Social Information Retrieval
Experiment & Exercise

Social Gaming / Social Computing (IN0036+IN2241)

Christoph Fuchs

2015-05-20

WHY ARE YOU HERE?

1. **You** have the chance to **apply theoretical social computing concepts** using a newly generated dataset (which you **help to build**)
2. **You** are encouraged to come up with **own ideas and suggestions** for Social Information Retrieval
3. **You** can **improve your overall grade** of the lecture by participating in the exercises and submitting all 5 homework assignments

WHO IS THIS?

Christoph Fuchs

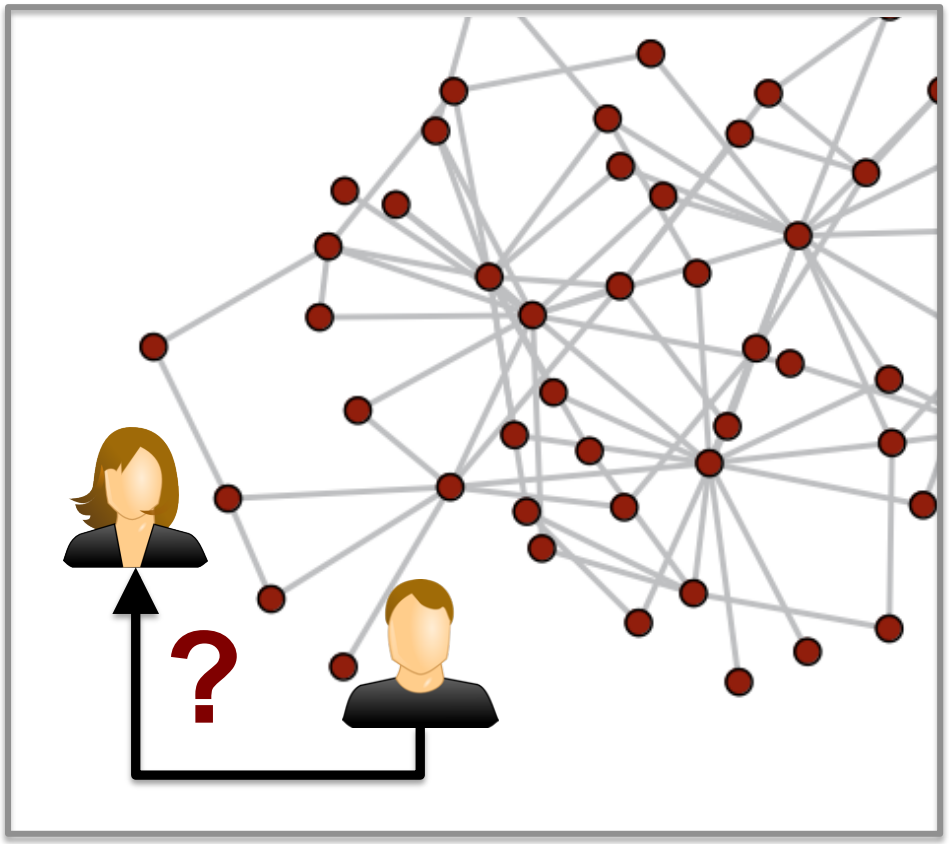
- 2007-08 Startup on **collaborative search** (andUNITE GmbH)
- 2009 Bachelor's thesis: **Social Networks in Science** (supervised by Georg Groh)
- 2011 Master's thesis: **Social Situation Detection** (supervised by Georg Groh)
- 2008-11 Developer for **real-time Linux kernel code** for specific requirements (atomic power plants, seismology, etc.)
- 2011- Consultant
- 2014- Educational leave; external PhD student: **Social Information Retrieval** (supervised by Georg Groh)

OUTLINE

1. **My research** on Social Information Retrieval
2. **Background information** on the Social Information Retrieval Experiment
3. **Homework Assignments** in Detail
4. **Questions** & Next Steps

SOCIAL INFORMATION RETRIEVAL

Access other people's private information spaces….



... in order to

- **Access** information **not publicly available**
- Use **social context as an amendment** for underdetermined queries
- Increase chance for **serendipity** due to social context

QUESTIONS OF INTEREST

1. How do **social context and interaction archetypes** influence **users' data sharing sensitivity**?

2. How **relevant** are **information items** taken from **non-public information spaces** of socially close people when satisfying information needs?

