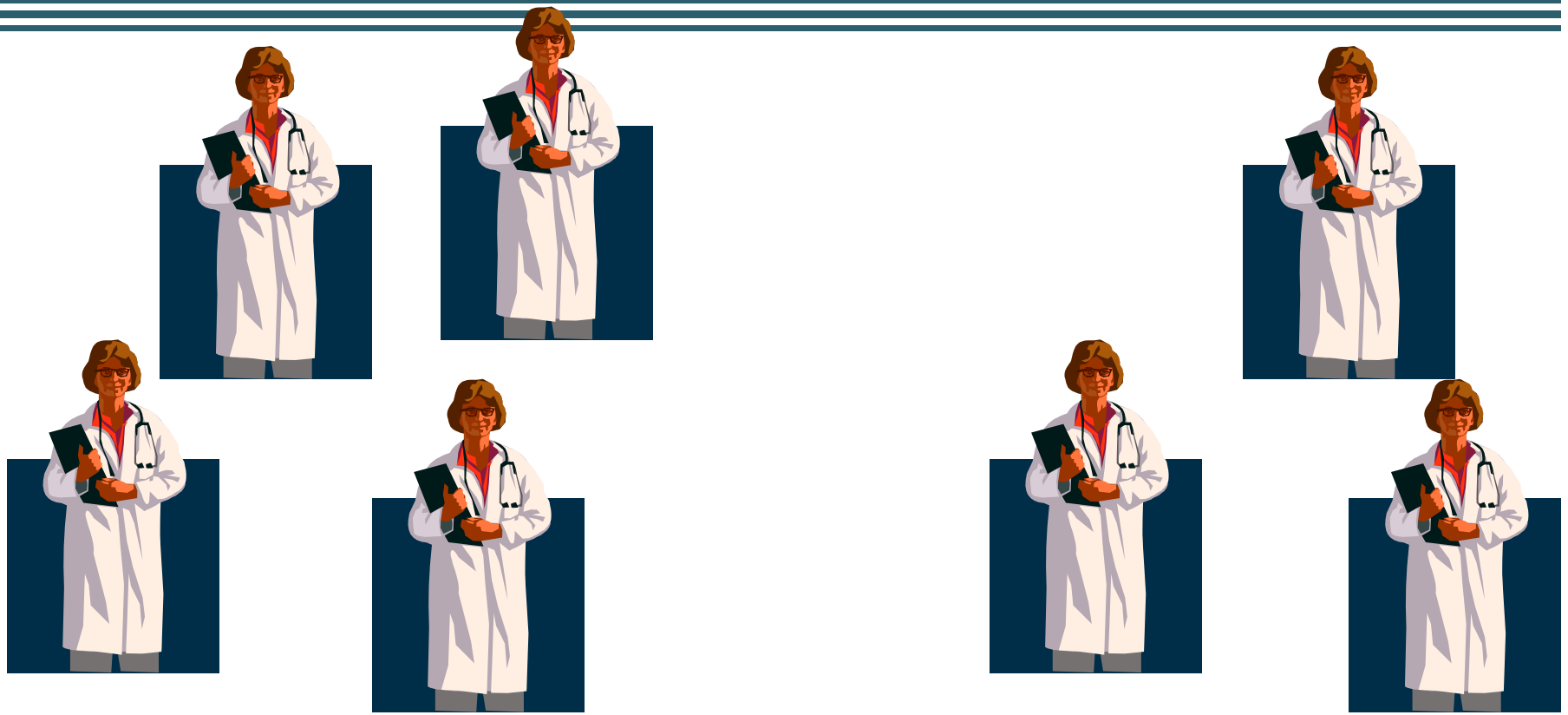
pool of actors with something in common (C1)