3. Does **social context imply** a valuable contribution to retrieving information from the **unconscious information need** (serendipitous information)?

4. Which **social concepts** impact the **users' routing decisions**?

5. Which **category of information needs** could **benefit** from **social information retrieval**?

OUTLINE

1. **My research** on Social Information Retrieval
2. **Background information** on the Social Information Retrieval Experiment
3. **Homework Assignments** in Detail
4. **Questions** & Next Steps

WHAT IS THE EXERCISE ABOUT?

Exercise = 5 homework assignments, incl.

- **Social Information Retrieval Experiment**
- **Hands-on tasks** related to **concepts from lecture**, e.g., random graphs, centrality measures, graph clustering, visualization, tie strength prediction, etc.

DISCLAIMER: NO WORRIES….

Following slides explain structure of the experiment (to show you the **big picture**)

– **later**: more detailed information!

BACKGROUND INFORMATION ON THE EXPERIMENT

Manual Mode

- Information seeker defines **query and recipients** for query
- **Recipients** of query **answer manually**

Automatic Mode

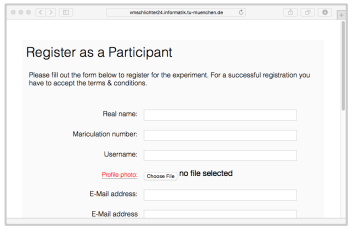
- Information seeker defines **query**
- **Recipients** of query **defined by system**
- Recipients get a **recommendation** which document from their information space might be a **suitable answer**
- **After confirmation**: Reply sent to Information Seeker through system

Social Semantic Product Search

- User searches for products
- Results are taken from Amazon
- **Products bought / viewed by friends** are added and/or highlighted accordingly (without sharing the name of the respective friend)

PREPARATION FOR EXERCISE

Register in web application

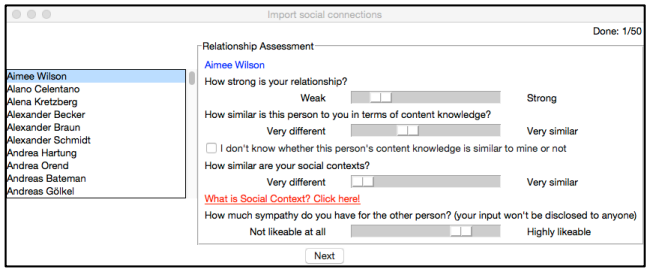


Upload anonymized network with weighted edges to web application

Use local application to identify & upload

- viewed products on Amazon (browser history)
- bought products (manual entry or custom JavaScript)

Download Facebook profile; assess tie strength, knowledge similarity, social context similarity and sympathy



- Parse browser history files for Chrome / Firefox,
- download visited URLs,
- calculate & upload individual topic model (LDA) on crawled corpus

<u>You are in control:</u>

- Edit list of Amazon products and URLs to crawl
- Names of friends only uploaded as hash
- More details about privacy: s. next slide!

PREPARATION – NOTES & REMARKS (1/2)

We take Data Privacy seriously: We don't do anything without your permission & we are transparent on what we do:

1. **We don't know the names of your Facebook friends**: We only ask you to upload the hashed version of the network file[1]
2. **None of your data** (evaluation of ties, names, products, etc.) **will ever be related to you**
3. **We don't get to know your browser history** – the uploaded topic models only contain document IDs and word vectors

1 The social network built from all individual ego-networks will get hashed again and randomly changed (edges and nodes changed/removed) before used in exercise to ensure that it is not possible to identify nodes

PREPARATION – NOTES & REMARKS (2/2)

4. The data will only be used to **conduct sound scientific research** (and student exercises)
5. You will use a (**obfuscated**) version of the dataset to **solve homework assignments**
6. Your **name, matriculation number, email address**, etc. **is only used to grade the homework** – it **<u>will be deleted afterwards</u>**
7. We will **never give** the data to anyone
8. Publications based on this dataset will only contain **highly aggregated information**
9. We will **hash the Amazon products after the experiment** and delete all copies of the original dataset

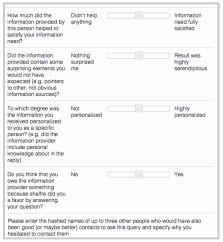
MANUAL MODE – OVERVIEW

Information Seeker (IS) **defines query**

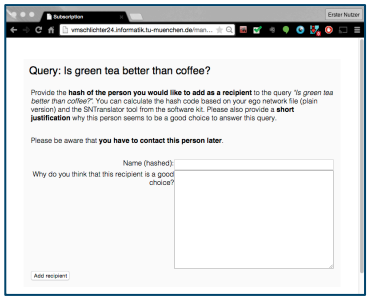
IS **sends query to designated IP**, incl. link to survey (outside of system)

IS **evaluates result** and fills out survey assessing

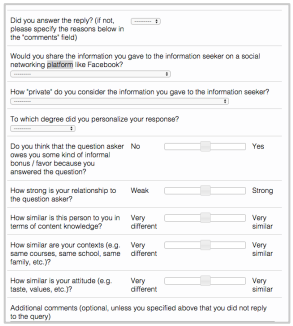
- Response **quality** (satisfaction, serendipity, personalization)
- Impact on **social capital**
- **Alternative** recipients



IS defines **recipients** (Information Providers, IP) for query (+rationale); IP is participant or member of social network



IP responds to query with information item from own information space and fills out survey assessing



- Data **sharing sensitivity**
- Degree of **personalization**
- Impact on **social capital**
- **Relation** to IS (tie strength, knowledge similarity, context similarity)

IP can also forward query

MANUAL MODE – SURVEYS (1/2)

Action

Information Seeker

Before sending request:

- Reasons why you chose this recipient
- Hashed name of recipient (you have to send request to potential information provider on your own, we don't even know her/his name…) – recipient must be part of your Facebook network or participant in lecture

Send request to potential information provider (e.g. via email using prepared email draft)

After receiving reply from information provider:

- Evaluate reply (w.r.t. general satisfaction, relevance, serendipity, personalization)
- Information about your feelings towards the person who answered (do you think that you owe her/him something for answering?)
- Up to 3 other contacts you would also consider asking, but you did not

Information Provider

Received your request with link to surveys. 2 options: reply to query or forward

Reply:

Clicks on link to reply to query & fills out survey covering the following topics:

- **Information sharing**: Did the information provider share information given as a reply already before on social media? Would information provider share it on social media channel?
- **Degree of privacy** for information
- Information about feeling towards Information Seeker: Do you think that she/he owes you something?
- **Assessment of tie to information seeker** (tie strength, knowledge similarity, social context similarity, sympathy)

MANUAL MODE – SURVEYS (2/2)

Action

Information Seeker

Before sending the request:

- Reasons why you chose this person
- Hashed name of the recipient (you have to send the request to the potential information provider on your own, we don't even know her/his name…) - the contact must be part of your Facebook network or participant in the lecture

Afterwards, send the request to the potential information provider (e.g. via email using a prepared email draft)

After receiving a reply from the information provider:

- Evaluate the reply (w.r.t. general satisfaction, relevance, novelty, personalization)
- Information about your feelings towards the person who answered (do you think that you owe her/him something for answering?)
- Up to 3 other contacts you would also consider asking, but you did not

Information Provider

Received your request with link to surveys. 2 options: reply to query or forward

Forward:

Fills out survey for information providers who forward the query

- Reasons why this request is forwarded
- Assessment of ties to second information provider (i.e., the receiver of forwarded request)

Sends email to new information provider (e.g. by using proposed email draft)

AUTOMATIC MODE – OVERVIEW

Detailed on next slides

Information Seeker (IS) **defines query**

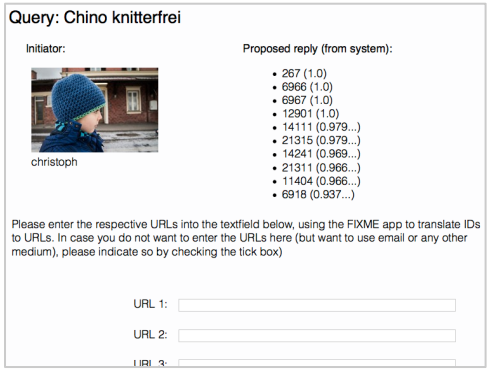
Background task
- **queries Topic Model/ LDA indexes** for each of the information providers