C1: Finding Similar Users



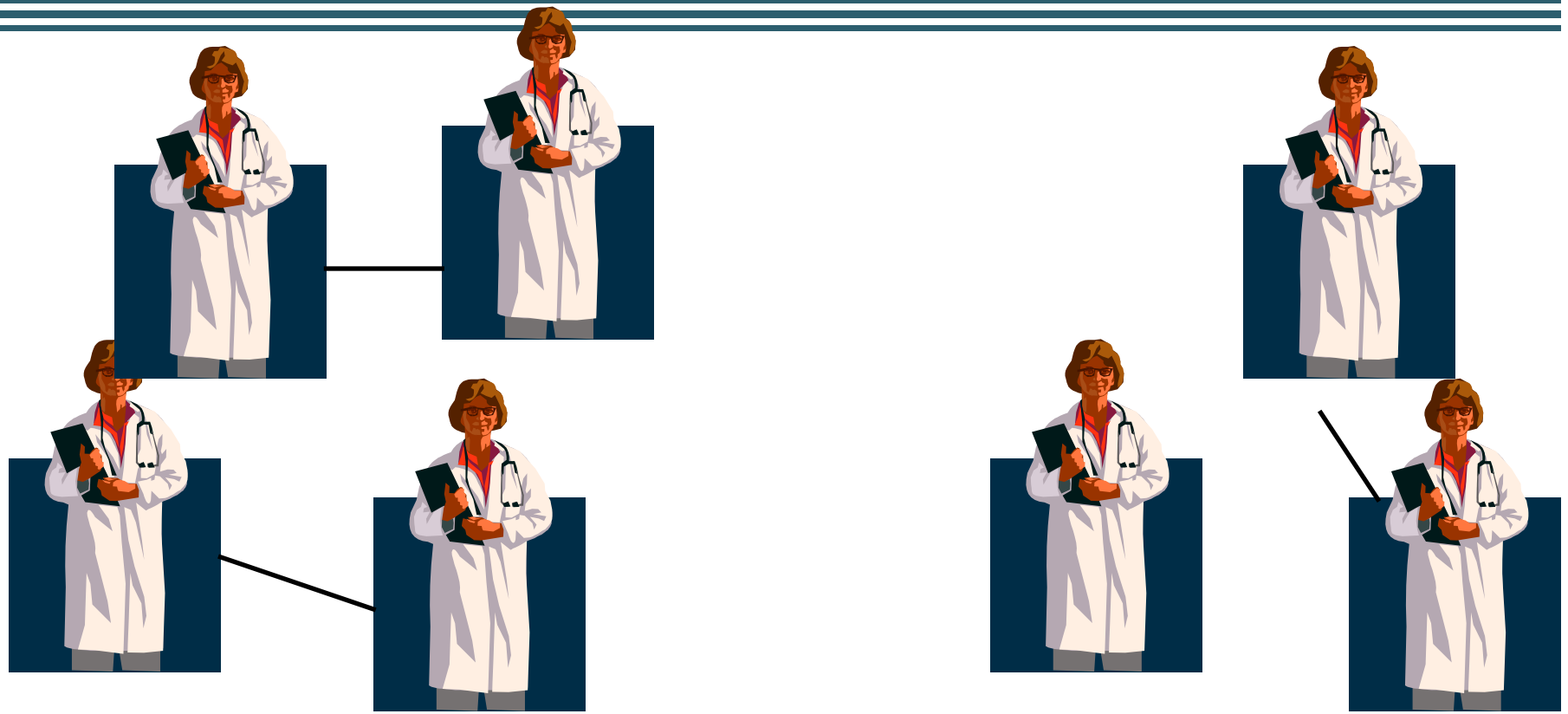
Criterion 1
+ mutual awareness (C2)
+ character / issue of interaction (C3)