IS reviews and rates results:
- Satisfaction, serendipity, personalization (for each URL)
- Impact on social capital (for each answer)

2

System **randomly chooses strategy** to select information providers (tie strenth, knowledge similarity, social context similarity, sympathy)

- **Designated IP confirm response** and **add URL details** (index contains only IDs, not URLs)
- IPs **can send reply anonymously**



1

AUTOMATIC MODE – OVERVIEW

1 System identified you as potential Information Provider

1. Participant searched for query **which matches URLs in your Topic Model**
2. System **only knows index position of URLs**, doesn't know URLs itself
3. Sends a "**Data request**" to you
4. You can use **local tool** on your computer to **translate index positions** given from the system to "real" URLs
5. Please **enter URLs in web system** and submit form to send real URLs as reply to information seeker (can be done anonymously)

AUTOMATIC MODE – OVERVIEW

2 You received reply from other participant

1. Please review returned URLs
2. Evaluate each of returned URL (relevance, serendipity, personalization)
3. Rate overall satisfaction of results

SOCIAL SEMANTIC PRODUCT SEARCH (1/2)

Which children's seat to buy?

- Don't want to do research
- Maximum safety
- No regret

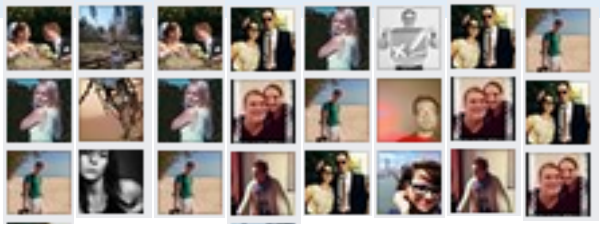
Traditional way: store

Primary interest: make $$$

Traditional experts

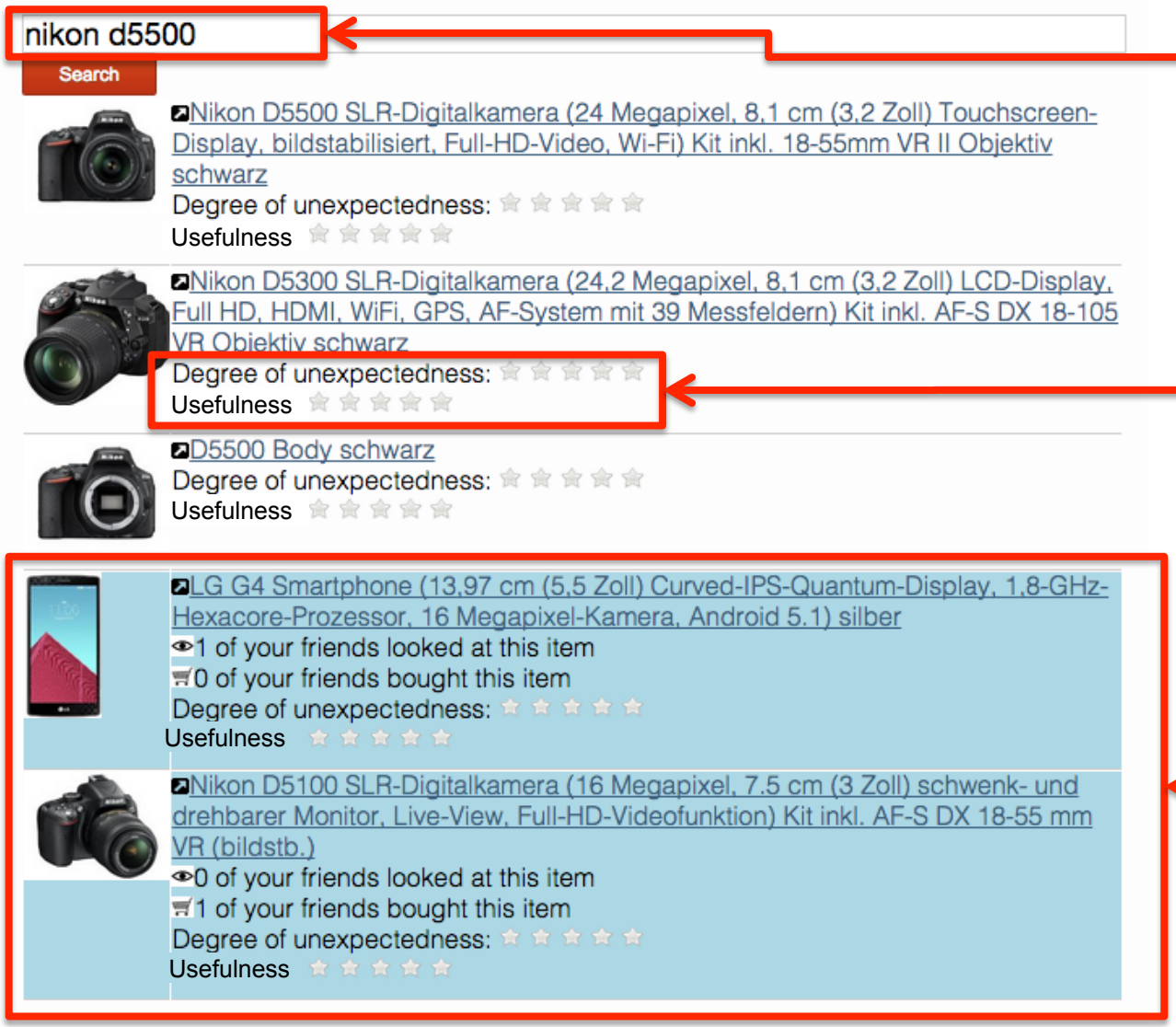
Give **general recommendations** (ease of installation, weight, ...)
Impartiality not guaranteed