C2: User-radar, User-page
C3: offer several interaction channels



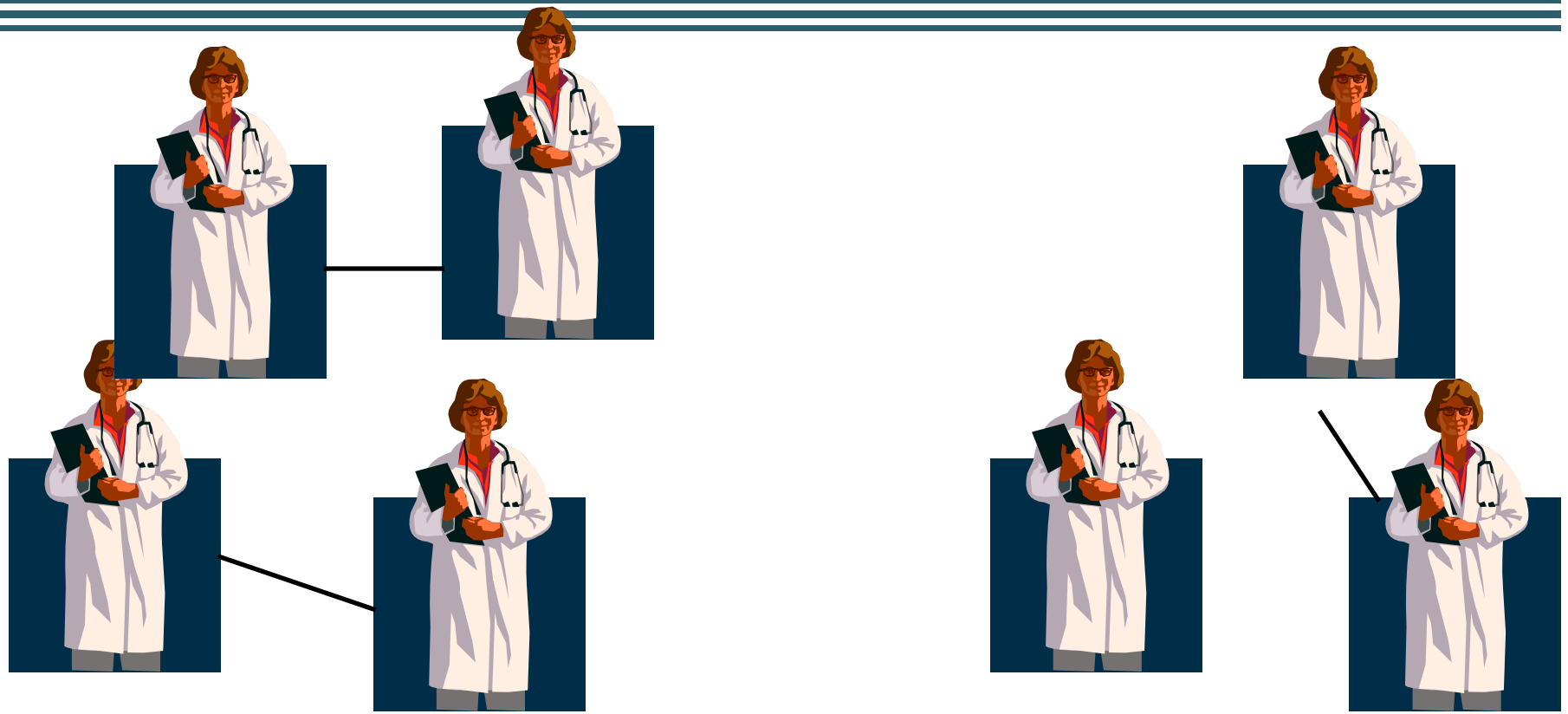
pool of actors with something in common (C1)

C1: Finding Similar Users



Criterion 1
+ mutual awareness (C2)
+ character / issue of interaction (C3)

C2: User-radar, User-page
C3: offer several interaction channels

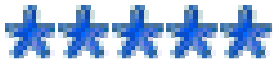


Criterion 1
+ mutual awareness (C2)
+ character / issue of interaction (C3)

C2: User-radar, User-page
C3: offer several interaction channels

Community Supporting Features

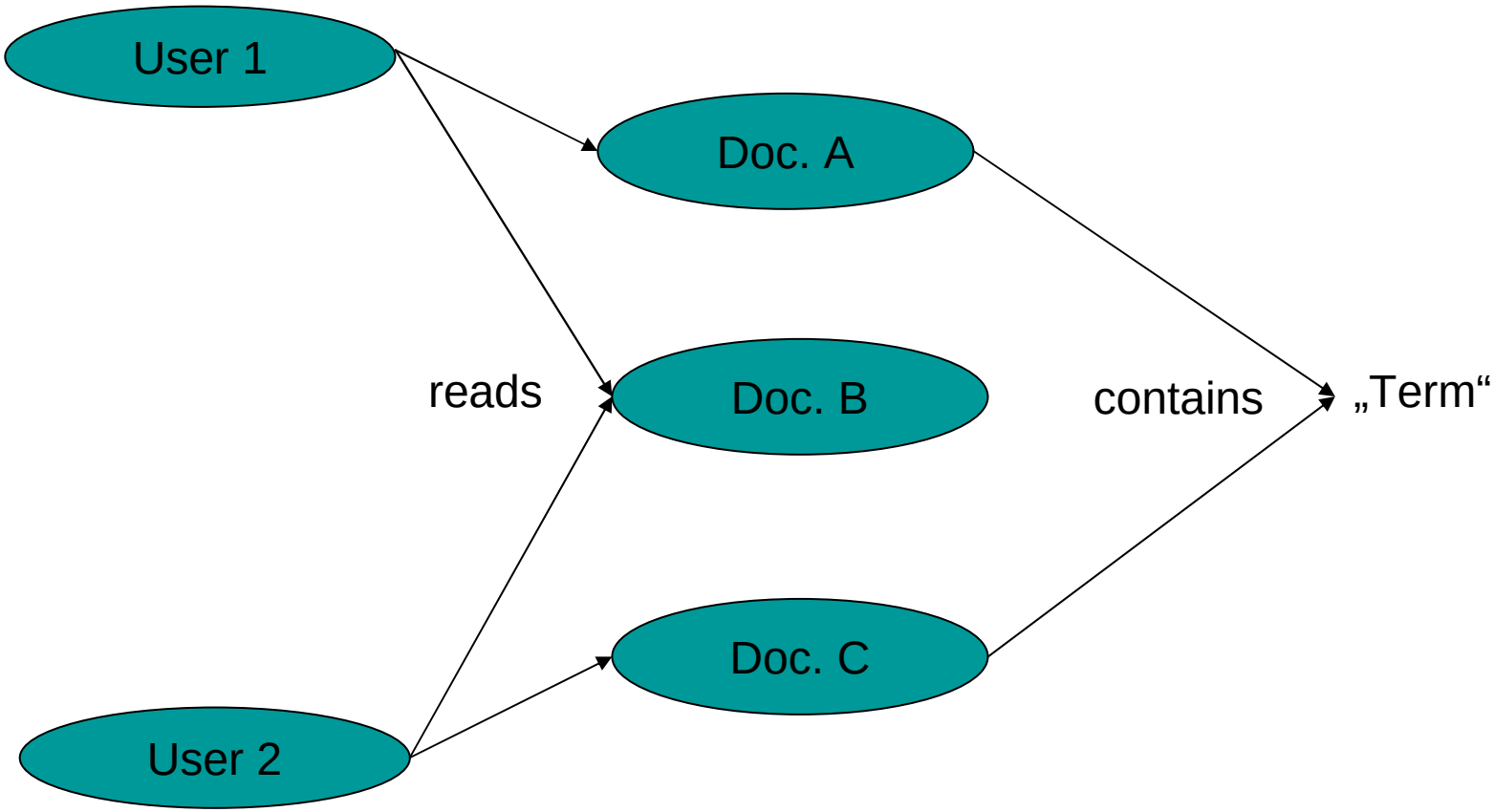
Community Support Features
Finding Similar Users

- ⇨ We want to determine how similar two users are to each other
- ⇨ Users read documents
- ⇨ Users rate documents 
- ⇨ Finding documents read in common: Not enough documents are present
- ⇨ Therefore: Finding similarities between documents