Selected part of Social Network



Similar requirements
Additional requirements:
- Easy to clean
- Material not too hot in summer

SOCIAL SEMANTIC PRODUCT SEARCH (2/2)



Product you consider to buy

Please rate unexpectedness and usefulness of each result item

"Similar" products bought/viewed by friends (using multiple strategies to determine them)

OUTLINE

1. **My research** on Social Information Retrieval
2. **Background information** on the Social Information Retrieval Experiment
3. **Homework Assignments** in Detail
4. **Questions** & Next Steps

Homework Assignments (1/2)

- Homework assignments **always due on Wednesday, 23.59 CEST** (first: May 27th) via upload (emergency: see contact details!)
- Homework assignments consist of
  - **Tasks from experiment** (e.g., data collection, sending/ replying to questions, performing searches)
  - **Scientific tasks** (e.g., reading background literature, expressing own opinion, verifying results from other papers using our dataset)
  - **Hands-on implementation** work on datasets (e.g., clustering, calculation and interpretation of graph properties, etc.),

New to Python?: David M. Beazley (2012): Learning Python: Python Essential Reference (ISBN: 0672329786), available as eBook in TUM library

Homework Assignments (2/2)

- **Screencast video** available **each week** to support you using tools required for experiment
- *In a hurry?* **Fact Pack** incl. screenshots from the screencast video also available **each week**
- Any other trouble? **Contact me!**
  - **Cell:** +49 152 3427 2661
  - **Skype:** christophfuchs1139
  - **E-Mail / Google Hangout:** christoph@fuchs.cc
  - **Facebook:** https://www.facebook.com/christoph.fuchs2
  - **In Person:** MI 01.05.055 (by arrangement)

DISCLAIMER: NO WORRIES….

Following slides explain homework assignments **to give you an idea about the tasks**

– **online**[1]: more detailed instruc-tions (screencast, fact pack)!

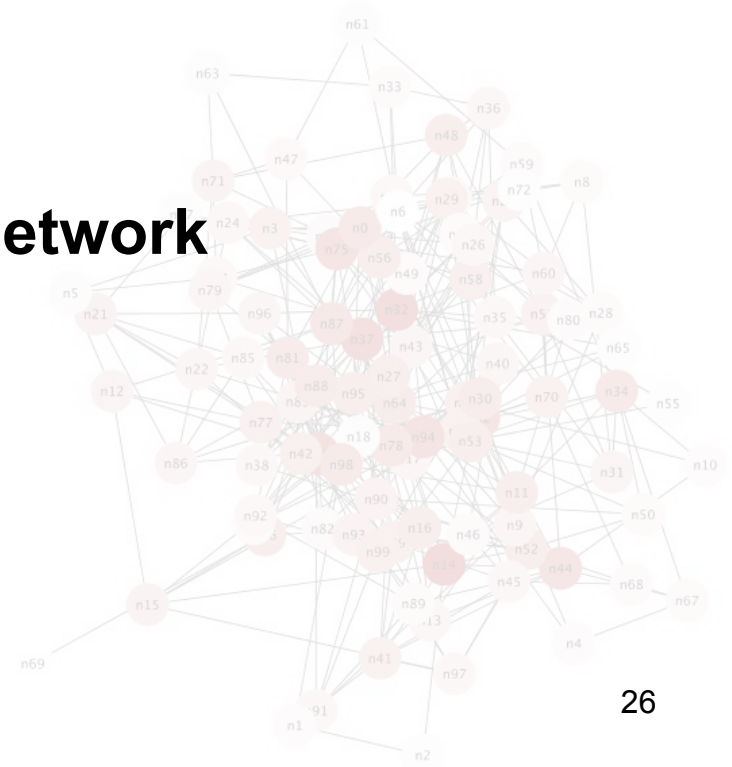
1 https://vmschlichter24.informatik.tu-muenchen.de