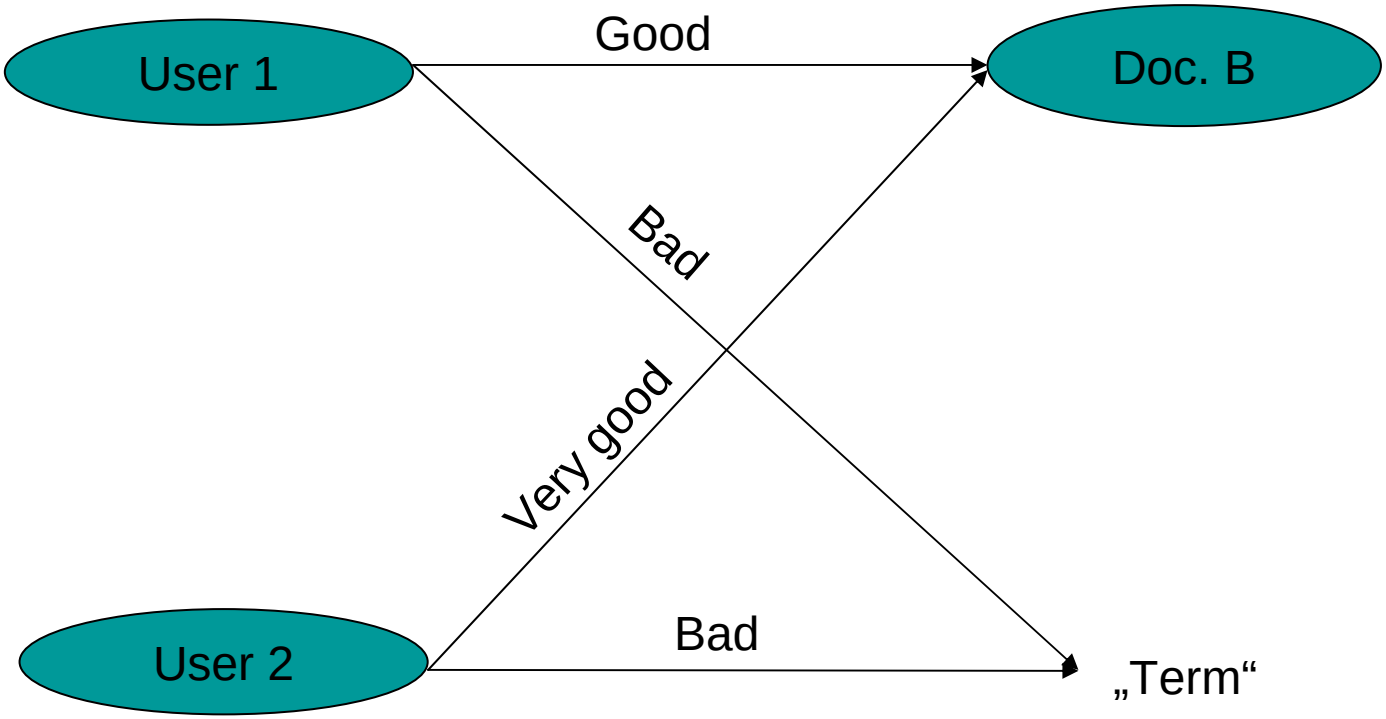
Community Support Features
Document Features

- ⇨ **Contained words**
- ⇨ **Weighted by number of appearences (term frequency)**
- ⇨ **Weighted by rarity (inverse document frequency)**
- ⇨ **Phrases**
- ⇨ **Synonyms, Metadata, References**

Community Support Features
Example

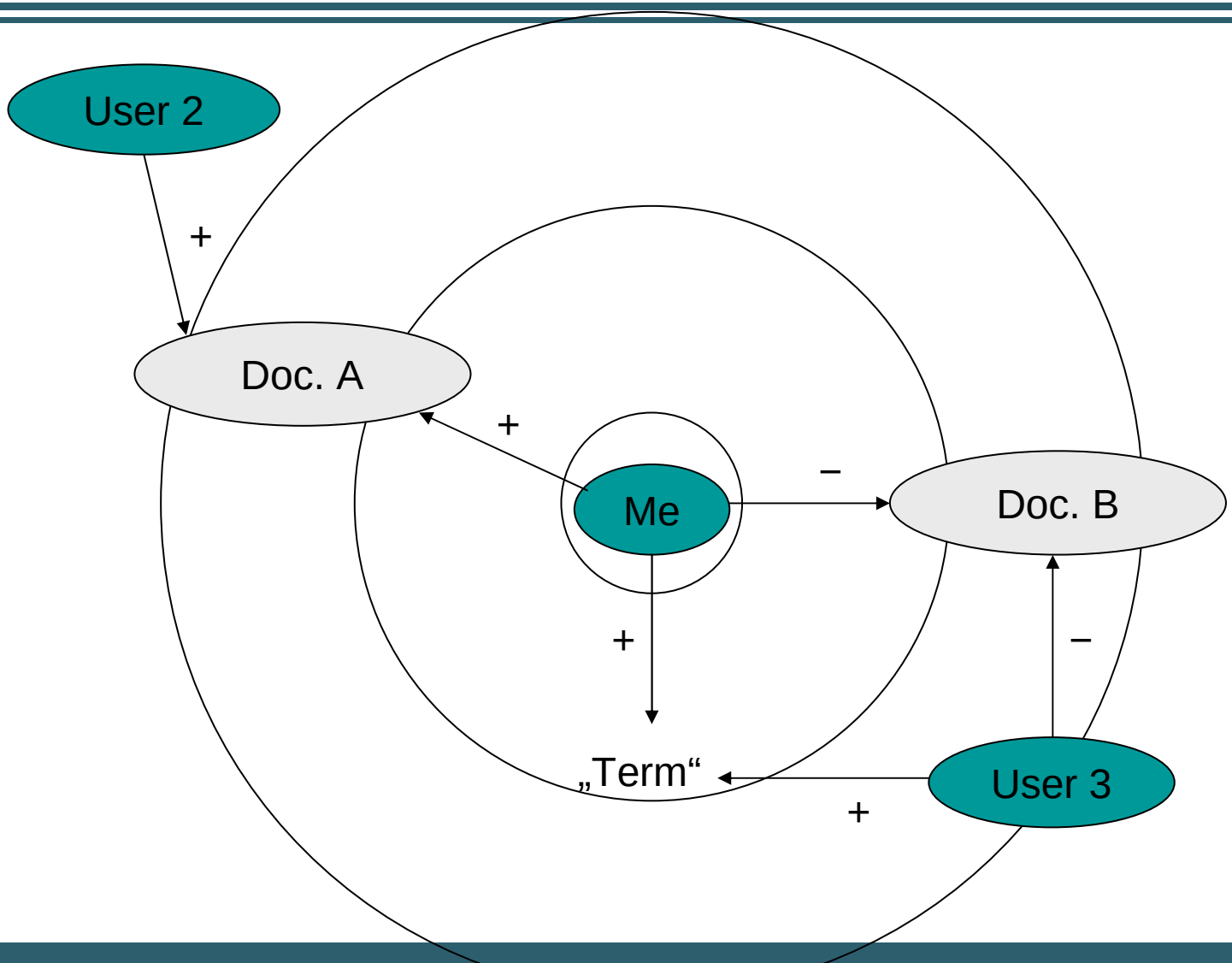


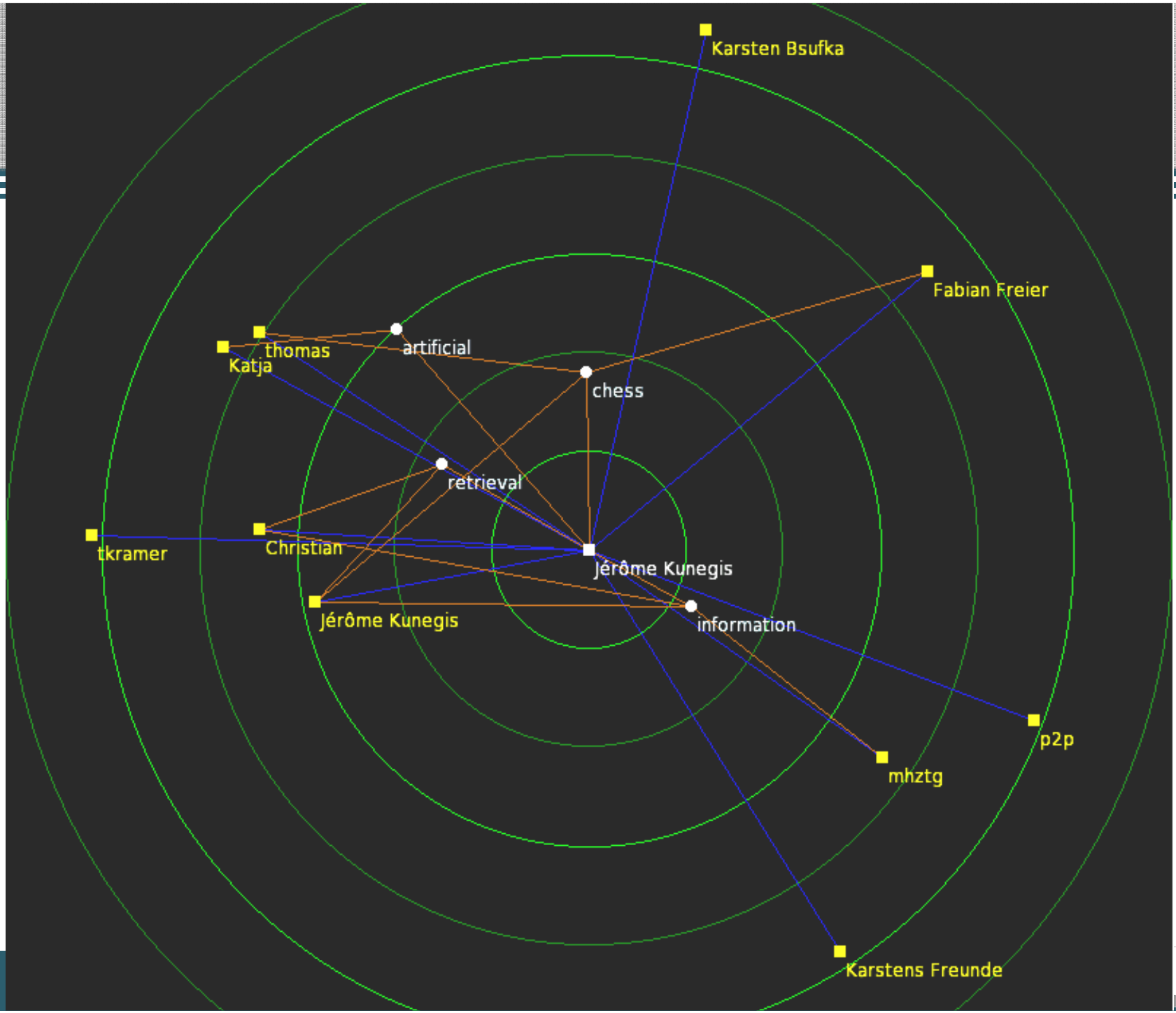
Community Support Features
Properties



Community Support Features

User-Radar




Karsten Bsufka
Fabian Freier
thomas
Katja
artificial
chess
retrieval
tkramer
Christian
Jérôme Kunegis
Jérôme Kunegis
information
p2p
mhztg
Karstens Freunde

Community Support Features
Community Based Document Recommendation

- ⇨ **Virtual communities tend to decay fast as member contributions decrease**
- ⇨ **Provide not only single users with documents, but find conitinuously up-to-date documents for the whole community!**
- ⇨ **This might help to keep communities alive**
- ⇨ **Train a classificator which identifies documents relevant for the whole community**