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 27.05. 23.59 CEST

Homework Assignment #1

- **Create your own account** at https://vmschlichter24.informatik.tu-muenchen.de
- **Upload a picture** of yourself to allow your fellow students to identify you!
- **Read background literature**
- **Generate and visualize random network**



Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 27.05. 23.59 CEST

Homework Assignment #1

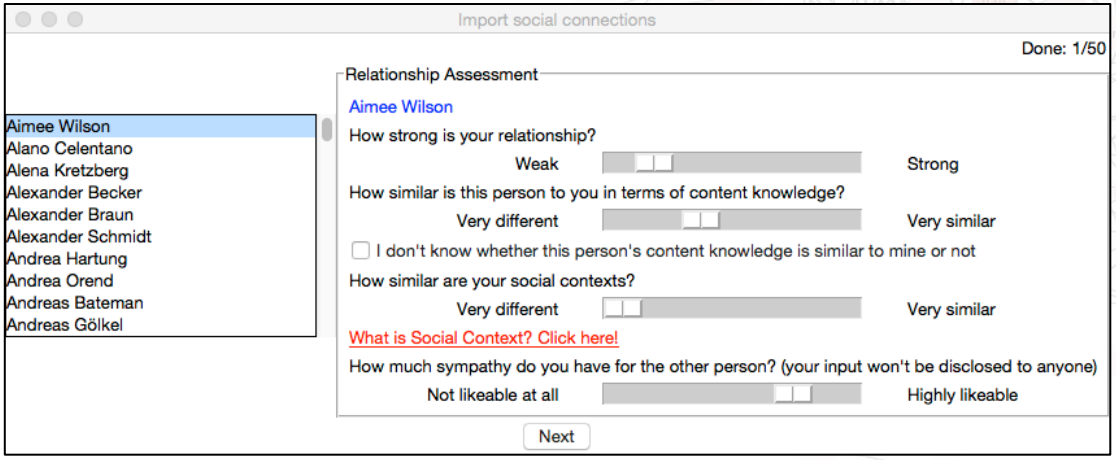
- **Create your own account** at https://vmschlichter24.informatik.tu-muenchen.de (=web system for exercise, there you will
  - find & upload homework assignments,
  - Download required **software kit** for experiment,
  - participate in experiment)
- Please **upload a picture** of yourself to allow your fellow students to identify you!
- **Read background literature** (available in web system)**:**
  - Oeldorf-Hirsch et al.: *To Search or to Ask: The Routing of Information Needs Between Traditional Search Engines and Social Networks*
  - Nagpal et al.: *Friends, Romans, Countrymen: Lend Me Your URLs. Using Social Chatter to Personalize Web Search*
- **Random networks:**
  - Install `Python 2.7` and `igraph` library on your computer
  - Generate and visualize random network (Watts-Strogatz)
  - Calculate betweeness centrality for each node
  - Visualize results (e.g., using Gephi, Cytoscape) & upload your code

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 03.06. 23.59 CEST

Homework Assignment #2

- Summarize your **personal perspective on Social Search** (300-400 words)
- **Generate & upload your individual ego-network**