Community Support Features
Community Based Document Recommendation

- ⇨ **Unusual information retrieval task**
- ⇨ **Usually: single user, using query**
- ⇨ **Here:**
  - ⇨ **Classify documents for a whole community**
  - ⇨ **No query given**
- ⇨ **Because no benchmarks are available, we have to use synthetic benchmarks**

Community Support Features
Community Based Document Recommendation

The training set consists of all documents which community members have

- ⇨ **Explicitly rated as relevant**
- ⇨ **Saved (implicitly relevant)**
- ⇨ **Explicitly rated as irrelevant**
- ⇨ **Documents in search results which have not been rated (implicitly irrelevant)**

Community Support Features
Potential Support Vector Machine P-SVM

Linear classification

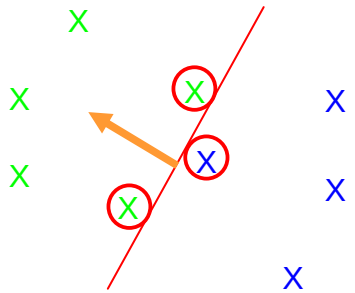
$y_i = \text{sign}(w \cdot x_i)$

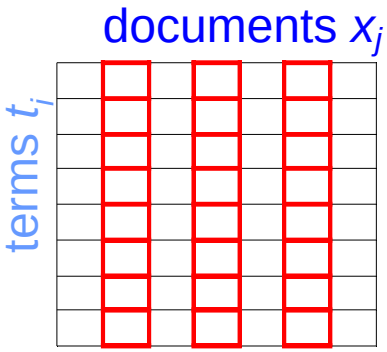
Nonlinear classification

$y_i = \text{sign}(k(w,x)_i)$

SVM:

$w = \sum \alpha_i x_i$



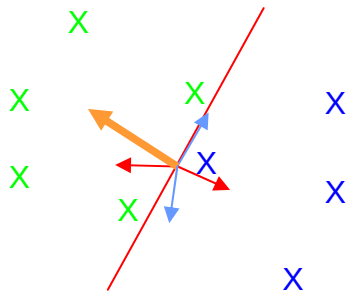


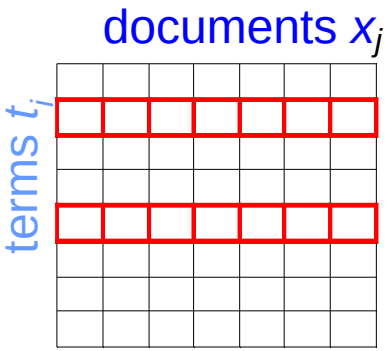
$K_{ij} = k(x_i , x_j)$

positive definite, square

P-SVM:

$w = \sum \alpha_i t_i$





$K_{ij} = k(t_i , x_j)$

no restrictions

Community Support Features
Potential Support Vector Machine P-SVM

Support Feature selection

- ⇨ **Saves time (more features than documents in training set)**
- ⇨ **Could help to characterize implicit community interests**

Conclusion and Outlook

Conclusion and Outlook

Type	Minimal community criteria	Features
potential communities	pool of actors with something in common (C1)	C1: Finding Similar Users
short-term pair relations	Criterion 1 + mutual awareness (C2) + character / issue of interaction (C3)	C2: User-radar, User-page C3: offer several interaction channels
long-term pair relations	Criterion 1 + C2 + C3 + continuous interaction (C4) + external input (K 5)	C4: repository C5: Community Based Document Recommendation
collective short-term relations	Criterion 1 + C2 + C3 + community interaction (C6) + positive community feeling (C7)	C6: coordination, information control, distribution C7: conflict management, comparison
collective long-term relations	Criterion 1 + C2 + C3 + C4 + C5 + C6 + C7 + + community stereotype (C8) + role differential (C9)	C8: external representation C9: internal representation