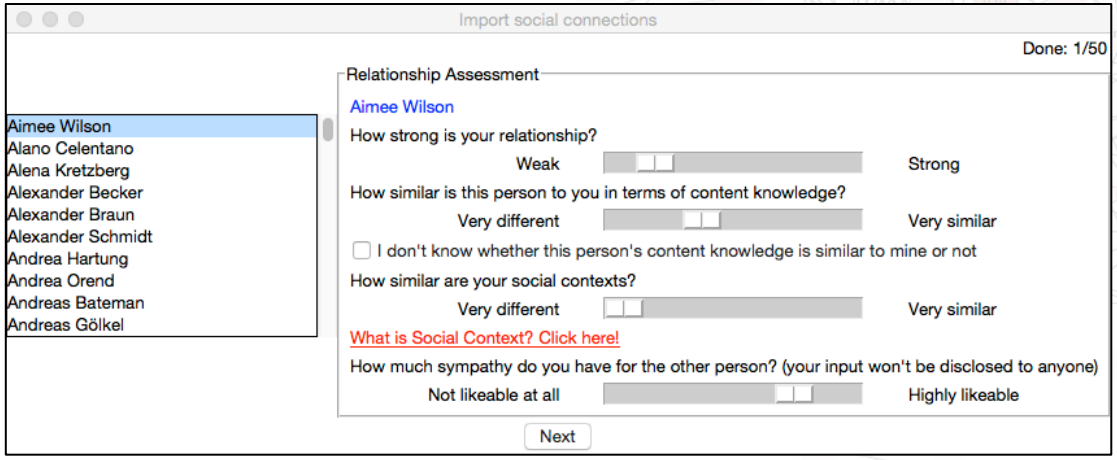


Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 03.06. 23.59 CEST

Homework Assignment #2

- **Take a stand**: Write document where you
  - Summarize your personal perspective on social search in approx. 300-400 words discussing at least 2 benefits and 2 drawbacks (last week's literature might be helpful)
  - List 3 potential information needs you could satisfy by asking people within your personal social network and 3 information needs which you would prefer not to solve by asking others. Please justify your decision briefly for each information need (~5-10 words)
- **Generate your individual ego-network** using `SNExtractor` tool from software kit, upload anonymized version (but don't delete plain version, you will need it later)

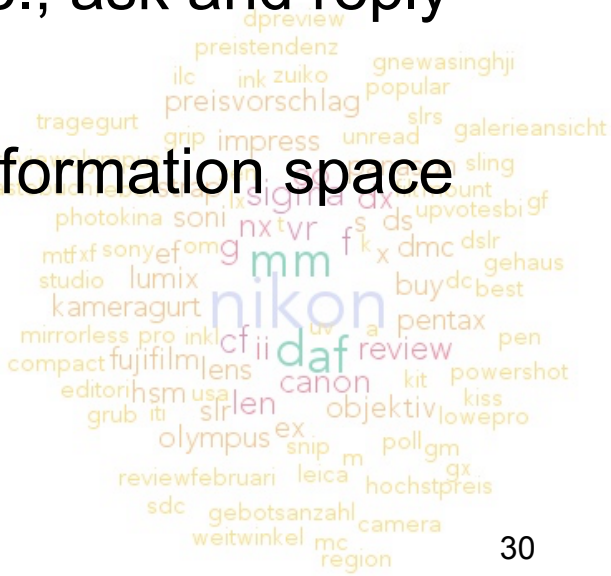


Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 10.06. 23.59 CEST

Homework Assignment #3 (1/2)

- **Build your private information space based on your browsing history**
- **Extract items you bought from and/or viewed** at at Amazon[1] (you can edit/adjust results before upload)
- **Start manual mode query process** (i.e., ask and reply to queries)
- **Analyze the detected topics** in your information space (50-100 words + rating)



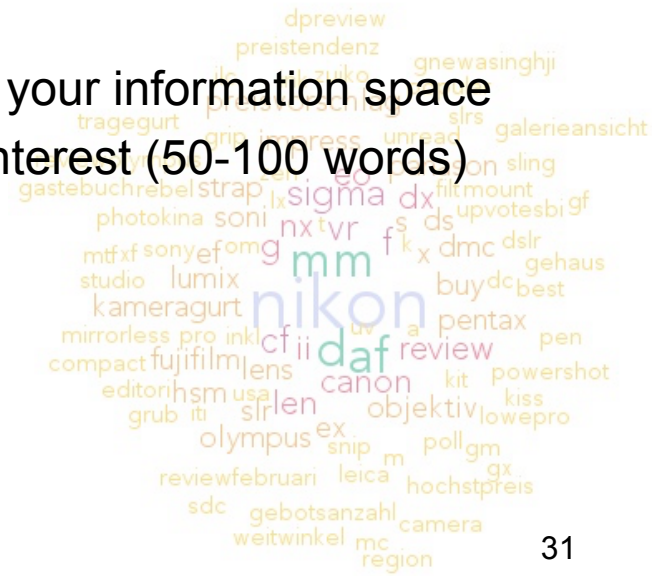
1 Part of the software kit – please see screencast and/or fact pack for details

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 10.06. 23.59 CEST

Homework Assignment #3 (1/2)

- **Build your private information space**
  - Download your web history using `history_crawler`[1]
  - Calculate topic models using `indexer`[1]
- **Extract items you bought from and/or viewed** at at Amazon[1] (you can edit/adjust results before upload)
- **Start the manual mode query process** (i.e., ask and reply to queries which get answered manually)
- **Analyze the detected topics** (`CloudCreator`) in your information space and explain whether they represent your areas of interest (50-100 words)



1 Part of the software kit – please see screencast and/or fact pack for details

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 10.06. 23.59 CEST

Homework Assignment #3 (2/2)

Homework Assignment #4

- **Start automatic query process**
- Continue to maintain manual query process (e.g., evaluating results)
- **Analyze “our” social network**

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 17.06. 23.59 CEST

Homework Assignment #4

- **Start automatic query process** (i.e., add 6 queries to system for automatic mode)
- Continue to maintain manual query process (e.g., evaluating results)
- **Analyze "our" social network:**
  - Use `igraph` to create a Python script that
    - Identifies number of components in network
    - Applies two different clustering algorithms
    - Calculates diameter, density, clustering coefficient of network
  - Comment on results in report (~50-100 words): Which results are surprising, which not? Give reasons for your answers.

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 24.06. 23.59 CEST

Homework Assignment #5

- Search at least for 10 items you consider to buy using the **Social Product Search Interface**
- Summarize in bullet points **what you like and what you don't like about social product search** in general
- Analyze the **correlation between spatial distance and tie strength for our dataset**

Liben-Nowell et al., 2005: Geographic routing in social networks, Proceedings of the National Academy of Sciences of the United States of America, Vol. 102 (33) p11623-11628

Primary (and binding) place for homework assignments is https://vmschlichter24.informatik.tu-muenchen.de

Due: 24.06. 23.59 CEST

Homework Assignment #5

- Try **Social Product Search Interface** with at least 10 queries (and rate results using stars) – search for items you plan to buy in near future
- Summarize in bullet points **what you like and what you don't like about social product search** in general
- **Finalize query sessions** in automatic and manual mode (i.e., conduct outstanding evaluations)
- Analyze **correlation between spatial distance and tie strength for dataset (**e.g., in R):
  - Download CSV file with given pairs of (spatial distance, tie_strength)
  - Define suitable threshold to determine binary friendship relation from tie_strength
  - Sample distance accordingly in fixed buckets and calculate probability for friendship for each bucket, interpret results (~10-20 words)
  - Compare your results with results of (Liben-Nowell et al., 2005) (~50 words)

Liben-Nowell et al., 2005: Geographic routing in social networks, Proceedings of the National Academy of Sciences of the United States of America, Vol. 102 (33) p11623-11628

OUTLINE

1. **My research** on Social Information Retrieval
2. **Background information** on the Social Information Retrieval Experiment
3. **Homework Assignments** in Detail
4. **Questions** & Next Steps

HELP & SUPPORT

1. **Screencast Video and Fact Pack each week**
2. **Piazza**
3. **Other**
   - **Cell:** +49 152 3427 2661
   - **Skype:** christophfuchs1139
   - **E-Mail / Google Hangout:** christoph@fuchs.cc
   - **Facebook:** https://www.facebook.com/christoph.fuchs2
   - **In Person:** MI 01.05.055 (by arrangement)

NEXT STEPS

Participate & submit Homework Assignments in time:

- HA #1 27.05.
- HA #2 03.06.
- HA #3 10.06.
- HA #4 17.06.
- HA #5 24.06.

It's not possible to join after HA #1!

THANK YOU!

QUESTIONS?